MONASH WARWICK ALLIANCE

MONASH University

WARWICK
THE UNIVERSITY OF WARWICK

Manifestations of Impulsivity and Self-Control

Simon Thomas van Baal

Doctor of Philosophy

A Thesis Submitted for the Degree of Doctor of Philosophy at
Monash University & University of Warwick in 2022

Cognition & Philosophy Lab
School of Philosophical, Historical and International Studies
Faculty of Arts

Abstract

Many decisions and behaviours involve impulsivity and self-control; people deal with urges, cravings, and temptations by deliberating, contemplating counterfactual or future scenarios, and committing themselves to policies for action. People who live in developed economies require more self-control than ever, as our environment relies on consumers succumbing to their impulses; fast food, addictive smartphone apps, one-click online purchasing, and other temptations are ubiquitous.

A vast body of research notwithstanding, we lack a comprehensive theoretical framework for understanding how impulsivity and self-control manifest in decision-making and behaviour. As a result, we are unable to provide a satisfactory answer to the question of how impulsivity and self-control relate to one another: is one the absence of the other, or is the relationship more complex? This gap in understanding limits our research efforts, and thereby our opportunities for implementing systemic change, decision aids and behavioural interventions.

Using methods from philosophy, psychology and economics, this interdisciplinary thesis examines, conceptually and empirically, manifestations of impulsivity and self-control. I develop a conceptual framework of what it is to be impulsive or self-controlled, such that we may better understand how these cognitive processes manifest in people's lives. The framework reveals conceptual limitations inherent in much theorising and experimental work, pointing to a more nuanced understanding, with the potential to open new avenues of research and intervention.

Aspects of the conceptual framework are then unfolded in empirical studies; I examine the roles of impulsivity, self-control, and deliberation in time preferences. I use ecological momentary assessment to study how people experience and deal with urges in everyday life and I investigate how people balance tending to their wants and needs against prosocial behaviour using survey experiments and an economic game.

I leverage my understanding of impulsivity and self-control to develop and test behavioural interventions designed to help people make better decisions by providing deliberation aids and changes in incentives. Specifically, I assess whether prompting people to contemplate a positive future or consequences to others can change the way they experience urges, and whether eliciting evaluative judgment or meta-preferences can change choice patterns. Moreover, I investigate in a public health setting whether different communication styles and a change in uncertain collective costs can promote prosocial behaviour. Findings reveal that these interventions indeed can have a positive effect on our decision-making, consistent with the conceptual analysis.

Overall, the thesis provides a nuanced picture of how impulsivity and self-control manifest in daily life and tests interventions aimed at helping people deal with their impulses, and with balancing needs and wants against prosociality and goals for the future.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes four original papers published in peer-reviewed journals, one submitted article that is in the peer-review process, and one article that has not yet been submitted to a journal.

The core theme of the thesis is an examination of the manifestations of impulsivity and self-control. The ideas, development, and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Philosophical, Historical, and International Studies at Monash University, and the Department of Psychology at University of Warwick under the supervision of Professor Jakob Hohwy (Monash), Associate Professor Lukasz Walasek (Warwick), and Professor Antonio Verdejo-García (Monash).

The inclusion of co-authors for all articles reflects the fact that the work was produced through active collaboration between researchers and acknowledges input into team-based research.

My contribution to Chapter 1 involved the following: conceptualisation, conducting the literature review, visualisation, drafting the initial version, and overseeing and implementing revisions of subsequent versions. For Chapters 2, 4, 5, and 6, my involvement entailed conceptualisation, designing the study, conducting the study, project administration, curating data, analysing data, visualisation, and drafting the initial version, and overseeing and implementing revisions of subsequent versions. For Chapter 3, my involvement entailed data curation, analysis, visualisation, interpretation of results, initial drafting of the Chapter, and overseeing and implementing revisions to subsequent versions.

Monash University requests that published papers are inserted in their published form. I have not renumbered sections of published papers, but I have added page numbers consistent with the rest of the thesis to those papers for ease of navigation.

TABLE 1: A table describing the publication status of the main thesis chapters (excluding the conclusion)

Ch.	Title of Submission	Status	Nature and % Student Contribution	Nature and % of Co-author's Contribution	Co-author Monash Student
1	*Impulsivity and self-control as timeless concepts*	Under review	75%; conceptualisation, writing.	1) Lukasz Walasek, input into manuscript, 10%; 2) Antonio Verdejo-García, input into manuscript, 5%; 3) Jakob Hohwy, input into manuscript, 10%.	N
3	*State impulsivity amplifies urges without diminishing self-control*	Published	60%; administration, data analysis, writing.	1) Neda Moskovsky, experimentation, input into manuscript, 15%; 2) Antonio Verdejo-García, conceptualisation, input into manuscript, 15%; 3) Jakob Hohwy, input into manuscript, 10%.	N
4	*Episodic future thinking and compassion reduce public health guideline noncompliance urges: A randomised controlled trial*	Published	75%; conceptualisation, experimentation, administration, data analysis, writing.	1) Antonio Verdejo-García, conceptualisation, input into manuscript, 12.5%; 2) Jakob Hohwy, conceptualisation, input into manuscript, 12.5%.	N
5	*Risk perception, illusory superiority and personal responsibility during COVID-19: An experimental study of attitudes to staying home*	Published	70%; conceptualisation, experimentation, administration, data analysis, writing.	1) Lukasz Walasek, input into manuscript, 5%; 2) Daniela Karanfilovska, validation, input into manuscript, 5%; 3) Allen Cheng, validation, input into manuscript, 5%; 4) Jakob Hohwy, conceptualisation, input into manuscript, 15%.	N
6	*Modelling pandemic behaviour using an economic multiplayer game*	Published	80%; conceptualisation, software, experimentation, administration, data analysis, writing	1) Lukasz Walasek conceptualisation, input into manuscript, 10%; 2) Jakob Hohwy conceptualisation, input into manuscript, 10%.	N

Student name: Simon Thomas van Baal

Date: 19 May 2023

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor names: Jakob Hohwy & Lukasz Walasek

Publications during enrolment

*__Van Baal, S.T.__, Verdejo-García, A., Hohwy, J. (2023). Episodic future thinking and compassion reduce non-compliance urges regarding public health guidelines: a randomised controlled trial. *BMC Public Health*. https://doi.org/10.1186/s12889-023-15031-0

*__Van Baal, S.T.__, Karanfilovska, D., Cheng, A.C., Walasek, L., and Hohwy, J. (2022). Risk perception, illusory superiority and personal responsibility during COVID-19: An experimental study of attitudes to staying home. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12554

*__Van Baal, S.T.__, Moskovsky, N., Hohwy, J., Verdejo-García, A. (2022). State impulsivity amplifies urges without diminishing self-control. *Addictive Behaviors*. https://doi.org/10.1016/j.addbeh.2022.107381

*__Van Baal, S.T.__, Walasek, L., Hohwy, J. (2022). Modelling pandemic behaviour using an economic multiplayer game. *Scientific Reports*. https://doi.org/10.1038/s41598-022-17642-w

__Van Baal, S.T.__, Le, S.T.T., Fatehi, F., Hohwy, J., Verdejo-García, A. (2022). Cory COVID-Bot: An Evidence-Based Behavior Change Chatbot for COVID-19, in *Informatics and Technology in Clinical Care and Public Health*, Amsterdam: IOS Press. https://doi.org/10.3233/SHTI210948

Hanegraaf, L., __Van Baal, S.T.__, Hohwy, J., Verdejo-García, A. (2021). A systematic review and meta-analysis of 'Systems for Social Processes' in borderline personality and substance use disorders. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2021.04.013

Michael, J., __Van Baal, S.T.__ (2021). Key themes in recent research on interpersonal functioning in borderline personality disorder. *Current Opinion in Psychiatry*. https://doi.org/10.1097/YCO.0000000000000649

Westra, E., Terrizzi, B.F., __Van Baal, S.T.__, Beier, J.S., Michael, J. (2021). Beyond avatars and arrows: Testing the mentalising and submentalising hypotheses with a novel entity paradigm. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218211007388

* = Included in the thesis

Acknowledgements

I would like to thank my supervisors, Jakob Hohwy, Lukasz Walasek, and Antonio Verdejo-García. During my PhD, I was continuously aware of my incredible privilege of having such an amazing team of supervisors. Not only were they generous with their time beyond what could be expected of any supervisor, but they have also shaped me as an academic and as a person. In particular during the pandemic, Jakob was more than an excellent supervisor, he was also my mentor and advocate.

I am also grateful to various other senior academics whose input has had large impact on me and my research, among whom are Manos Konstantinidis, John Michael, Wenjia Joyce Zhao, and Suong Le. They all made time to share their knowledge and advice on how to be a good academic.

My fellow lab members at the Monash Centre for Consciousness and Contemplative Studies have also been an important part of my journey. I highly appreciate their support, their openness in discussing our multidisciplinary work, and their wonderful attitudes towards science. In particular, I would like to thank Niccolo Negro, who was always generous with his time discussing visa antics, Andrew Corcoran, who made room in his house during the first months of my PhD, Kelsey Perrykkad and Jonno Robinson, who used their experience to pave the way for the supportive atmosphere in the lab, and Beth Fisher and Tom Darling, who are arguably the main drivers of the lab's general liveliness. This is not to discount my appreciation of the other members of the lab, who have all added greatly to my experience and ability to complete this thesis.

Fellow PhD students at the Warwick Psychology Department greatly contributed to me feeling welcome and valued during the last phase of my PhD; something that is especially at a premium when moving overseas for only six months. In particular I would like to thank Lucas, Yuqi, and Elif for their warmth and good conversations.

I would also like to extend my gratitude to my friends and family, who were important sources of support during my PhD. Their kindness, understanding, and compassion were of great help during the many COVID-19 lockdowns, travel restrictions, and other difficulties that sprung up over the years. Notably, my academic aspirations deprived my sister, Sophie, of her brother's presence in difficult times, but she remained supportive of me throughout—I am grateful.

Please note that I did not engage the services of a professional editor for any part of this thesis.

This research was supported by the Monash Warwick Alliance Accelerator Scheme.

Contents

Abbreviations

Acronym	**W**hat (it) **S**tands **F**or
ACT	**A**cceptance and **C**ommitment **T**herapy
AIC	**A**kaike **I**nformation **C**riterion
CI	**C**onfidence **I**nterval
CoViD-19 or **COVID-19**	**Co**rona **Vi**rus-20**19**
EFT	**E**pisodic **F**uture **T**hinking
FFBI-SF	**F**ive **F**actor **B**orderline **I**nventory–**S**hort **F**orm
LL	**L**arger-**L**ater (reward)
OR	**O**dds **R**atio
SS	**S**maller-**S**ooner (reward)

Chapter 1

Introduction: Dissociating Impulsivity, Self-Control, and Time Preferences

1.1 Background

The tradition of analysing and reflecting upon impulsivity and self-control goes back to the Ancient Greek philosophers. Aristotle, in his *Nichomachean Ethics*, contemplated *akrasia*, consisting of *propeteia* and *astheneia* (Ameriks & Clarke, 2000). To him, it appeared that sometimes we do things that are not in our best interest, either because we do not think through the consequences properly before we decide or act (propeteia) or because the temptation to do what we know is detrimental to us is too strong (astheneia).

Aristotle's understanding of these concepts mirrors much of the contemporary pre-theoretical view of what impulsivity and self-control in daily life entail. To be impulsive would be, for instance, to say 'yes' when your friend asks if you want to go to an all-you-can-eat shrimp buffet, while momentarily forgetting you promised to pick up your child from childcare. This corresponds to Aristotle's propeteia—impulsivity in an action stems from a lack of consideration of the consequences. A lapse of self-control would be to know one should sit down and start reading a lengthy thesis document (because

it has to happen at some point anyway), but succumb to the temptation to read something more entertaining instead. This corresponds to Aristotle's astheneia—a lack of self-control is constituted by not acting in accordance with one's better judgment.

The key distinction between impulsivity and self-control in these two cases lies in the cause of the behaviour, not the behaviour itself. For the former case, the lack of consideration causes the agent's behaviour, but it is possible that, had they deliberated the consequences, they might have declined the shrimp buffet. Likewise, deciding to read something more entertaining would be un-self-controlled because the agent believes that a different course of action would be better (Mele, 2010), not because of a disregard for consequences. I will discuss this distinction in more detail further on in this Chapter.

Note that while these examples are relatively small-scale, such decisions can be highly consequential when the decision-maker holds high public office or when the environment is unforgiving. This makes it important to understand impulsivity and self-control, since such understanding enables us to predict when agents are likely to be impulsive and when they will control themselves, providing us with the opportunity to intervene appropriately where necessary.

There are several interesting developments in the last decades of scientific research on impulsivity and self-control that have changed the way the constructs are understood in research as opposed to daily life. First, researchers from various disciplines started to develop their own understandings of the computational or neural underpinnings of the concepts and operationalised them in different ways, which then fed back into how they defined the constructs in their research. Second, on a background of increasingly individualistic framing of behaviour in the research literature—particularly with the rise of so-called "nudging" (Thaler & Sunstein, 2009)—much impulsivity research focuses on eradicating or minimising it, while research on self-control focuses on bolstering it. These studies often do not take into account the context that may contribute to these behaviours. Although the idea of adaptive impulsivity (Fenneman & Frankenhuis, 2020; Gullo & Dawe, 2008; Stevens & Stephens, 2010) and maladaptive self-control is gaining traction (e.g., Steinglass et al., 2012).

Further on in this first introductory Chapter of the thesis, I will explore how these and other developments have changed our understanding of impulsivity and self-control, and see if these changes are desirable from a theoretical and practical perspective. Before I do

that, though, I will provide a short overview of how researchers from various disciplines tend to describe and write about impulsivity and self-control.

In psychology, impulsivity is often conceptualised as a lack of inhibition, especially for quick responses and habits (Evenden, 1999; Whiteside & Lynam, 2001). Impulsivity is an especially popular concept in psychology because it has been identified as a major underlying factor for various mental disorders in various iterations of the Diagnostic and Statistical Manual for Mental Disorders (DSM; American Psychiatric Association, 2010) and addictive disorders (Verdejo-García et al., 2007) but also because it underlies risky behaviours in general (Bakhshani, 2014).

Self-control is often referred to as a composition of willpower and using planning to avoid exerting willpower (e.g., by removing tempting options) (Duckworth et al., 2018; Rachlin & Green, 1972). In developmental psychology, self-control is often equated with the capacity to delay gratification (Mischel & Ebbesen, 1970).

In behavioural economics and psychology, impulsivity is often implicated in time preferences, the degree to which people value rewards and costs in the present compared to in the future, in particular when characterised as *decreasing devaluation* of rewards as they move further into the future (i.e., rewards lose more subjective value when their receipt is postponed by one day from today, than when it is postponed by one day a month from now; Ainslie, 1975). This phenomenon, often referred to as *present bias*, is common among humans and animals, and across different domains of rewards (Mazur, 1987). The perceived tie between impulsivity and present bias has led to a situation where measuring time preferences to make inferences about agents' impulsivity is common in neuroscience (Madden et al., 2004) and neuroeconomics, in humans and in non-human animals such as mice, rats, and pigeons (Vanderveldt et al., 2016).

Self-control is often discussed in economics as the lack of present bias or a low rate of discounting future outcomes more generally. This connection is made because failures to act in one's best interests are normally associated with accepting immediate rewards (Laibson, 1997). However, this leads to the (often tacit) implication that impulsivity is the converse of self-control, which is not obviously correct (I will discuss this in more detail later on in this Chapter). But there are also conceptualisations of self-control that are based on thwarting temptation (Dekel et al., 2009; Gul & Pesendorfer, 2001), though this type of view often also explicitly relies on temporal aspects of temptation

(i.e., only temptations that pertain to present rewards) and self-control (i.e., only for goals for the future).

One issue with maintaining different conceptualisations and operationalisations of impulsivity and self-control across these various fields is that it has led to the incorporation of increasingly tangential phenomena into the understanding of impulsivity and self-control (e.g., see Caswell et al., 2015). An example from the psychology literature is 'cognitive impulsivity'—the inability to inhibit prepotent responses—often measured using a *Go-No Go* task. In the Go-No Go task, participants need to press a key on a computer keyboard when they are presented with a 'Go' stimulus (e.g., food), but not respond when presented with a 'No Go' stimulus (e.g., non-food). Participants are instructed to go as fast as possible, and to try to be as accurate as possible (a trade-off that is difficult to navigate). The experimental software is usually set up such that the Go stimuli occur far more often than No Go stimuli, lulling participants into a false sense of security until the No Go surprises them. If participants are unable to sufficiently suppress their prepotent responses in the No Go trials, they are said to be impulsive.

In this task, and many other tasks designed to measure impulsivity and self-control, it is unclear whether the resulting measurements are tapping into a meaningful aspect of these concepts. What does it mean when someone performs poorly on a Go/No Go task? Is this really an important part of what it is to be impulsive in real life? Considering the lack of conceptual agreement on impulsivity and self-control, it is difficult to assess this theoretical fit between the various measures and the two concepts.

The lack of conceptual clarity prompts similar questions about manifestations of impulsivity and self-control in real life behaviours: which decisions and behaviours are impulsive? Which are self-controlled? If a behaviour is impulsive, or if it was caused by a self-control failure, is that always maladaptive or counterproductive? If someone chooses to delay gratification to obtain a future reward, is that always a marker of self-control? Currently, we have no satisfactory framework to answer these questions, and thus no good way of predicting when people's dispositions lead them to do something that they will regret or that leads to decreases in (individual and societal) welfare or wellbeing.

In sum, the research literature lacks a comprehensive theoretical framework for understanding impulsivity and self-control, which means that we are also unable to provide

a satisfactory answer to the question of how impulsivity and self-control relate to one another: is one the absence of the other, or is the relationship more complex? This lack of a comprehensive framework limits us in our research efforts, in implementing systemic change, decision aids, and behavioural interventions; if we do not know what role impulsivity or self-control play in causing the behaviour, we cannot hope to intervene in these cognitive processes to prevent it from happening.

1.2 Thesis Plan

In this thesis, I develop a novel conceptual framework for impulsivity and self-control, use that to assess how the concepts relate to time preferences, see how people deal with impulses in everyday life, and evaluate how we can intervene in these cognitive processes. I do this in an attempt to answer two questions:

1 How do impulsivity and self-control manifest?

2 How might we help individuals deal with their impulsivity and limited self-control to obtain better outcomes for the individual or the collective?

To be able to answer these questions I will first consider the background of impulsivity and self-control, building the foundation for the conceptual framework. To do so, I will build on both pre-theoretical conceptions and commonly used definitions from the literature to arrive at an understanding of the two concepts. This understanding will allow me to evaluate the connection between impulsivity, self-control, and time preferences (which are often suggested to represent impulsivity and self-control)—the task for the rest of this Chapter.

Chapter 2 will build on this framework, evaluating how higher-order preferences can influence intertemporal choice (the most common experimental means for measuring time preferences). This is relevant to both of the research questions since a lack of deliberation about consequences is a core feature of impulsivity and since higher-order preferences are thought to be central to self-control. (The thought being that an agent needs self-control only if they have overarching preferences (e.g., about who they want to be, or what they want to want) that conflict with 'lower-order' preferences, often these pertain to bodily needs or anything deemed a tempting vice by the agent.) Thus,

understanding how higher-order preferences influence intertemporal choice will bring us closer to understanding how impulsivity and self-control manifest, as well as how we may help people deal with their impulsivity and limited capacity for self-control.

In Chapter 3 and Chapter 4, guided by the framework in Chapter 1, I will elucidate the role of impulsivity and self-control in experiencing and dealing with urges in daily life, and evaluate the possibility of intervening in these processes by influencing people's mental state.

In Chapter 5 and Chapter 6, I investigate the learning process and epistemology of balancing wants and needs against the collective good, assessing what communication, social norms, and changes to the environment can contribute to better decision-making. These two Chapters are meant to extend the examination of impulsivity and self-control, and the model developed in Chapter 1 to the domain of decisions where people trade off selfish options with prosocial ones.

The thesis, then, can be seen as consisting of the current introduction, three distinct but related sections, the first being the current Chapter and Chapter 2, concerning impulsivity, self-control, and their connection to time preferences; the second section consists of Chapter 3 and Chapter 4, concerned with how people deal with urges; the third consists of Chapter 5 and Chapter 6, extending impulsivity and self-control to the selfish/prosocial domain, and Chapter 7, the Conclusion. The thesis will transition from the theoretical, across to the fundamental, and over to the applied and situational. Another distinction within the thesis splits it in half: Chapters 1-3 are Chapters that pertain to tendencies in preferences, choices, and decisions that underpin many everyday situations in life, whereas the work conducted in Chapters 4-6 is done in the context of the COVID-19 pandemic, which has disrupted many people's habits, incentives, and options, making for an interesting case study that aids in answering the research questions in an applied context, critical for public health.

The methodology for this thesis is comprised of tools from philosophy, psychology, and economics. I use logical arguments to dissociate concepts, use intertemporal choice experiments to conduct foundational research on time preferences, surveys and ecological momentary assessment to investigate beliefs and manifestations of impulsivity and self-control in everyday life, and finally an economic game to investigate how people learn

to make decisions in an unfamiliar setting, and how environmental parameters influence this process.

1.3 Building the Theoretical Foundation for the Thesis

The rest of this Chapter will start us off on the journey of understanding manifestations of impulsivity and self-control. I will build the theoretical foundation for the rest of the thesis in this Chapter and discuss how impulsivity, self-control, and time preferences relate to one another.

This Chapter aims to bring us closer to understanding the manifestations of impulsivity and self-control by enabling us to tell apart cases of impulsivity from a lack of self-control, present-orientation from impulsivity, and future-orientation from self-control. This is a good place to start in understanding impulsivity and self-control because in the literature on time preferences, it has traditionally been assumed that future-oriented decision-making is indicative of self-control, and present-oriented decision-making is indicative of impulsivity. The conceptual foundation I build in this Chapter will lead to a challenge to these notions, for which I rely on logical arguments reinforced with evidence from the literature.

The work in this Chapter will also form the foundation for answering the second research question of the thesis, of how we might help individuals to deal with their impulsivity and self-control to achieve better outcomes for themselves and the collective good. Developing this new, more comprehensive account of how impulsivity, self-control, and time preferences come apart will be highly useful in practical scenarios because it helps us identify what may happen when we intervene in these cognitive processes. For example, if we aim to increase future-oriented decision-making, we must know whether stimulating more deliberation (i.e., less impulsivity) will necessarily result in more future-oriented decision-making, or that this intervention may backfire in certain contexts. Gauging the conceptual overlap of the concepts as well as the scenarios where they come apart will, then, also help drive theoretical predictions on whether affecting one psychological process might affect the other and in what direction.

In addition to the contribution this Chapter makes to answering the research questions for this thesis, it has already proven to provide fertile ground for future work—it

prompted an invitation to submit a commentary for *Psychological Bulletin*, discussing a review of the conditions that make impulsivity adaptive (available here). The Chapter is currently under review at a journal.

Impulsivity and Self-Control as Timeless Concepts: A Conceptual Analysis of Intertemporal Choice

Simon T. van Baal, Lukasz Walasek, Antonio Verdejo-García, Jakob Hohwy

Abstract

Impulsivity and self-control are central in theories and empirical research concerning addiction, substance use, psychopathology, organisational behaviour, among many. Despite their importance, researchers disagree about the definition and measurement of these constructs. A common method for measuring impulsivity or self-control is inspired by time preferences. Accordingly, preferences for smaller rewards received sooner, over larger rewards received later, have been linked to impulsivity and self-control. This paper investigates whether impulsivity and self-control are measured by time preference parameters obtained through intertemporal choice paradigms. We first provide a historical overview of research on time preferences tracking the theoretical link between intertemporal choice, impulsivity and self-control. Conceptualising impulsivity and self-control building on various disciplines, we consider whether impulsivity and self-control concern time preferences. Using these insights, we suggest a possible direction for formal models of impulsivity and self-control. Our conceptual analysis reveals that impulsivity concerns a lack of reflection on one's choices, not a lack of concern with the future, and self-control concerns deliberation versus temptation, rather than future-orientedness. People may, and do, use self-control to choose a 'smaller-sooner' reward or impulsively select a 'larger-later' reward. This implies these constructs cannot be measured using the standard intertemporal choice paradigm. We suggest that to study impulsivity and self-control in a temporal context, more information is needed about agents' motivation and deliberation. We propose a future direction for decision-making models with a separate causal role for impulsivity and self-control, providing the basis for a new understanding of how they influence choices across different domains of behaviour.

keywords: impulsivity, self-control, intertemporal choice, time preferences, delay discounting

1.4 Introduction

Many everyday choices have consequences that occur at different timescales. Do I buy a coffee or save my money? Do I cycle to work or take the car? Will I eat pizza or something healthier? Similar dilemmas are faced by organisations and governments, some with the potential to shape nations. Do we lower fuel taxes to combat inflation or reduce dependency on fossil fuels by building green energy capacity and facilitating alternative modes of transportation?

The ubiquity and importance of these choices have resulted in a wealth of research on intertemporal choice, where rewards or costs of options occur at different moments in time. Generally, humans and non-human animals prefer rewards that arrive sooner rather than later, a pattern economists refer to as time preferences. The most prominent economic framework to account for this phenomenon is called delay discounting (Chabris et al., 2010), a method that seeks to explain these time preferences by diminishing the utility of an option as it moves further away into the future by some "discount factor".

Delay discounting was not developed as a model of the human decision-making process for trading money through time. Fisher (1930) suggested that discounting was fundamental: the price of capital was merely the discounted value of its expected future income streams. Thus early delay discounting models (notably, Fisher, 1930; Samuelson, 1937) initially served as a normative account of intertemporal choice to assist with economic modelling for how interest rates come to be. Delay discounting can be described an 'as-if' model, that is, if agents discount outcomes in the future with some function *f(t)*, where t denotes time, then their utility function would look like some function *u(f,X)*, where X are, broadly, sources of enjoyment (positive or negative). Currently delay discounting models are commonly taken to be both a normative and descriptive account of how people make decisions.

Researchers interested in the decision-making process (e.g., in behavioural science and psychiatry) are interested in individual differences, and how they lead to different outcomes. To learn more about people's time preferences and how they arrive at them, behavioural scientists, psychologists, and psychiatrists widely use intertemporal choice tasks. In this line of work, researchers often elicit choices between options that are comprised of rewards at different points in time. The rewards are usually based on

(hypothetical) monetary amounts for humans and food rewards for non-human animals. Researchers then fit people's choices to a delay discounting model and estimate discount factors to quantify individual differences.

Central to the thesis of the current work is that researchers in this space want to make claims about impulsivity and self-control based on time preferences elicited through intertemporal choice tasks, often as quantified by delay discounting parameters. Namely, those who discount the future strongly are often denoted as impulsive, whereas those who do not are often denoted as self-controlled.

In this paper, we challenge the preconception that impulsivity and self-control are implicated in time preferences, putting pressure on the notion that they can be measured in intertemporal choice tasks. To illustrate the basic logic of our proposal, we first provide an example of the suggested involvement of impulsivity and self-control in an intertemporal choice—precisely the sort of scenario that researchers want to describe in their studies. We then provide a historical overview of research on time preferences to track the link between impulsivity, self-control, and time preferences. We conceptualise impulsivity and self-control using definitions and theories from various relevant disciplines, and using these definitions, we evaluate whether impulsivity and self-control are implicated in traditional intertemporal choice tasks. Finally, using these insights, we propose a new direction for formal models of impulsivity and self-control.

Consider the case of Charlie, who is an accountant at a large company and is planning on moving to a new city for a different position. She is looking to buy a new house, and she knows that a colleague, David, wants to sell his house. After showing her the property, David says he would accept an offer from Charlie of $1 million if she moved in in December, but Charlie notes that she might want to move in right away when she arrives to start her new job in October. David tells her that they could arrange for this to happen, but that this would be inconvenient for him because he and his family would need to stay somewhere else and move all their things twice. Therefore, David suggests that he would accept Charlie's offer and let her move in in October if she added $25,000 to the $1 million initially proposed. That is, she can choose to speed up the reward by two months for a 2.5% increase in cost. At this point, her moving date is still a long time away so Charlie says she would be fine with December.

However, a few weeks before she moves to the new city, Charlie reconsiders accepting David's proposal for expediting her move. She thinks it will be a great home to start her new job from: it is within cycling distance to her office; the house was recently renovated, thus needing minimal work; given the size and location, it is also reasonably priced. If she chose to wait, she would have to move twice, and start her new job in temporary housing (which she would also have to spend time looking for). She would like to avoid this because she wants to be in a good headspace for her new job. She decides that it would be best to take the option to move in earlier. In other words, she would prefer to incur the $25k cost to forego these inconveniences. (This is a common pattern in intertemporal choice, broadly consistent with Construal Level Theory, see Trope & Liberman, 2010).

There is nothing inherently impulsive about Charlie's decision to incur additional costs and move in sooner. Equally, it would appear odd to claim that Charlie lacks self-control. Indeed, her decision seems deliberate and based on sound and careful evaluation of pros and cons of each option. Her decision can also be explained with respect to Charlie's goals. Yet, these decisions are analogues of what is studied in intertemporal choice tasks. In this literature, a decision to opt for a smaller-sooner reward (or in this case, larger cost sooner) is typically associated with an individual being impulsive or lacking (or failing to exercise) self-control.

Certainly, given her preference reversal (she prefers the later reward initially, then changes her preference to the sooner reward), her decision-making would be better described by hyperbolic discounting (Ainslie, 1975; Thaler, 1980, 1981) where the subjective value of future rewards rises fast when they approach immediacy, than by the Discounted Utility Model (commonly referred to as exponential discounting; Fisher, 1930; Samuelson, 1937)—generally taken to be the normative intertemporal decision-making model in economics. Hyperbolic discounting is often suggested to be the underlying reason for decisions like Charlie's. It is not entirely clear, though, why exactly this behaviour is impulsive or why her choice indicates her lack of self-control, especially since she deliberated the choice carefully and came to a decision that suited her best judgment.

Are future-oriented decisions necessarily self-controlled? Does the economic future-orientation of Charlie's initial decision preclude it from being an impulsive one? Currently, the dominant theory answers in the affirmative: impulsivity manifests through present bias, or excessive delay discounting, while self-control manifests through future-orientation, and thus little delay discounting.

In the following sections, we will provide an overview of the relationship between intertemporal choice, impulsivity, and self-control in the literature, provide a conceptual framework for the two concepts, and then we will argue that time preferences are largely orthogonal to impulsivity and self-control. We also discuss the broader implications of our argument and propose options for new directions for the study of impulsivity and self-control.

1.5 Impulsivity and Self-Control in the Intertemporal Choice Literature

Here we present a brief overview of how impulsivity and self-control are conceptualised and how they are related to delay discounting measurements in the literature; we provide examples of these occurrences in influential and highly cited articles in the various fields working with intertemporal choice tasks. Table 1 provides examples of representative, influential articles from the literature discussing impulsivity and self-control in the context of time preferences.

1.5.1 Examples in Influential Works

To exemplify, consider the following formulations. A long-time leading researcher in the field, Ainslie (2021) formulated the role of self-control (though he uses 'willpower', seemingly taking it to be equivalent) as follows, "choices that evoke willpower typically compare options that pay off over different time courses, with poorer but faster paying ones weighed against the better but slower paying".

In clinical psychology and neuroscience this type of formulation is also common, (for a prominent example, see "Impulsivity: the neurological and behavioural study of discounting"; Madden & Bickel, 2010). Similarly, Nigg (2017) writes: "Impulsivity: Non-reflective stimulus-driven action when a later-rewarding goal-relevant response was also available... Mediated by both bottom-up processes (e.g. spontaneous reward valuation/discounting) and top-down processes (e.g. biasing from prior goals; response inhibition)".

This view of the constructs is also not uncommon in non-human animal behaviour research; Laude et al. (2012) write: "The steepness of the discounting function can be taken as a measure of the degree to which an animal is characterized as impulsive or, the degree to which it lacks self-control ... That is to say, behavioural measures of discounting are often interpreted as indicating the degree of impulsivity". In other texts, the equivalence between present-oriented and future-oriented intertemporal choice patterns, and impulsivity and self-control are tacitly accepted rather than explicitly asserted (e.g., Doyle, 2012; Stevens & Stephens, 2010).

It is worth examining why impulsivity and self-control have come to be conceptualised in this way. Many authors, including some mentioned in Table 1, either reference Ainslie (1975), or Logue (1988); Logue, in turn, references Ainslie (1975) who is concerned with preference reversals like Charlie's and calls this 'impulsiveness'. Ultimately, this view of time preferences appears to date back to Fisher (1930), who names weakness of will as one of several drivers of future discounting. That line of reasoning is subsequently continued by Strotz (1955), who theorises about why people do things against their best interests.

TABLE 1.1: Sample of representative articles from the intertemporal choice literature discussing impulsivity and self-control.

Authors, Year, Journal	**Title**	**No. Cit.**	**Quote**
Concept: Impulsivity			
Baumann and Odum (2012). *Behavioral Processes*	Impulsivity, risk taking, and timing	166	"Steep delay discounting is also known as impulsive decision making"
Continued on next page			

Table 1.1 – Continued from previous page

Authors, Year, Journal	Title	No. Cit.	Quote
De Wit (2009). *Addiction Biology*	Impulsivity as a determinant and consequence of drug use: a review of underlying processes	1512	"The most commonly used behavioral measures of impulsivity are delay discounting, which assesses impulsive decision-making, and behavioral-inhibition tasks."
Rubia et al. (2009). *Philosophical transactions of the Royal Society of London*	Impulsiveness as a timing disturbance: neurocognitive abnormalities in attention-deficit hyperactivity disorder during temporal processes and normalization with methylphenidate	344	"We argue that impulsiveness is characterized by compromised timing functions such as premature motor timing, decreased tolerance to delays, poor temporal foresight and steeper temporal discounting."
Sharma et al. (2014). *Psychological Bulletin*	Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures	578	"Another aspect of neuropsychologists' definition of impulsivity, termed "choice impulsivity" by some ... is the inability to delay gratification or to choose smaller, immediate rewards over larger, distant ones ..."
Nigg (2017). *Journal of Child Psychology and Psychiatry*	Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology	947	"Impulsivity: Nonreflective stimulus-driven action when a later-rewarding goal-relevant response was also available. May be adaptive or maladaptive depending on context and degree of inflexibility as context changes. Mediated by both bottom-up processes (e.g.spontaneous reward valuation/discounting) and top-down process (e.g. biasing from prior goals; response inhibition)."
Wiers et al. (2010). *Frontiers in Psychology*	Impulsivity, impulsive and reflective processes and the development of alcohol use and misuse in adolescents and young adults	117	"Aspects of impulsivity and sensation seeking can also be assessed with behavioral performance measures such as delay discounting."
Coffey et al. (2003). *Experimental and Clinical Psychopharmacology*	Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals	710	"However, the behavioral economics literature provides a behavioral method of assessing impulsivity by assessing the reduction in the subjective value of delayed rewards as a function of delay interval."

Continued on next page

Table 1.1 – Continued from previous page

Authors, Year, Journal	Title	No. Cit.	Quote
Kirby and Herrnstein (1995). *Psychological Science*	Preference reversals due to myopic discounting of reward	749	"Their delay discount rates do not remain constant; their relative preferences change; they are indifferent between outcomes at some points in time but not at others; and they reverse preference as their temporal vantage point changes. All of these apparently haphazard features of human decision making are consistent with the ordinary definition of impulsiveness, in which a choice is made on the basis of a temporary, and often sudden, change in preference."
Concept: Self-Control			
Rachlin and Green (1972). *Journal of the experimental analysis of behaviour*	Commitment, choice and self-control	1635	"The preference for the large delayed alternative with long durations of T parallels everyday instances of advance commitment to a given course of action. Such commitment may be seen as a prototype for self-control."
Thaler (1981). *Economic Letters*	Some empirical evidence on dynamic inconsistency	2953	"This hypothesis is that the discount rate will vary inversely with the size of the reward for which the individual must wait. This hypothesis is derived from viewing intertemporal choice as problem [sic] in self-control."
Both Concepts			
Stevens and Stephens (2010). (in Madden & Bickel, 2010)	The Adaptive Nature of Impulsivity	120	"We define impulsivity as choosing a smaller-sooner option when a larger later option produces a better outcome."; "In self-control studies, the investigator trains subjects ... to choose between a small reward the subject can obtain quickly ... and a larger reward it must wait a bit longer to obtain."
Logue (1988). *Brain and Behavioral Sciences*	Research on self-control: An integrating framework	1026	"... self-control referring only to the choice of a larger, more delayed reinforcer over a smaller, less delayed reinforcer and impulsiveness referring to the opposite."[1]
Van den Bergh et al. (2008). *Journal of Consumer Research*	Bikinis instigate generalized impatience in intertemporal choice	416	"Visceral factors may drive impatient and impulsive behaviors ... and demonstrate less self-control (defined by a greater preference for smaller, less delayed access to apple juice ...)."

Continued on next page

[1]She does not commit to this without weighing the advantages and disadvantages, but she never touches on any of the substantial disadvantages we will discuss below.

Table 1.1 – Continued from previous page

Authors, Year, Journal	Title	No. Cit.	Quote
Bickel and Marsch (2001). *Addiction*	Toward a behavioral economic understanding of drug dependence: delay discounting processes	1281	"Impulsivity has been defined as the selection of a smaller more immediate reward over a larger more delayed reward (self-control has been defined as the opposite . . .)."

It should be noted that Ainslie (1975) initially writes that the concavity of an agent's discount function (i.e., hyperbolic discounting, which we will discuss in more detail below) is what makes choices impulsive or self-controlled—not their general preference for sooner rewards over later rewards. He acknowledges that there are many situations where a preference for smaller-sooner rewards is adaptive and justified, and thus these situations were not of interest to his project.

According to Ainslie, one is impulsive if by choosing the smaller-sooner reward one reduces their lifetime utility. In this, he focuses on what we now call *preference reversals* (or the common delay effect): an agent prefers larger-later option A over smaller-sooner option B at time t when both options are still far away; as time goes on (and no new information presents itself in the interim), when option B becomes more immediate at time T, the agent prefers B to waiting for option A. The story of Charlie's home purchase is an example of this phenomenon because she reverses her earlier preference, and chooses to speed up her reward (moving into the house) by paying an extra \$25k. Ainslie (1975) suggests that this type of choice is due to curvature in the delay discount function, which was later formalised in the following equation (Mazur, 1987):

$$f(D) = \frac{1}{1 + kD}, \tag{1.1}$$

where f is the discount factor applied to a reward received after delay D and k is the discount parameter. This is in contrast with Samuelson (1937) his discounted utility model function, based on Fisher (1930) theory of interest:

$$f(D) = e^{-kD}, \tag{1.2}$$

where f and D mean the same as above, k is the discount parameter, and e is the base of the natural logarithm. Discounting future rewards through this exponential

function never leads to preference reversals. See Figure 1.1 for a depiction of the different parametric forms of these models.

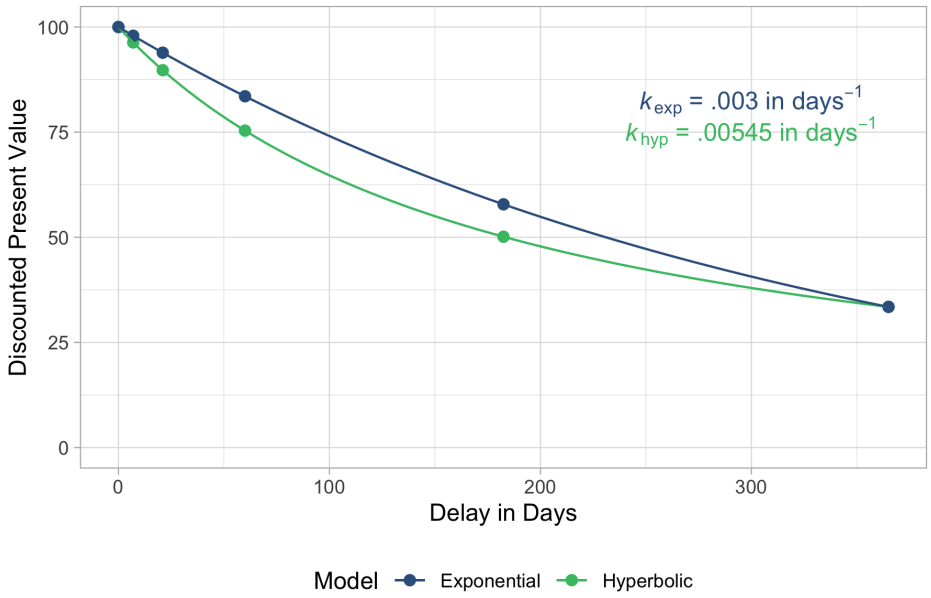

FIGURE 1.1: The exponential discounted utility model (blue) and the hyperbolic discounting model (green). As the reward moves further away, the decision-maker values it less in the present. In the hyperbolic model this devaluation is stronger when the reward is closer to the present and attenuates when rewards are further into the future. For the discounted utility model, the devaluation occurs at a constant rate.

In the clinical psychology and neuroscience literature on 'choice impulsivity' (i.e., delay discounting), however, this more nuanced point that impulsivity and self-control are proposed to be related to the parametric form (i.e., convexity) of the discount function is often lost. What is denoted as 'choice impulsivity' is instead often associated with the amount of discounting (i.e., the value of k). Crucially, in the hyperbolic discounting model, the steepness of the function and the parametric form are indissociable because they are both captured by one parameter. The hyperbolic model therefore has no 'impulsivity' parameter (in Ainslie's view of the construct), but a single time preference + impulsivity parameter. The other common measurement taken is the area under the curve, a mostly theory-neutral measure of the value of future rewards lost to delay discounting (Myerson et al., 2001), which is also uninformative about the exact shape of the convexity of the discount function. This is important because the conflation between steepness and convexity makes it difficult for researchers to pinpoint what they are looking for in an intertemporal choice task when they study impulsivity.

In their review on choice impulsivity, Hamilton et al. (2015) write that choice impulsivity refers to "making impulsive decisions and involves tendencies to select smaller-sooner rewards over larger-later rewards (e.g., the choosing of immediate but smaller versus delayed and larger rewards)", and that it can be measured through indifference points, the discount rate, area under the curve, or even, in a model-free way, just by taking the percentage of larger later choices made, noting that "a higher k reflects greater CI [choice impulsivity]". None of these measures directly deal with the convexity of the discount function. See Figure 1.2 for an illustration of how different rates of exponential discounting would lead to a different amount of convexity if the data were fit with the hyperbolic model.

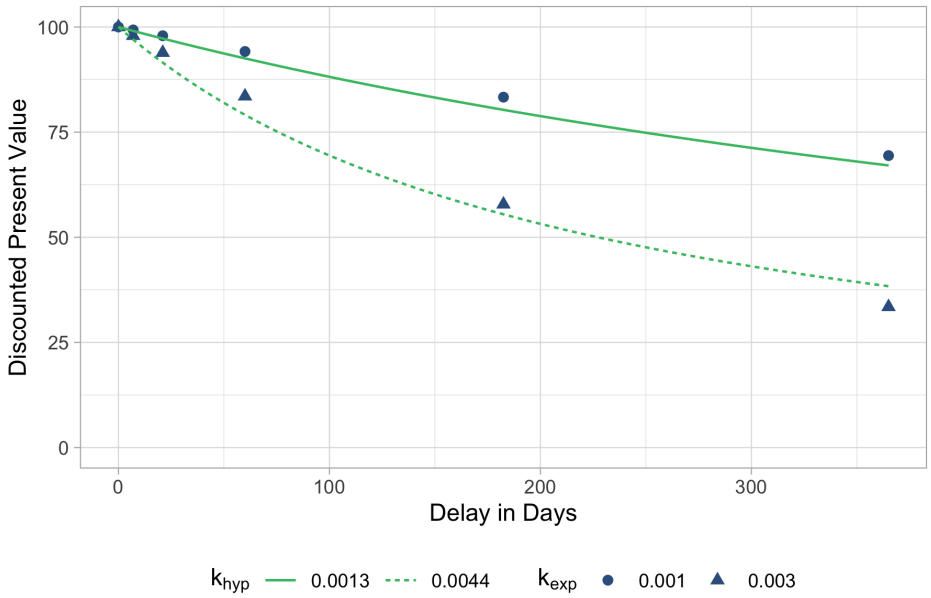

FIGURE 1.2: Hyperbolic model fits (the curved lines) to simulated indifference points generated based on exponential discounting (without noise; the points). The shapes represent different generating parameters, and the difference in line type (solid/dashed) represents the fitted parameter.

Another view, though less commonly explicitly endorsed, is that impulsivity and self-control are antitheses. That is, they are not only both implicated in various ways in intertemporal choice, but they are opposites—if you are high in impulsivity, you are low in self-control and vice versa. Authors endorsing this view explicitly put the concepts on opposite ends of the future-oriented decision-making spectrum. Along these lines, Duckworth and Kern (2011) write "Several authors have noted the challenge of defining and measuring self-control . . . and its *converse*, impulsivity or impulsiveness" (emphasis added).

Thus, considering there are differences in what researchers take to be impulsive or self-controlled in intertemporal choice experiments, our current work will address the various possible relationships between the constructs and intertemporal choice.

1.5.2 The Need for Conceptual Consensus

There have been many discussions about the relationships between intertemporal choice, impulsivity, and self-control, and some of the arguments and conceptual analysis we present may on reflection seem obvious to some readers. Indeed, Bulley et al. (2021) and Loewenstein (2018) take an important step in arguing, much as we will, that self-control does not equate to present bias. Strickland and Johnson (2020) have suggested that impulsivity should be rejected as a psychological construct, and thus that it cannot be measured using intertemporal choice tasks. We are, however, currently unaware of researchers arguing that time preferences, and impulsivity and self-control are largely orthogonal. Therefore, we intend to go further and examine the possible links impulsivity and self-control might have to time preferences. Given the proliferation of research in this area, we believe a comprehensive discussion is warranted because conceptual clarity on this topic will define boundaries and future directions for scientific inquiry (Kaplan, 1964).

One of the problems with the lack of conceptual consensus is that it could lead to conclusions that do not translate to other domains, or only apply to impulsivity and self-control in limited circumstances. For example, Bickel and Marsch (2001) note in one of their papers: "By identifying such behavioural processes [that result in impulsivity and loss of control] delay discounting may also suggest potential interventions for modifying impulsivity and self-control failure". If delay discounting is not an important part of impulsivity and self-control, their claim could be challenged because we would not be able to modify impulsivity (we will discuss other potential roles for time preferences in lapses of self-control later).

Another issue with tying impulsivity to short-term rewards and self-control to long-term rewards is that few studies examine people who are troubled by frequent impulsive future-oriented behaviours (such as in Kivetz & Keinan, 2006) or a lack of self-control in attaining short-term rewards. These patterns could play a role in anorexia nervosa, compulsive behaviour patterns such as attentional capture and cognitive hijacking (Muela

et al., 2022), and workaholism. It has been argued that in some domains many people are too future-oriented (Loewenstein, 2018). Workaholism is a good example of this because workaholics frequently put work, an activity that predominantly rewards them in the future, before any immediate rewards. In doing so, they often harm their connections with family, community, and their mental health. For workaholics, it can take considerable self-control to tear themselves away from their work (more on this in the next section) and do something that is more immediately rewarding.

Finally, tying this back to the example we started with, the preconception that binds time preferences to impulsivity and self-control costs us the ability to call Charlie's earlier decision to wait for the house impulsive, and prescribes that we call that decision self-controlled. Yet from the details of that situation, we saw that this view is unsatisfactory.

To address these issues, we will discuss the conceptual framework for impulsivity and self-control to be used for the rest of the paper in the next section.

1.6 Conceptualising Impulsivity and Self-Control

Impulsivity and self-control both enjoy longstanding philosophical inquiries (for some historical insight, see Madden & Bickel, 2010; Stroud, 2021). The constructs were already discussed among the ancient Greeks. The Aristotelian term akrasia (loosely translatable to 'weakness of will'), comes in two forms: *astheneia*, where one has deliberated and decided on the best course of action, but then succumbed to a passion (i.e., a self-control lapse), and *propeteia*, where the same result occurs but one never deliberated before the act (i.e., impulsivity) (Stroud, 2021). However, it is not until recently that research on impulsivity and self-control gained momentum.

The phenomenon of 'concept-creep' (Haslam, 2016) refers to the application of constructs to tangential phenomena, causing the understanding of those constructs to broaden, as has been observed in the psychology and neuroscience literature. This seemingly applies to impulsivity and self-control, but a key difference is that research in impulsivity and self-control has historically been limited by a lack of consistency in their conceptualisation, leading to difficulty disentangling exactly how the concepts are expanding. To that effect, De Wit (2009) wrote that different definitions of impulsivity incorporate "seemingly unrelated maladaptive behaviours including, for example, inability to wait,

difficulty withholding responses and insensitivity to negative or delayed consequences". Since then, we have made little progress on the front of conceptual clarity. It appears that it would be prudent to develop our understanding of what impulsivity and self-control are.

1.6.1 A Conceptual Framework for Impulsivity

Currently, the most used meanings of impulsivity in the literature are derived from various influential works published around the change of the millennium (Evenden, 1999; Moeller et al., 2001; Whiteside & Lynam, 2001). Impulsivity is widely considered to be multidimensional. According to Whiteside and Lynam (2001), impulsivity consists of five dimensions: sensation-seeking, negative urgency, lack of perseverance, lack of premeditation, and positive urgency. Some dimensions can be loosely mapped onto other conceptualisations, for example, one where impulsivity consists of motor impulsivity (response inhibition), cognitive impulsivity (making quick cognitive decisions), and choice impulsivity (delay discounting) (e.g., Caswell et al., 2015; Dick et al., 2010; Green & Myerson, 2013; Patton et al., 1995; Vassileva & Conrod, 2019).

There are also widely differing definitions of impulsivity. In Stedman (1995): "Impulsive: relating to or activated by an impulse rather than controlled by reason or careful deliberation". Evenden (1999) writes "Impulsivity refers to behaviours or acts that are unduly hasty, risky, and inappropriate, leading to negative outcomes", and this conceptualisation is still commonly used. Given the diverging schools of thinking on impulsivity, it is important to settle on a conceptualisation that captures the key components of the construct to facilitate clear discussion. We will conceptualise impulsivity as:

- A predisposition toward unplanned reactions to stimuli.
- A disregard for the consequences of these reactions.

This conceptualisation largely follows Moeller et al. (2001), which is one of the most commonly used definitions and reflects the complexity of impulsivity (Stanford et al., 2009). Since this conceptualisation is widely used, we use it here as our target (and will discuss below if another conceptualisation is in fact more apt).

1.6.2 A Conceptual Framework for Self-Control

There is also some disagreement about the proper conceptualisation of self-control. An important factor for this disagreement is that some researchers think of a self-control conflict as tension between short-term temptations and long-term goals (see e.g., Duckworth et al., 2018; Milkman et al., 2008).

Common conceptualisations are that self-control entails a want/should conflict (Milkman et al., 2008), or a conflict between affect and deliberation where one has emotional desires, drives and motivational feeling states (such as pain) on the one hand, and deliberation on the other hand (Loewenstein, 2018; Loewenstein et al., 2015). These views correspond with elements of our common-sense (pre-theoretical) concept of self-control but are not without their disagreements. The want/should conflict is unduly cast in a short-term versus long-term conflict, and even when abstracting away from that issue, it appears that the deliberative option in such a self-control conflict might not always entail a 'should'. The deliberative option in a self-control conflict often also represents or stems from, a 'want' (unless wants are taken to be equivalent to affect).

Baumeister and Alghamdi (2015) are not far off this type of view. They write "self-control is what enables people to override impulses and responses so as to do something else, especially something that is more highly valued". Self-control could otherwise be understood as the opposite of *weakness of will* (an over-readiness to revise a resolution; Holton, 2009), or as *enkrateia* – the opposite of *akrasia* (Kraut, 2001; Mele, 2010, 2012), where an agent fails to act in accordance with reason.

For the present discussion, we would like to remain neutral on the various theories of self-control because they seem conceptually closer than those of impulsivity. We will broadly conceptualise self-control conflicts as those that concern deliberation versus temptation, that is, situations where the agent has a view on the best course of action but has conflicting desires. We would like self-control to entail both what would often be called willpower; adopting mental strategies to avoid giving into temptation, but also *precommitment strategies*; changing the decision-making environment such that one does not need to rely on *willpower* (e.g., not buying any snacks to take home)—which is often judged as more effective than willpower (Ariely & Loewenstein, 2006; Baumeister & Vohs, 2003).

As with our conceptualisation of impulsivity, we will remain neutral initially on whether this is the 'correct' way to conceptualise self-control, but we emphasise that we take self-control not to involve any temporal aspects.

1.6.3 The Relationship Between Impulsivity and Self-Control

Based on the above conceptualisations and definitions, it appears that with any combination of these conceptualisations of either construct, impulsivity cannot be the converse of self-control. Impulsivity mainly consists of acting with a disregard to consequences or lack of deliberation, or premeditation, and crucially does not involve an imbalance of 'want' over 'should', nor of temptation over judgment. It is possible, for instance, to use self-control to be impulsive: "I should do more things that I would enjoy doing without thinking about all the consequences it might have; I should live a little!". Or, for the alternative understandings of self-control we may *think it best* to enjoy life and discard the weight of society's expectations rather than think about consequences all the time. If we then act impulsively, we have clearly not experienced a lapse in self-control[2].

1.7 Necessary and Sufficient Conditions for Impulsivity and Self-Control

To understand the roles of impulsivity and self-control in intertemporal choice, we must evaluate whether intertemporal choice patterns fulfil the necessary and sufficient conditions of impulsivity and self-control. In other words, we must evaluate whether time preferences map onto impulsivity and self-control in a meaningful way, to see if we can measure individual differences in impulsivity and self-control using intertemporal choice tasks.

[2]If one is committed to equating self-control with Holton's weakness of will, one needs to set boundaries to the deliberation limits for impulsivity here. The agent needs to deliberate beforehand to make a resolution, so if we take the lack of deliberation required for impulsivity literally, anything that follows can no longer be impulsive. This would be undesirable, we think. Even if this point is not conceded, it is not crucial for the rest of the work.

1.7.1 Necessary and Sufficient Conditions of Impulsivity

Steep delay discounting, or temporal myopia (short-sightedness), is implicated in decision-making patterns that are commonly associated with impulsivity, such as substance use and addictive disorders (Verdejo-García et al., 2008). People with substance use disorders, for example, are regarded as impulsive in their inability to suppress the urge to consume their drug of choice (i.e., succumbing to the smaller-sooner reward). However, this correlation does not mean that steep delay discounting is necessary or sufficient for an individual to be impulsive. It is just as likely that impulsivity and time preferences contribute separately to addiction (De Wit, 2009).

One example of this would be a person who is prompted by their employer to pick a pension scheme. They are prompted with a multiple-choice question on the proportion of their salary they would like to put into their pension fund. Courtesy of some behavioural economists working for the company, the default option is the one with the highest proportion of income that goes to their pension. Without thinking too much about their budget constraints, the person proceeds with the default option (Jachimowicz et al., 2019). Here, their choice of a future-oriented option was impulsive. This shows that making present-oriented decisions, or excessively discounting time delays of rewards, is not *necessary* for someone to be impulsive.

Next, to see whether excessive delay discounting or preference reversals are *sufficient* for impulsivity, we will use the example we started this paper with. Charlie had clearly thought through her decision to change her earlier decision to choose the smaller-sooner reward, at a cost of $25,000. Her preference reversal should tell us her discounting fits the hyperbolic model better than the exponential model, and the amount she pays for speeding up her move indicates steep discounting. Yet to say she was impulsive in her later decision to move in at the earlier date would be inaccurate because there was no lack of deliberation or hasty reaction associated with her choice. Therefore, strong delay discounting is not *sufficient* for impulsivity.

Indeed, would we not say her initial decision to wait until December was impulsive? She may not have thought through the consequences concerning her wellbeing in her decision to move in later. She would be starting a demanding new job, but she may not have considered how moving twice in that period would affect her performance and

learning. Does her choosing the larger-later reward preclude her initial decision from being impulsive? We think not.

Therefore, even though impulsive behaviours may correlate with delay discounting in some cases, that does not mean that delay discounting reflects impulsivity.

1.7.2 Necessary and Sufficient Conditions of Self-Control

In the intertemporal choice literature, a tendency to choose larger-later options is often denoted as a self-controlled choice pattern. A plausible explanation of this time-bound conceptualisation of self-control is that choosing the larger-later option connects with some aspect of self-control, namely delay of gratification (such as in the marshmallow test; Mischel et al., 1989). However, the temporal aspect of these delay of gratification experiments is not the part that indicates self-control in the subject.

If the most famous version of the marshmallow test experiment is changed marginally, such that the first reward is an apple, while the second reward that requires waiting ten minutes is two marshmallows, it is easy to see that choosing the apple (i.e., the smaller-sooner option, hardly as rewarding for most children as two marshmallows) could be the reward that indicates self-control. Children are generally aware that the apple is something they should eat, whereas the marshmallows are something they want to eat—even if they must wait for them (as evidenced by the findings of Mischel and colleagues). This indicates that waiting for a larger-later reward is not *necessary* for self-control; one can be self-controlled without choosing the larger-later option.

To establish whether shallow delay discounting (i.e., the discounting curve has a gentle gradient) is *sufficient* for self-control, consider an extremely future-oriented (i.e., hyperopic) agent, who wishes they made more present-oriented decisions because they want to enjoy themselves more in the present moment (Kivetz & Keinan, 2006). This sophisticated hyperopic agent may judge it best to enjoy themselves today and go to dinner with their friend. They therefore intend to not do any penny pinching; to just be in the moment. When the waiter comes, however, they cannot help but succumb to temptation, even though they realise their goal was to forget about these considerations. As a result, they choose the cheapest wine on the menu—the others are so expensive!

Such hyperopic behaviour would be indicative of a lack of self-control because the tempting option wins out over the deliberative option. Yet most intertemporal choice research would have us conclude that this person was showing self-control through their future-oriented decision of choosing to save money for later. This story shows that the contrary is true: shallow delay discounting, or a lack of curvature in the delay discounting function is not *sufficient* for self-control because it was a lapse of self-control that led the agent to this future-oriented choice.

Finally, taking this reasoning one step further, we posit that eliciting preferences via an intertemporal choice task without knowledge of contextual factors is incompatible with the study of self-control. A preference, in the economic sense relevant to intertemporal choice, is a combination of wants, shoulds, and other social and contextual factors (i.e., there is no way of telling whether there is a want/should conflict, or a discrepancy between deliberation and temptation). If a participant chooses between $80 now and $100 in a month in an intertemporal choice task, it is not necessary to exert willpower or employ precommitment devices to limit their ability to claim a tempting immediate $80 in the meantime, so their choice brings us no closer to knowing whether they have self-control or not.

Thus, unless we can prove that a participant feels that they *should* pick the $100, but *want* the $80 today, and ends up choosing the latter, it is difficult to argue that they lack self-control if they choose the smaller-sooner option. Therefore, time preferences on their own cannot be seen as self-controlled or not self-controlled.

Considering that time preferences do not satisfy the necessary and sufficient conditions of impulsivity and self-control, we can conclude that the constructs are dissociable. In future writing about time preferences (e.g., in intertemporal choice research), we recommend using specific descriptors about the patterns we observe instead, such as present bias, degrees of future-orientation, and apparent uses of heuristics.

1.8 What Can Be Learnt About Impulsivity and Self-Control Through Intertemporal Choice?

If we accept that time preferences are neither necessary nor sufficient for impulsivity and self-control, or even that the study of these constructs with a standard intertemporal choice task is uninformative, then we should critically evaluate what we are learning about impulsivity and self-control through intertemporal choice tasks. In this section, we discuss the possible ways that intertemporal choice might be informative on impulsivity and self-control, as well as lay the foundation for a potential new model that could aid further work on the relationship between time preferences, impulsivity, and self-control.

It is possible that even though delay discounting does not meaningfully represent impulsivity or self-control, it may be used as an instrumental variable in cases where the validity of self-report responses on clinical questionnaires is especially tenuous (e.g., for individuals with memory deficits). For this to work, impulsivity and self-control measures need to correlate strongly with delay discounting parameters.

However, research correlating impulsivity with delay discounting has produced mixed results (for extensive discussion, see e.g., Odum, 2011). Some studies have found no correlation, or even negative ones (e.g., Lane et al., 2003; Mitchell, 1999; Reynolds et al., 2006; White et al., 1994). Other studies did report a positive correlation of varying magnitudes (e.g., Cherek et al., 1997; De Wit et al., 2007; Kirby et al., 1999). More generally, Cyders and Coskunpinar (2011) conducted a meta-analysis of 27 studies comparing self-report and behavioural measures of impulsivity and failed to demonstrate a strong relationship between 'delay response' tasks, of which intertemporal choice is one, and dimensions of impulsivity, even before correcting for publication bias (for Sensation Seeking: r = 0.031–0.094).

Evidence of convergent validity (strong correlations between different instruments designed to assess a common construct) is the minimum requirement for the validity of any psychological test (Fiske, 1971, p.164). Although one could also attribute this lack of convergent validity to poor reliability or validity of measurement, or method invariance (see e.g., Enkavi et al., 2019). Additionally, behavioural measures are generally designed to maximise within-subject variability rather than between-subject variability, and tend to measure granular state information rather than acculumated trait information (but

see, Odum, 2011; Odum & Baumann, 2010). Regardless of which interpretation is correct, it appears unlikely that delay discounting measures are directly measuring impulsivity, and further, that they might not be good instrumental variables.

A potential reason for this lack of correlation between time preferences, impulsivity and self-control is that the effect of hyperbolic delay discounting on real-life behaviour is modulated by whether one is naïve or sophisticated about future self-control problems (O'Donoghue & Rabin, 1998) (further discussion in O'Donoghue & Rabin, 2015). That is, if one realises that their preferences will revert when they approach a tempting reward, they can employ strategies to reduce their chance of succumbing to temptation (e.g., by precommitment strategies, or by removing the temptation). This illustrates that an important aspect of impulsivity and self-control is taken away in intertemporal choice experiments: there is no temptation to go back on the initial decision—there is no deliberation-temptation conflict. In other words, we find out whether Ulysses decides to tie himself to the mast, but we attain no information on the journey past the sirens.

Moreover, Gabaix and Laibson (2017), among others (e.g., Farmer & Geanakoplos, 2009), suggest that hyperbolic discounting can be economically rational. They argue that this is the case if simulation of future rewards has noise that increases linearly with time, which is further evidence that context is important to know how time preferences will manifest in people's lives. For some, it may be adaptive to discount hyperbolically because they do not trust opportunities to materialise (Jachimowicz et al., 2017). This all is to say there may not be a disconnect between hyperbolic discounting and agents' evaluative judgment, and more strongly, nor should there be.

Another possible reason for the lack of evidence on convergent validity is that the influence of time preference on impulsive behaviours and self-control varies depending on the domain. That is, time preferences are likely to be important for whether and to what extent people deliberate the consequences of their actions, and how tempted they are to choose wants over shoulds or temptations over better judgment, but the direction of the influence likely depends on what the stimuli, rewards (Hursh & Schwartz, 2022; Weatherly & Terrell, 2010), shoulds and wants are. If we again consider our hyperopic agent: their time preferences play a role in causing them to make (potentially impulsive) future-oriented decisions, whereas, for someone with a substance use disorder, their time preferences contribute to high valuation of rewards that are immediately available. Time

preferences may inhibit both types of agents' ability to consider the consequences at the time or to act in accordance with their better judgment. Therefore, both future-oriented time preferences and present-oriented time preferences can cause impulsivity and (lapses of) self-control.

In all, intertemporal choice research appears to be an ineffective way to investigate impulsivity and self-control. This is partly because of a lack of consideration of the broader context in typical experiments. Yet this need not deter us from researching the impact of time preferences on impulsive behaviour and lapses in self-control, since time preferences may have a different role in the manifestation of these behaviours that cannot be captured by traditional intertemporal choice research.

1.9 Alternative Roles for Time Preferences in Impulsive Behaviour and Lapses in Self-Control

The main point of this paper was to put pressure on the notion that impulsivity and self-control are captured by current measures of time preferences. Addressing this issue naturally raises the question of what a possible positive account of the connection between time preferences, impulsivity, and self-control should look like.

A critic of our perspective might point out that excessive discounting of delayed rewards is a core feature of many phenomena commonly referred to as *impulsive behaviours* or *lapses in self-control* in certain domains, such as substance use disorder (Bickel & Johnson, 2003; Madden et al., 1997), gambling disorder (Ioannidis et al., 2019), and borderline personality disorder (Barker et al., 2015). This could lead one to conclude that time preferences are, in some way, important for impulsivity and self-control.

We do not deny that time preferences are important for impulsivity and self-control, and this is not incompatible with what we have argued. As alluded to before, the most straightforward way to interpret this relationship is that time preferences play a *causal role* in some impulsive behaviours and lapses in self-control. In this section, we discuss some of the candidate approaches to this type of view.

1.9.1 Candidate Approaches for the Causal View

Given the history of the intertemporal choice literature, it would be natural to assume that hyperbolic discounting plays this causal role for the occurrence of the kind of impulsive behaviours and lapses in self-control we see in behaviour patterns that are triggered by immediate rewards (as in addiction). Perhaps by increasing the relative value of immediate rewards so much that it precludes agents from considering future consequences. However, there is the question of whether hyperbolic discounting is an accurate descriptive account of human decision-making. Read (2001) has argued, using a food-based example, that the predictions of the hyperbolic discounting model do not agree with real-world observations that preference reversals usually occur when the reward is immediate (in the standard hyperbolic discounting model this reversal can occur weeks before the smaller-sooner becomes available). This is the reason why, he argues, the most interesting prediction of hyperbolic discounting, that people display *decreasing impatience*, is incorrect.

Several models have been developed to improve on the hyperbolic model by incorporating psychologically plausible parameters. Some of the most commonly used ones belong to a class called "hyperboloid" models (for a comprehensive list, see He et al., 2022). These usually add a time sensitivity parameter, usually denoted as s. However, the functions that are based on the hyperbolic model suffer from the same parameter interpretability problems mentioned before (see Figure 1.2) because the values parameter s needs to take are psychologically implausible and because it is ill equipped to decrease the convexity of these functions. The Myerson-Green model (Myerson & Green, 1995) and the Modified Rachlin model (Rachlin, 2006)—the original suffers from muddled units so using that one to estimate discount factors is ill-advised (Vincent & Stewart, 2020)— are examples of this.

An alternative explanation for these preference reversals when in proximity of immediate rewards can be accounted for by quasi-hyperbolic discounting, where discounting is composed of two parameters (Laibson, 1997):

$$f(0) = 1 \tag{1.3}$$

$$f(D) = \beta * \delta^D, \tag{1.4}$$

where β is present bias, δ is the discount rate, and D is again the delay. Present bias, which predicts preference reversals when the smaller-sooner option is immediate, could be counted as the cause of these preference reversals.

Another option is the constant sensitivity function (Ebert & Prelec, 2007):

$$f(D) = exp(-(aD)^b), \tag{1.5}$$

where a is analogous to k in Equation Equation 2.1, D is again delay and b is the time sensitivity parameter. This model is attractive because it can achieve a similar convex shape as the hyperbolic model ($b < 1$) but if $b = 1$, the equation resolves to the exponential discount function, Equation 2.1. When b nears zero, the model's predictions start to functionally resemble Equation 1.4, allowing for immediacy effects.

Equating choice impulsivity or self-control with delay discounting parameters, or parameters of models more generally carries risk, however: the intertemporal choice field has not made up its mind yet about the best descriptive theory of intertemporal choice (it does not look promising for delay discounting in general, and hyperbolic discounting in particular; Ericson et al., 2015; He et al., 2019). There have been many reports of anomalies and biases in intertemporal choice that cannot be accounted for by traditional delay discounting models, including findings of subadditive discounting (Read, 2001), query-order effects (Loewenstein & Thaler, 1989; Weber et al., 2007), date/delay effects (Read et al., 2005), similarity effects (Rubinstein, 2003), and delay/speedup asymmetry (Loewenstein, 1988), suggesting that intertemporal choice is not well described by either exponential or hyperbolic delay discounting models. Alternatives for a descriptive account of intertemporal choice are cognitive process models (e.g., Konstantinidis et al., 2020).

If it turns out that delay discounting is not the right way to describe what humans do because it is not the mechanism through which they 'solve' intertemporal choices (which is likely, given that delay discounting models were not developed as cognitive process models), then we might need to seriously reconsider claims made about human cognition and behaviour based on delay discounting parameters (including area under the curve)—especially in a clinical context. It might be, for instance, that what most people actually do can be captured by an attribute-based additive model (Ericson et al., 2015); the fields applying these notions will then need to figure out what it is about

the application of this model (e.g., the weights allocated to attributes) that is indicative of addictions and other disorders. Alternatively, limited inferences could be made by relying on a limited set of standard intertemporal choice tasks and reporting the model-free outcomes, such as response times and the proportion of larger-later options chosen over the smaller-sooner options chosen.

It appears, therefore, that a positive view might be more robust if we abstract away from delay discounting and talk about time preferences more generally in a model-free way. In other words, a satisfactory positive account for the causal role of time preferences in impulsivity and self-control should ideally be theory-neutral about how time preferences are formed. However, currently there are no theories we know of that satisfy our criteria.

1.9.2 Future Directions

In this final section, we briefly canvass a possible avenue for developing a positive account of the relationship between time preferences, impulsivity, and self-control. In this type of account, time preferences play a causal role in impulsive behaviour and self-control, but the influence exerted by time preferences depends on the agent as well as the context. Specifically, the valence of the effect of time preferences depends on the domain of the reward, the agent's goals, beliefs, intentions, and temptations.

For example, assume we know that an agent is predisposed to implement policies to get reward X but would, after some deliberation about consequences and consideration of their higher-order preferences (i.e., preferences about themselves, preferences about their preferences, for example: "I wish I were more spontaneous, what option fits best with that?") choose to implement a policy to get Y instead. In this case, the impulsive action is X. Therefore, if X is a future reward (e.g., a comfortable retirement), then more future-oriented time preferences would partially cause the agent to be more impulsive in this context.

Likewise, suppose A is a tempting future reward (e.g., a vacation on an island in the Pacific over the winter holidays), but the agent judges that, all things considered, it is better to attain B right now (e.g., purchase an e-bike). Here, B is the self-controlled option and present-oriented time preferences would partially cause a self-controlled action.

Time preferences are, in this account, an input to the subjective value of rewards, but their influence varies across different modalities (Cubitt et al., 2018). Time preferences change the subjective value of an outcome, which influences the likelihood that that option or course of action is chosen.

Impulsivity and self-control then moderate the influence that reward valuation has on our decisions and choices. This would mean that being in an impulsive state strengthens the relationship between reward valuation and behaviour (for evidence of this, see van Baal et al., 2022), whereas exerting self-control weakens the relationship between reward valuation and behaviour. One can, for instance, use self-control to resist succumbing to a highly valued reward (temptation), thereby disrupting the relationship between reward valuation and behaviour. However, an agent's willingness to exert self-control is also affected by the subjective value of the reward, as is the agent's propensity to act impulsively. This type of causal account contrasts with the conventional conception, which is unable to distinguish between the influence of time preferences from the influence of impulsivity and self-control on the outcome. See Figure 1.3 for a sketch of such a model.

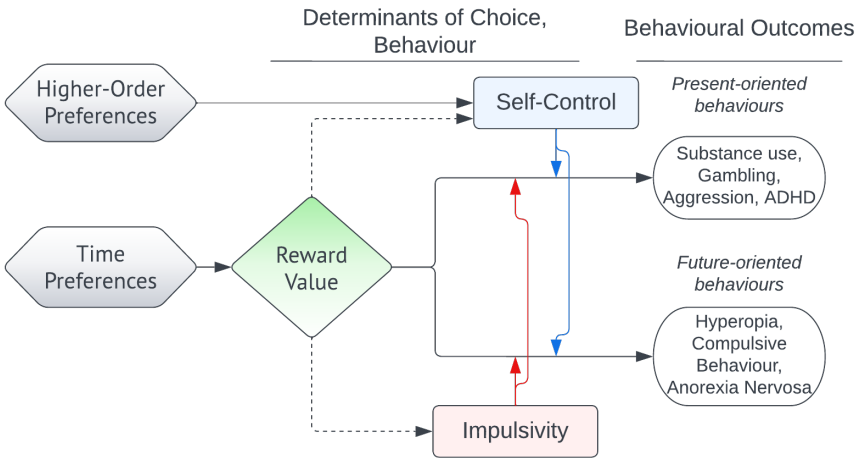

FIGURE 1.3: Sketch of a causal model for the influence of time preferences on behaviour, depicting time preferences as an input for reward valuation. Impulsivity and self-control can moderate the influence of reward valuation on behaviour. Reward valuation also feeds into impulsivity and self-control. The red arrows signify a strengthening of the moderated relationship, and the blue arrows signify an attenuation.

Whether the effects time preferences have on an agent's life are positive depends on whether their preferences within a certain domain are aligned or misaligned with their

higher-order preferences. Excessive future-orientation can be harmful in some eating disorders but stimulating more future-oriented preferences could be beneficial for someone battling gambling disorder.

Thus, we suggest that the role for time preferences in impulsive behaviour and lapses in self-control may be causal, and we have briefly sketched how one may begin to work up such a model. We think a satisfactory positive account is one where the valence of the influence of time preferences on impulsivity and self-control depends on the agent and the domain. Nonetheless, because there is no consensus on the best descriptive model of human intertemporal choice, we should be careful about describing exactly what role is being played and what aspect of time preferences is of interest.

Although our conceptual model is informal, it is possible and perhaps useful to conceive of several testable predictions. Models that regard time preferences as manifestations of impulsivity and self-control are unable to distinguish between the influence these three concepts have on behaviour, unlike the model we propose here. First, the model we propose suggests that if an individual is in an impulsive state, the drive to engage in a behaviour will be strengthened, but their capacity for self-control ought not be directly weakened. Second, as the model is currently conceived, an impulsive state would not necessarily predict a difference in the reward valuation process, only in the intensity of the drive to engage in a behaviour. Third, the use of self-control (successful or not) will be partially predictable if the reward is known to conflict with the higher-order preferences of the individual. And fourth, because self-control is separate from time preferences, the predictive value of the difference between higher-order preferences and the reward will be independent of whether the reward is a future-oriented one (such as saving) or a present-oriented one (such as having an alcoholic beverage). Several other predictions that go against conventional models of impulsivity and self-control are possible, but we limit ourselves to these for the sake of brevity.

1.10 Conclusion

We discussed how impulsivity and self-control are conceptualised in the intertemporal choice literature, and how they are used to denote choice patterns in intertemporal choice and thus time preferences. We considered whether various time preferences satisfy the

necessary and sufficient conditions for impulsivity and self-control, and whether they may play a different role in impulsive and self-controlled behaviours.

We suggest that common ways of conceptualising impulsivity and self-control would require us to conclude that time preferences, impulsivity, and self-control contribute independently to real-world choices between options involving different timescales. Resulting from this observation, we conclude that impulsivity and self-control are unlikely to play an important role in laboratory intertemporal choice tasks. We also discuss whether we should attach any cognitive labels to delay discounting parameters because delay discounting models were not designed to capture cognitive processes in intertemporal choice.

We conclude that we should, therefore, a) dissociate time preferences from impulsivity and self-control, and thus b) not assume that we measure impulsivity and self-control through delay discounting parameters, and thus also not c) assume that people who tend to be impulsive (self-controlled) tend to prefer present-oriented (future-oriented) options.

We suggest that time preferences play a separate causal role from impulsivity and self-control in manifestations of impulsive behaviours and lapses in self-control through changing the subjective value of various rewards. We speculate that the direction of this influence depends on the stimulus and context. This would mean that, for example, future-oriented time preferences could cause more impulsive behaviour in certain domains. This type of causal model strikes us as a useful future research target, which is however difficult to discern as long as impulsivity and self-control remain tied too closely with time preferences.

Our conceptualisation puts pressure on preconceptions about the connections between intertemporal choice, impulsivity, and self-control. We are hoping that this invites scrutiny, where this scrutiny leads us to a clearer understanding of the role of time preferences in impulsivity and self-control.

1.11 Declarations

1.11.1 Data Availability

The code for the simulations used in this manuscript is publicly available here.

1.11.2 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1.11.3 Acknowledgments

This research was supported by the Monash-Warwick Alliance Accelerator Fund. We would like to thank Emmanouil Konstantinidis and Toby Handfield for their comments on an earlier version of this paper.

1.12 References

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

Ainslie, G. (2021). Willpower with and without effort. *Behavioral and Brain Sciences*, *44*.

Ameriks, K., & Clarke, D. M. (2000). *Aristotle: Nicomachean Ethics*. Cambridge University Press. Retrieved December 13, 2022, from https://books.google.co.uk/books?hl=en&lr=&id=p1NmFDSlMfgC&oi=fnd&pg=PR6&dq=nicomachean+ethics&ots=mXq23hDZbs&sig=t8AxhSqAzqHJI7QEuqSJ1HWg8fA&redir_esc=y#v=onepage&q=nicomachean%20ethics&f=false

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*(2), 87–98. https://doi.org/10.1002/bdm.501

Association, A. P. (2010). *Diagnostic and statistical manual of mental disorders, text revision (DSM-IV-TR®)*. American Psychiatric Publishing.

Bakhshani, N.-M. (2014). Impulsivity: A Predisposition Toward Risky Behaviors. *International Journal of High Risk Behaviors & Addiction*, *3*(2), e20428. https://doi.org/10.5812/ijhrba.20428

Barker, V., Romaniuk, L., Cardinal, R., Pope, M., Nicol, K., & Hall, J. (2015). Impulsivity in borderline personality disorder. *Psychological medicine*, *45*(9), 1955–1964. https://doi.org/10.1017/S0033291714003079

Baumann, A. A., & Odum, A. L. (2012). Impulsivity, risk taking, and timing. *Behavioural Processes*, *90*(3), 408–414. https://doi.org/10.1016/j.beproc.2012.04.005

Baumeister, R. F., & Alghamdi, N. G. (2015). Role of self-control failure in immoral and unethical actions. *Current Opinion in Psychology*, *6*, 66–69. https://doi.org/10.1016/j.copsyc.2015.04.001

Baumeister, R. F., & Vohs, K. D. (2003). Willpower, choice, and self-control.

Bickel, W. K., & Johnson, M. W. (2003). Delay discounting: A fundamental behavioral process of drug dependence. In *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 419–440). Russell Sage Foundation.

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction*, *96*(1), 73–86.

Bulley, A., Lempert, K., Conwell, C., Irish, M., & Schacter, D. L. (2021). Intertemporal choice reflects value comparison rather than self-control: Insights from confidence judgments. https://doi.org/10.31234/osf.io/w5zuk

Caswell, A. J., Bond, R., Duka, T., & Morgan, M. J. (2015). Further evidence of the heterogeneous nature of impulsivity. *Personality and Individual Differences*, *76*, 68–74. https://doi.org/10.1016/j.paid.2014.11.059

Chabris, C. F., Laibson, D. I., & Schuldt, J. P. (2010). Intertemporal choice. In *Behavioural and Experimental Economics* (pp. 168–177). Springer.

Cherek, D. R., Moeller, F. G., Dougherty, D. M., & Rhoades, H. (1997). Studies of violent and nonviolent male parolees: II. Laboratory and psychometric measurements of impulsivity. *Biological Psychiatry*, *41*(5), 523–529. https://doi.org/10.1016/S0006-3223(96)00426-X

Coffey, S. F., Gudleski, G. D., Saladin, M. E., & Brady, K. T. (2003). Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals. *Experimental and clinical psychopharmacology*, *11*(1), 18. https://doi.org/10.1037/1064-1297.11.1.18

Cubitt, R., McDonald, R., & Read, D. (2018). Time Matters Less When Outcomes Differ: Unimodal vs. Cross-Modal Comparisons in Intertemporal Choice. *Management Science*, *64*(2), 873–887. https://doi.org/10.1287/mnsc.2016.2613

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical psychology review*, *31*(6), 965–982. https://doi.org/10.1016/j.cpr.2011.06.001

De Wit, H. (2009). Impulsivity as a determinant and consequence of drug use: A review of underlying processes. *Addiction biology*, *14*(1), 22–31. https://doi.org/10.1111/j.1369-1600.2008.00129.x

De Wit, H., Flory, J. D., Acheson, A., McCloskey, M., & Manuck, S. B. (2007). IQ and nonplanning impulsivity are independently associated with delay discounting in middle-aged adults. *Personality and Individual Differences*, *42*(1), 111–121. https://doi.org/10.1016/j.paid.2006.06.026

Dekel, E., Lipman, B. L., & Rustichini, A. (2009). Temptation-Driven Preferences. *Review of Economic Studies*, *76*(3), 937–971. https://doi.org/10.1111/j.1467-937X.2009.00560.x

Dick, D. M., Smith, G., Olausson, P., Mitchell, S. H., Leeman, R. F., O'Malley, S. S., & Sher, K. (2010). REVIEW: Understanding the construct of impulsivity and its relationship to alcohol use disorders. *Addiction Biology*, *15*(2), 217–226. https://doi.org/10.1111/j.1369-1600.2009.00190.x

Doyle, J. R. (2012). Survey of time preference, delay discounting models. *Delay Discounting Models (April 20, 2012)*.

Duckworth, A. L., Milkman, K. L., & Laibson, D. (2018). Beyond Willpower: Strategies for Reducing Failures of Self-Control. *Psychological Science in the Public Interest*, *19*(3), 102–129. https://doi.org/10.1177/1529100618821893

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259–268. https://doi.org/https://doi.org/10.1016/j.jrp.2011.02.004

Ebert, J. E. J., & Prelec, D. (2007). The Fragility of Time: Time-Insensitivity and Valuation of the Near and Far Future. *Management Science*, *53*(9), 1423–1438. https://doi.org/10.1287/mnsc.1060.0671

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472–5477. https://doi.org/10.1073/pnas.1818430116

Ericson, K., White, J. M., Laibson, D., & Cohen, J. (2015). *Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better than Delay Discounting* (tech. rep. w20948). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w20948

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*(4), 348–361. https://doi.org/10.1007/pl00005481

Farmer, J. D., & Geanakoplos, J. (2009). Hyperbolic discounting is rational: Valuing the far future with uncertain discount rates.

Fenneman, J., & Frankenhuis, W. E. (2020). Is impulsive behavior adaptive in harsh and unpredictable environments? A formal model. *Evolution and Human Behavior*, *41*(4), 261–273. https://doi.org/10.1016/j.evolhumbehav.2020.02.005

Fisher, I. (1930). *The Theory of Interest, as Determined by Impatience to Spend Income and Opportunity to Invest it.*

Fiske, D. W. (1971). Measuring the concepts of personality.

Gabaix, X., & Laibson, D. (2017). *Myopia and discounting* (Report). National bureau of economic research.

Green, L., & Myerson, J. (2013). How many impulsivities? A discounting perspective. *Journal of the experimental analysis of behavior*, *99*(1), 3–13. https://doi.org/10.1002/jeab.1

Gul, F., & Pesendorfer, W. (2001). Temptation and Self-Control. *Econometrica*, *69*(6), 1403–1435. Retrieved June 30, 2022, from https://www.jstor.org/stable/2692262

Gullo, M. J., & Dawe, S. (2008). Impulsivity and adolescent substance use: Rashly dismissed as "all-bad"? *Neuroscience and Biobehavioral Reviews*, *32*(8), 1507–1518. https://doi.org/10.1016/j.neubiorev.2008.06.003

Hamilton, K. R., Mitchell, M. R., Wing, V. C., Balodis, I. M., Bickel, W. K., Fillmore, M., Lane, S. D., Lejuez, C. W., Littlefield, A. K., Luijten, M., Mathias, C. W.,

Mitchell, S. H., Napier, T. C., Reynolds, B., Schütz, C. G., Setlow, B., Sher, K. J., Swann, A. C., Tedford, S. E., … Moeller, F. G. (2015). Choice impulsivity: Definitions, measurement issues, and clinical implications. *Personality Disorders: Theory, Research, and Treatment*, *6*(2), 182–198. https://doi.org/10.1037/per0000099

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, *27*(1), 1–17. https://doi.org/10.1080/1047840X.2016.1082418

He, L., Golman, R., & Bhatia, S. (2019). Variable time preference. *Cognitive psychology*, *111*, 53–79. https://doi.org/10.1016/j.cogpsych.2019.03.003

He, L., Zhao, W. J., & Bhatia, S. (2022). An ontology of decision models. *Psychological Review*, *129*(1), 49–72. https://doi.org/10.1037/rev0000231

Holton, R. (2009). *Willing, wanting, waiting*. Oxford University Press on Demand.

Hursh, S. R., & Schwartz, L. P. (2022). A general model of demand and discounting. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*. https://doi.org/10.1037/adb0000848

Ioannidis, K., Hook, R., Wickham, K., Grant, J. E., & Chamberlain, S. R. (2019). Impulsivity in gambling disorder and problem gambling: A meta-analysis. *Neuropsychopharmacology*, *44*(8), 1354–1361. https://doi.org/10.1038/s41386-019-0393-9

Jachimowicz, J. M., Chafik, S., Munrat, S., Prabhu, J. C., & Weber, E. U. (2017). Community trust reduces myopic decisions of low-income individuals. *Proceedings of the National Academy of Sciences*, *114*(21), 5401–5406. https://doi.org/10.1073/pnas.1617395114

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*(2), 159–186. https://doi.org/10.1017/bpp.2018.43

Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler Publishing Company.

Kirby, K. N., & Herrnstein, R. (1995). Preference Reversals Due to Myopic Discounting of Delayed Reward. *Psychological Science*, *6*(2), 83–89. https://doi.org/10.1111/j.1467-9280.1995.tb00311.x

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental psychology: general*, *128*(1), 78. https://doi.org/10.1037/0096-3445.128.1.78

Kivetz, R., & Keinan, A. (2006). Repenting Hyperopia: An Analysis of Self-Control Regrets. *Journal of Consumer Research*, *33*(2), 273–282. https://doi.org/10.1086/506308

Konstantinidis, E., van Ravenzwaaij, D., Güney, Ş., & Newell, B. R. (2020). Now for sure or later with a risk? Modeling risky intertemporal choice as accumulated preference. *Decision*, *7*(2), 91. https://doi.org/10.1037/dec0000103

Kraut, R. (2001). Aristotle's ethics.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478. https://doi.org/10.1162/003355397555253

Lane, S. D., Cherek, D. R., Rhodes, H. M., Pietras, C. J., & Tcheremissine, O. V. (2003). Relationships among laboratory and psychometric measures of impulsivity: Implications in substance abuse and dependence. *Addictive Disorders & Their Treatment*, *2*(2), 33–40. https://doi.org/10.1097/00132576-200302020-00001

Laude, J. R., Pattison, K. F., & Zentall, T. R. (2012). Hungry pigeons make suboptimal choices, less hungry pigeons do not. *Psychonomic bulletin & review*, *19*(5), 884–891. https://doi.org/10.3758/s13423-012-0282-2

Loewenstein, G. (1988). Frames of Mind in Intertemporal Choice. *Management Science*, *34*(2), 200–214. https://doi.org/10.1287/mnsc.34.2.200

Loewenstein, G. (2018). Self-control and its discontents: A commentary on Duckworth, Milkman, and Laibson. *Psychological Science in the Public Interest*, *19*(3), 95–101.

Loewenstein, G., O'Donoghue, T., & Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, *2*(2), 55–81. https://doi.org/10.1037/dec0000029

Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal choice. *Journal of Economic perspectives*, *3*(4), 181–193. https://doi.org/10.1257/jep.3.4.181

Logue, A. W. (1988). Research on self-control: An integrating framework. *Behavioral and Brain Sciences*, *11*(4), 665–679. https://doi.org/10.1017/S0140525X00053978

Madden, G. J., & Bickel, W. K. (2010). *Impulsivity: The behavioral and neurological science of discounting*. American Psychological Association.

Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control patients: Drug and monetary rewards. *Experimental and clinical psychopharmacology*, *5*(3), 256. https://doi.org/10.1037/1064-1297.5.3.256

Madden, G. J., Raiff, B. R., Lagorio, C. H., Begotka, A. M., Mueller, A. M., Hehli, D. J., & Wegener, A. A. (2004). Delay discounting of potentially real and hypothetical rewards: II. Between-and within-subject comparisons. *Experimental and clinical psychopharmacology*, *12*(4), 251. https://doi.org/10.1037/1064-1297.12.4.251

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Quantitative analyses of behavior*, *5*, 55–73.

Mele, A. R. (2010). Weakness of will and akrasia. *Philosophical Studies*, *150*(3), 391–404. https://doi.org/10.1007/s11098-009-9418-2

Mele, A. R. (2012). *Backsliding: Understanding weakness of will*. Oxford University Press.

Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making - Katherine L. Milkman, Todd Rogers, Max H. Bazerman, 2008. *Perspectives on Psychological Science*, *3*(4). https://doi.org/10.1111/j.1745-6924.2008.0008

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, *16*(2), 329. https://doi.org/10.1037/h0029815

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, *244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Mitchell, S. H. (1999). Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology*, *146*(4), 455–464. https://doi.org/10.1007/PL00005491

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American journal of psychiatry*, *158*(11), 1783–1793. https://doi.org/10.1176/appi.ajp.158.11.1783

Muela, I., Navas, J. F., Ventura-Lucena, J. M., & Perales, J. C. (2022). How to pin a compulsive behavior down: A systematic review and conceptual synthesis of

compulsivity-sensitive items in measures of behavioral addiction. *Addictive Behaviors*, *134*, 107410. https://doi.org/10.1016/j.addbeh.2022.107410

Myerson, J., & Green, L. (1995). Discounting of Delayed Rewards: Models of Individual Choice. *Journal of the Experimental Analysis of Behavior*, *64*(3), 263–276. https://doi.org/10.1901/jeab.1995.64-263

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, *76*(2), 235–243. Retrieved June 28, 2022, from https://onlinelibrary.wiley.com/doi/abs/10.1901/jeab.2001.76-235

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of child psychology and psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

O'Donoghue, T., & Rabin, M. (1998). Procrastination in preparing for retirement. *University of California-Berkeley Working Paper*.

O'Donoghue, T., & Rabin, M. (2015). Present bias: Lessons learned and to be learned. *American Economic Review*, *105*(5), 273–79. https://doi.org/10.1257/aer.p20151085

Odum, A. L. (2011). Delay Discounting: I'm a k, You're a k. *Journal of the Experimental Analysis of Behavior*, *96*(3), 427–439. https://doi.org/10.1901/jeab.2011.96-423

Odum, A. L., & Baumann, A. A. (2010). Delay discounting: State and trait variable. In *Impulsivity: The behavioral and neurological science of discounting.* (pp. 39–65). American Psychological Association.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6⟨768::AID-JCLP2270510607⟩3.0.CO;2-1

Rachlin, H. (2006). Notes on Discounting. *Journal of the Experimental Analysis of Behavior*, *85*(3), 425–435. https://doi.org/10.1901/jeab.2006.85-05

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the experimental analysis of behavior*, *17*(1), 15–22. https://doi.org/10.1901/jeab.1972.17-15

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of risk and uncertainty*, *23*(1), 5–32.

Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four Score and Seven Years from Now: The Date/Delay Effect in Temporal Discounting. *Management Science*, *51*(9), 1326–1335. https://doi.org/10.1287/mnsc.1050.0412

Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and individual differences*, *40*(2), 305–315. https://doi.org/10.1016/j.paid.2005.03.024

Rubia, K., Halari, R., Christakou, A., & Taylor, E. (2009). Impulsiveness as a timing disturbance: Neurocognitive abnormalities in attention-deficit hyperactivity disorder during temporal processes and normalization with methylphenidate. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1525), 1919–1931. https://doi.org/10.1098/rstb.2009.0014

Rubinstein, A. (2003). "Economics and psychology"? The case of hyperbolic discounting. *International Economic Review*, *44*(4), 1207–1216. https://doi.org/10.1111/1468-2354.t01-1-00106

Samuelson, P. A. (1937). A note on measurement of utility. *The review of economic studies*, *4*(2), 155–161.

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological bulletin*, *140*(2), 374. https://doi.org/10.1037/a0034418

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and individual differences*, *47*(5), 385–395. https://doi.org/10.1016/j.paid.2009.04.008

Stedman, T. L. (1995). *Stedman's Medical Dictionary: 26th Edition* (M. Spraycar, Ed.; 26th). Williams & Wilkins.

Steinglass, J., Figner, B., Berkowitz, S., Simpson, H. B., Weber, E. U., & Walsh, B. T. (2012). Increased Capacity to Delay Reward in Anorexia Nervosa. *Journal of the International Neuropsychological Society : JINS*, *18*(4), 773–780. https://doi.org/10.1017/S1355617712000446

Stevens, J. R., & Stephens, D. W. (2010). The adaptive nature of impulsivity. https://doi.org/10.1037/12069-013

Strickland, J. C., & Johnson, M. W. (2020). Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. *Psychological review*, *128*. https://doi.org/10.1037/rev0000263

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The review of economic studies*, *23*(3), 165–180.

Stroud, S. (2021). Weakness of Will. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. Retrieved December 20, 2022, from https://plato.stanford.edu/archives/win2021/entries/weakness-will/

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Thaler, R. H. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207. https://doi.org/10.1016/0165-1765(81)90067-7

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440. https://doi.org/10.1037/a0018963

Van den Bergh, B., Dewitte, S., & Warlop, L. (2008). Bikinis instigate generalized impatience in intertemporal choice. *Journal of Consumer Research*, *35*(1), 85–97. https://doi.org/10.1086/525505

van Baal, S. T., Moskovsky, N., Hohwy, J., & Verdejo-García, A. (2022). State impulsivity amplifies urges without diminishing self-control. *Addictive Behaviors*, *133*, 107381. https://doi.org/10.1016/j.addbeh.2022.107381

Vanderveldt, A., Oliveira, L., & Green, L. (2016). Delay Discounting: Pigeon, Rat, Human – Does it Matter? *Journal of experimental psychology. Animal learning and cognition*, *42*(2), 141–162. https://doi.org/10.1037/xan0000097

Vassileva, J., & Conrod, P. J. (2019). Impulsivities and addictions: A multidimensional integrative framework informing assessment and interventions for substance use disorders. *Philosophical Transactions of the Royal Society B*, *374*(1766), 20180137. https://doi.org/10.1098/rstb.2018.0137

Verdejo-García, A., Benbrook, A., Funderburk, F., David, P., Cadet, J.-L., & Bolla, K. I. (2007). The differential relationship between cocaine use and marijuana

use on decision-making performance over repeat testing with the Iowa Gambling Task. *Drug and Alcohol Dependence*, *90*(1), 2–11. https://doi.org/10.1016/j.drugalcdep.2007.02.004

Verdejo-García, A., Lawrence, A. J., & Clark, L. (2008). Impulsivity as a vulnerability marker for substance-use disorders: Review of findings from high-risk research, problem gamblers and genetic association studies. *Neuroscience and biobehavioral reviews.*, *32*(4), 777–810. https://doi.org/10.1016/j.neubiorev.2007.11.003

Vincent, B. T., & Stewart, N. (2020). The case of muddled units in temporal discounting. *Cognition*, *198*, 104203. https://doi.org/10.1016/j.cognition.2020.104203

Weatherly, J. N., & Terrell, H. K. (2010). Delay Discounting of Different Commodities II: Confirmatory Analyses. *The Journal of General Psychology*, *138*(1), 35–48. https://doi.org/10.1080/00221309.2010.532522

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric Discounting in Intertemporal Choice: A Query-Theory Account. *Psychological Science*, *18*(6), 516–523. https://doi.org/10.1111/j.1467-9280.2007.01932.x

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J., & Stouthamer-Loeber, M. (1994). Measuring impulsivity and examining its relationship to delinquency. *Journal of abnormal psychology*, *103*(2), 192. https://doi.org/10.1037/0021-843X.103.2.192

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and individual differences*, *30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Wiers, R., Ames, S. L., Hofmann, W., Krank, M., & Stacy, A. (2010). Impulsivity, impulsive and reflective processes and the development of alcohol use and misuse in adolescents and young adults. *Frontiers in psychology*, *1*, 144. https://doi.org/10.3389%2Ffpsyg.2010.00144

Chapter 2

Directed Deliberation in Intertemporal Choice

2.1 Linking Text for Chapter 2

In the previous Chapter, I developed a conceptual framework for understanding impulsivity, self-control, and time preferences. I then used this framework to put forward an account of the relationship between these three concepts. Based on logical arguments and evidence from the literature, I concluded that the three constructs dissociate theoretically and practically, and that this has implications for how we understand the claims made about impulsivity and self-control in intertemporal choice research, and about operationalising impulsivity and self-control more generally. One of these conclusions was that it is not evident whether more deliberation will lead to more future-oriented decision-making—this will depend on the decision-maker and the context.

Another, more tentative proposal is that since a difference between higher-order preferences and first-order preferences may motivate individuals to use their self-control (thus attenuating or breaking the connection between first-order desires and behaviour), people might change how they make decisions by focusing on their higher-order preferences. That is, if I want to eat fast food, but I don't *want to want* to eat fast food—mainly because I wish to be a healthy person—it could be that focusing on those latter two preferences either desensitises me toward my first-order desire or strengthens my resolve to control myself.

In the following Chapter, I will explore the implications of these conclusions further; I will see how people's decisions change when they think about intertemporal choices in ways that draw attention to one's higher-order preferences. This investigation will bring us closer to our goal of understanding how impulsivity and self-control manifest by showing whether deliberating more, in meaningful ways, will change decision processes and outcomes predictably. It will also bring us some way to understanding whether and how we can intervene in processes of impulsivity and self-control, and what implications this may have in real-life scenarios.

This Chapter will investigate preferences in intertemporal choice in light of two particular ways of thinking about decisions that not only relate to higher-order preferences in the model proposed in Chapter 1, but are also philosophically relevant. I will examine how people make decisions when they use their evaluative judgment (which is an essential component for Aristotle's account of *akrasia*; see Kraut, 2001), and when they use their meta-preferences (akin to second-order volitions; Frankfurt, 1988). Examining these two concepts in this context is interesting as a philosophical project as well as an empirical one because aside from shedding light on our proposed model of how impulsivity and self-control affect behaviour, it gives us insight into people's capacity to be akratic (weak-willed) and their capacity for having second-order desires that are different than their first-order desires, which is a fundamental part of what makes a person, according to Frankfurt.

Directed Deliberation Alters Intertemporal Choice

Simon T. van Baal, Emmanouil Konstantinidis, Joyce W. Zhao, John Michael, Daniel Read, Jakob Hohwy, Lukasz Walasek

Abstract

Many choices are made by deliberating, using heuristics, intuition, rules, or out of habit. While much decision-making research is devoted to finding the best descriptive model for people's choices on aggregate, less is known about how engaging different processes for making choices can influence outcomes. It is therefore unclear if there would be systematic differences in outcomes when people deliberate about what the normative option is, or about how the options fit with the preferences they have for themselves, compared to how they normally make such a decision. We examine this question using two intertemporal choice experiments. In the first study, we elicit meta-preferences and counterfactual judgments, and compare them to what participants normally do in incentivised choices. Using this setup, we assess whether these differences in the decision-making process influence participants' likelihood of choosing future rewards. To gauge the usefulness of directed deliberation for relevant sub-samples, we test whether eliciting meta-preferences produces similar results for those with high borderline personality disorder trait scores. In a second experiment, we elicit participants' evaluative judgment and compare it to regular choices. We assess whether prompting participants to use evaluative judgment changes their preferences, and we examine whether it can reduce the common delay effect. We also compare model fits of two commonly used discounting models across conditions. We find that participants' choices are significantly more future-oriented when they think of their meta-preferences. When evaluative judgment is elicited, participants make fewer common delay preference reversals. These results show that deliberating a choice in light of different aspects of one's beliefs and preferences about oneself might systematically change people's decisions between rewards that arrive at different time points.

keywords: intertemporal choice, delay discounting, time preferences, deliberation, higher-order preferences

2.2 Introduction

Trade-offs between outcomes that occur sooner and later feature in most choices in life. Our food choices, decisions on whether to obtain another degree, leaving our job for a different one, choosing a partner—all have outcomes that occur at different times. Choices that involve these tradeoffs are called intertemporal choices, and much thought and research are devoted to how people make these choices. The most common and simple experimental method for investigating people's preferences in intertemporal choice entails presenting participants with choices between monetary amounts that would be received at different points in time (Cohen et al., 2020).

Much decision-making research is concerned with creating good descriptive models of how we, in aggregate, make decisions. In the context of intertemporal choice, this means that researchers have traditionally examined whether people generally make intertemporal choices in accordance with the discounted utility model (Samuelson, 1937), the traditional benchmark for economic rationality in intertemporal choice. A prominent example is when researchers found people would often commit so-called *common delay preference reversals* (Frederick et al., 2002; Green et al., 1994; Kirby & Herrnstein, 1995; Solnick et al., 1980). This phenomenon entails that an agent may prefer a large reward that arrives in the future over a smaller reward that would arrive sooner, but when the smaller option becomes immediate, they would revert their preference, choosing the smaller-sooner option (Strotz, 1955).

In order to account for common delay preference reversals, the hyperbolic discounting model was developed (Ainslie, 1975; Loewenstein & Prelec, 1992; Mazur, 1987), which can accommodate these preference reversals by accounting for *decreasing impatience*; people tend to discount delayed rewards most when they are close to the present (i.e., decision-makers care more about a delay in the receipt of their reward from now to tomorrow, than about a delay from next month to one day after). Since then, a large number of different kinds of models were developed to account for various phenomena

similar to this one, where people's behaviour deviates from what the discounted utility model, or indeed the hyperbolic model predicts (for an overview, see He et al., 2022).

Although having an accurate descriptive model that predicts what people most often do is highly useful, it is also important to examine the underlying mechanisms of intertemporal choice (Dai & Busemeyer, 2014), and what happens when people use different processes to construct their preferences and make decisions; this area has traditionally received less attention. This area of research is crucial because people are known to modulate how they make decisions. Sometimes, for instance, people will rely on established habits, heuristics, or intuition (Ericson et al., 2015; Gigerenzer & Gaissmaier, 2011; Kahneman, 2011). These ways of making decisions are particularly useful for reducing the cognitive resources for choices that are not deemed too consequential. For more impactful choices, such as when an individual is deciding whether to take a job in a different country or to stay put at their current organisation, they might deliberate about their options more. Someone may, for example, think about what preferences they have for themselves. Who do I want to be in ten years? Which workplace fits better with that picture? What working environment do I want for myself? The thinking is that these meta-preferences may then help elucidate the right course of action.

In the literature on heuristics, some argue that in many cases using certain heuristics is rational and that they can outperform other ways of making decisions, in particular more complex models (Dosi et al., 2020; Gigerenzer, 2016). However, there is little such research on what effect different types of deliberation or contemplation have on choices. Even though researchers already provide widely varying instructions on how to make choices, for example, Amasino et al. (2019) tell their participants to "remember to think about what advice your friend would give before making your decision". It is, therefore, important we figure out how deliberation influences the decision-making process, so we can judge which ways of deliberating are useful in which situations.

Research on the influence of deliberation in the context of intertemporal choice is particularly likely to be fruitful because there is overwhelming evidence that preferences in this context are highly susceptible to subtle manipulation. For example, framing effects (Lempert & Phelps, 2016), changes in preferences due to sub-additive discounting, when delays are presented differently (i.e., people change their preferences when a delay is segmented into multiple parts; see Read, 2001), or preference reversals when

the absent reward for each option is made explicit (e.g., $80 now and $0 in one month, or $0 now and $100 in one month; Magen et al., 2008; Naudé et al., 2018).

Several researchers have, therefore, called for more research on the influence of deliberation on decisions (see e.g., Bulley & Schacter, 2020; Dai & Busemeyer, 2014). Bulley and Schacter (2020) found that when people are required to justify their response to an intertemporal choice, they were less biased towards taking a smaller-sooner reward in choices with small stakes. This common phenomenon is called the magnitude effect (Read, 2004)—where people tend to prefer smaller-sooner rewards for small stakes but larger-later rewards for large stakes. The result that the magnitude effect can be attenuated using a simple justification procedure shows that this kind of deliberation, and perhaps other kinds, can alter outcomes in intertemporal choice considerably.

It could be beneficial, however, to make our deliberation processes more targeted; what if a subset of decision-makers is looking in the wrong place for their justification? Further, as of yet it is unclear what aspect of this type of deliberative manipulation causes the observed attenuation of the magnitude effect. Discovering which types or elements of deliberation are sufficient for a change in behaviour, and under what conditions, will help pinpoint practices that lead to better outcomes with the least time cost. It might be, for example, the act of deliberating itself, a sense of accountability (Lerner & Tetlock, 1999), or instead the specific way of mentally framing choices (in this case: actively coming up with reasons for what one should decide) that drives the attenuation of the magnitude effect.

One way to direct the deliberative process is to have people ask themselves which option is best, all things considered. There is reason to believe that this may be effective because oftentimes our *evaluative judgment* (i.e., our judgment on what is the best course of action) is not fully formed (Holton, 2009). Instigating this deliberative process could be a way of making sure that we do not rely on a heuristic to break a tie if we do not know for sure which option to choose. For example, if someone is deciding between moving to a cheaper apartment further from their work, or to a more expensive one that is closer to their work, one might consider the annual costs of driving a car to work, the disutility of longer commutes etc., instead of quickly breaking the tie by the looking at the house price or rent.

Another way to direct the deliberative process would be to match the options with one's meta-preferences (closely related to second-order volitions, see Frankfurt, 1988). That is, to see how each option fits with one's preferences for one's own desires and actions. In the context of smoking, for example, when faced with the choice to light a cigarette, someone might ask themselves: "Do I want to be a smoker?" and "Does this action fit with whom I want to be?". This type of deliberation is easily applied to a standard intertemporal choice task too, as one might ask: "Do I want to make sure I am all set for the future, or should I enjoy my short time on this earth a little more?".

However, discerning how directing the deliberation process in these ways will influence people's choices is not a trivial matter. Each decision-maker might have different wishes for themselves, and certainly not all people will agree on what they consider is best. We will be concerned, therefore, with cases where it is known that people's decision-making is often sub-optimal or irrational. In addition, we are interested to see whether these deliberative processes cause people to follow a different decision-making model than they would normally.

In this paper, then, we investigate whether and how directing the decision-making process through meta-preferences and evaluative judgment might systematically change people's decisions in the context of intertemporal choice.

To do this, we conduct two studies aimed to discover whether each way of directing deliberation can change people's choices. In the first study, we investigate whether directing deliberation toward meta-preferences changes decision-making in a model-free way. We anticipate that meta-preferences will elicit more future-oriented responses for small stakes because the small rewards will be cast in a wider frame that emphasises the importance of acting in accordance with one's preferences about one's own actions. To gauge whether directed deliberation might be of use to relevant sub-populations (i.e., groups that have different intertemporal choice patterns, especially when related to psychopathology), we test whether eliciting meta-preferences differentially affects choices of those with high borderline personality disorder trait scores.

In the second study, we assess whether evaluative judgment can not only increase future-oriented decision-making, but also reduce common delay preference reversals, decrease dominated choices, and make people's choices fit the discounted utility model better. Here we expect that getting participants to deliberate on what the best option for them

is, will make them more "rational". That is, it would reduce the number of decisions that negatively impact their lifetime utility.

2.3 Experiment 1

In this experiment, we investigate whether directing the decision-making process towards meta-preferences can increase future-oriented decision-making for small stakes. We use a model-free method to analyse participants' choices; we analyse the likelihood of participants choosing a smaller-sooner or larger-later option.

Participants are instructed to either think about the choice in terms of their meta-preferences, engage in counterfactual reasoning, or make an incentivised choice (with minimal directions). The design for this experiment was constructed such that it was possible to partition the effect of the manipulation and any effect that might be due to abstract counterfactual reasoning about rewards.

Thus, our first research question was: "Do people their meta-preferences differ from their incentivised choices, and if so, is this due to abstract counterfactual reasoning?"

Another objective of this study was to see whether being prompted to deliberate about meta-preferences has a lasting influence on subsequent decisions where the meta-preference prompts are absent. This would be tentative evidence that participants' meta-preferential mindset carries over into subsequent thinking in some way, or possibly even that they keep applying this type of directed deliberation because they deem it helpful.

Our second research question was, then: "Does meta-preferential deliberation have carryover effects in subsequent choices?"

To evaluate the usefulness of our directed deliberation intervention, we investigated how using meta-preferences could impact groups of people who are most likely to benefit from a change in how they make intertemporal choices. One population that is known to strongly prefer rewards in the present over those in the future are people with borderline personality disorder, this preference is often suggested to be an important factor in the aetiology of the disorder (Barker et al., 2015; Lawrence et al., 2010). Tools that promote future-oriented decision-making could, therefore, be of particular use to them.

As such, our final research question, was: "Is the difference between meta-preferential choices and incentivised choices the same for those who score high on borderline personality disorder, as for those with low scores?"

2.3.1 Methods

2.3.1.1 Design

This study was conducted online using Prolific and Qualtrics survey software. We used a within-participant design with three conditions: the Meta-Preference condition, the Counterfactual condition, and the (incentivised) Control condition. All participants were required to complete the three conditions sequentially. Participants were randomly allocated to the six possible permutations of the block ordering.

All participants got a base rate payment of £1.00 for participating in the experiment. In the Control condition, one of the choices was incentivised, which meant that participants received a bonus payment after the option delay had elapsed. For instance, if a participant were presented with a choice between £3.00 today and £5.00 in four weeks, and they chose the latter, then if this choice were selected, the experimenter would complete the bonus payment to their Prolific account four weeks later.

2.3.1.2 Participants

To obtain a sufficient sample size with our design, we calculated that a sample of 400 participants would be sufficient to detect a 5 percentage point difference in the likelihood of selecting a larger-later with 90% power.

We excluded participants who chose only smaller-sooner or larger-later options throughout the experiment (i.e., no switching) because such a choice pattern would provide no information about the effectiveness of the manipulation, and since this would indicate that the range of options in the choice set was not right for them. That is, they preferred all the smaller-sooner options over what we offered in the paired larger-later options, or vice versa. Considering that we used a within-subjects design, this did not lead to any further biases due to uneven exclusions between conditions.

To ensure sufficient power, an initial sample of 637 subjects who reside in the United Kingdom was recruited from the Prolific database. Participants were screened for English fluency, no dyslexia, no dyscalculia, and normal or corrected-to-normal vision.

2.3.1.3 Procedure

Participants were presented with three blocks of twelve intertemporal choices each and these blocks were presented in random order (six permutations). The three blocks had slightly different instructions and choice prompts, asking participants to make meta-preferential judgments, counterfactual judgments, or incentivised choices. See Table 2.1 for the condition-specific instructions and prompts.

TABLE 2.1: The instructions and prompts by condition for the first experiment.

Condition	Meta-Preference	Counterfactual	Control
Instructions	We want you to imagine you actually were confronted with these choices. If you were, what is it that you would like yourself to choose?	We want you to imagine you actually were confronted with these choices. If you were making this choice, what is it that you expect to choose?	Now you will actually be confronted with these choices. One of the questions will be randomly selected for payment. The date and amount of your reward will be determined by the answer you provide to the selected question.
Choice prompt	What would you like yourself to choose?	What do you imagine you would choose?	What do you choose?

Each intertemporal choice consists of two options that entail monetary rewards, of which one would pay out on the day of the experiment, while the other would pay out in either 21, 28, 35, or 42 days. The rewards the participants chose between were paired over the three conditions. That is, the twelve questions in each were not identical, but comparable: both options would be either £0.25 higher or lower in magnitude than the corresponding question in one of the other conditions. See Table S1 for a table showing all the option pairs.

After the participants completed the intertemporal choice task, they were asked to complete the Five Factor Borderline Inventory – Short Form (DeShong et al., 2016).

2.3.1.4 Analysis

For the main analysis, we used a binomial GLMM to calculate the likelihood of a participant reporting a larger-later choice. We incorporated the participant as a random intercept and the effect of individual questions and the condition as random slopes. The fixed effects were the condition and the daily earnings—how much the participant would earn per day by waiting for the larger-later reward.

We also tested whether there were any carryover effects from the Meta-Preference condition. To do this, we analysed whether participants who started with the Meta-Preference condition were more likely to choose larger-later options in consecutive conditions than other participants. For this analysis, we created another variable indicating the position of the Meta-Preference condition in the order. We added this variable to the model and created an interaction term with the condition variable. The full model consisted, thus, of the probability of choosing a larger-later as the dependent variable, the condition, the position of the Meta-Preference condition in the order (with an interaction term), the daily earnings as a control variable, the participant as the random intercept, with the question as the random slope.

In addition, we investigated whether the effect of the Meta-Preference condition would be of a different magnitude in people who scored high on borderline personality disorder traits using another GLMM and including an interaction term between their score on the FFBI-SF (DeShong et al., 2016).

We used the R package `afex` (Singmann et al., 2015) to run the GLMM analyses and likelihood ratio tests, and we used the `emmeans` package (Lenth et al., 2018) to test for pairwise differences between conditions. We assessed the significance of tests according to a false discovery rate adjusted α of .05.

2.3.2 Results

The final sample consisted of 193 females (48.25%) with an average age of 33.0 (SD = 12.4) and 207 males (51.75%) with an average age of 28.3 (SD = 10.3). Participants who selected only one type of reward (only smaller-sooner or only larger-later) were excluded (n = 237).

In the analysis for the likelihood of selecting the larger-later, there was significant support for including the condition factor in the model, $\chi^2(2) = 59.56$, $p < .0001$. Participants were more likely to select larger-later options in the Meta-Preference condition ($M = .616$, $SD = .276$) than the Control condition ($M = .532$, $SD = .276$), OR = 4.600, 95% CI [2.633, 8.037], $z = 6.548$, $p < .0001$, and than in the Counterfactual condition ($M = .514$, $SD = .251$), OR = 6.170, 95% CI [3.452, 11.03], $z = 7.504$, $p < .0001$. Participants were less likely to select larger-later options in the Counterfactual condition than the Control condition but this difference was not statistically significant, OR = 7.46, 95% CI [.503, 1.106], $z = -1.782$, $p = .07$. See Figure 2.1 for a depiction of the pairwise comparisons of the estimated marginal means for the likelihood of selecting a larger-later by condition.

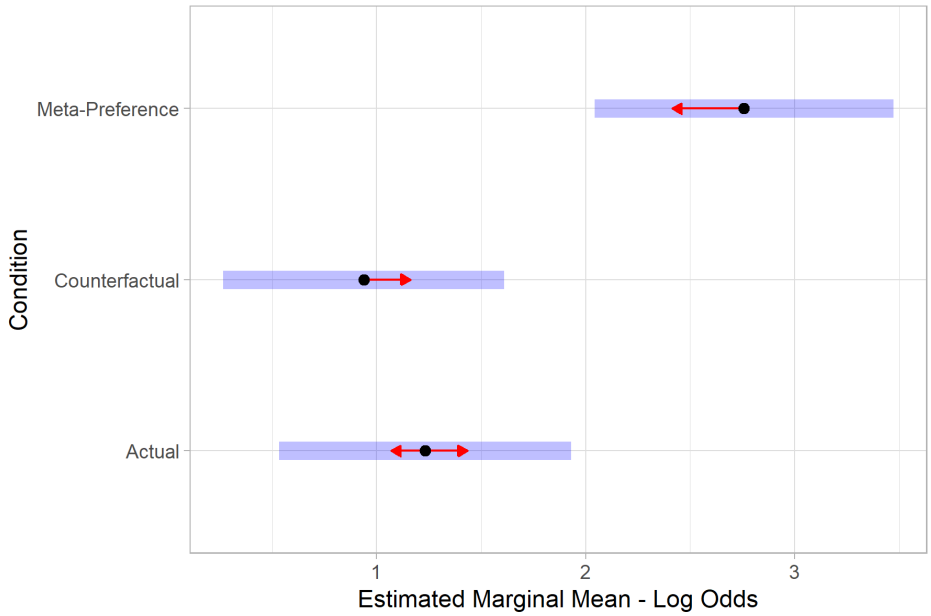

FIGURE 2.1: The estimated marginal means, with pairwise comparison arrows.

We found no evidence that the effect of the Meta-Preference condition carried over into the subsequent choice blocks. There was no support for including a term for the position of the Meta-Preference block in the order, $\chi^2(2) = 4.02$, $p = .13$, but there was support for the model with the interaction term between the position and the condition, $\chi^2(4) = 140.11$, $p < .001$. Post-hoc tests show that this interaction was driven by a difference between the position of the Meta-Preference condition on the effect of that same condition. That is, the odds of reporting a larger-later in the Meta-Preference condition were higher when that condition came second ($M = .638$, $SD = .265$), OR = 3.629, 95% CI [1.589, 5.669], $z = 4.493$, $p < .001$, or third ($M = .655$, $SD = .261$), OR

= 4.787, 95% CI[2.067, 7.508], $z = 5.400$, $p < .001$, than if it was administered first (M = .554, SD = .293); all other ps > .1.

We found significant support for a model including borderline personality disorder trait scores, $\chi^2(1) = 9.31$, $p < .01$. People with high borderline personality disorder trait scores were significantly less likely to choose larger-later options, b = -.263, 95% CI [-.430, .096], $z = 3.078$, $p < .01$. We found no support, however, for including an interaction term between the borderline personality disorder trait score and the condition factor, $\chi^2(2) = .93$, $p = .63$. This means that there is no evidence that there is an interaction effect between the condition and borderline personality disorder traits.

2.3.3 Discussion

The results of this experiment show that participants are more likely to choose larger-later options when they are directed to make decisions with their meta-preferences in mind. Moreover, the increased future-oriented decision-making in the Meta-Preference condition over the Counterfactual condition suggests that this effect is not an artefact due to abstract reasoning.

There are, however, multiple possible explanations for this effect. First, this type of directed deliberation may increase cognitive control; Wilson and Schooler (1991) found that cognitive control can be stimulated by asking participants to provide reasons for their choices, and (Bulley & Schacter, 2020) found that this type of manipulation can decrease the magnitude effect.

Alternatively, by thinking about their preferences for their own behaviour, people could have reduced their present bias, making their discounting follow the discounted utility model (Samuelson, 1937) more closely, or merely decreased their discounting parameter k—this is what is normally assumed to happen in studies that test manipulations for intertemporal choice.

The setup of the current experiment, however, does not allow us to identify which models of intertemporal choice participants' responses are most closely aligned with. Therefore, we are unable to determine which of these alternative theories has the most explanatory power using the current experiment. To illustrate: there is no way of testing cognitive control with the current task (i.e., what does this look like in a two-alternative choice

task with 12 questions per block?), and because the choice set is relatively small and all smaller-sooner options pertain to a reward in the present, while the larger-later options are all fairly close in the future, it is difficult to differentiate between intertemporal choice model fits.

Further, the effect of meta-preferences in this experiment occurred in decisions with small stakes. It is not yet clear whether the effect of directed deliberation would be generalisable to decisions with larger stakes, especially given the evidence that providing reasons for intertemporal choices can reduce the magnitude effect (Bulley & Schacter, 2020).

The results provide no evidence that making decisions with meta-preferences in mind increased future orientation more or less for participants who scored high on borderline personality disorder traits than for those who scored low. Therefore, we may still expect meta-preferences to be effective for people who struggle with excessive future discounting, but not necessarily more so than for others.

We also found no evidence for a carryover effect, but we found that the meta-preference condition predicted higher proportions of larger-later choices if the manipulation was administered later in the experiment. The most plausible explanation for this, it seems, is that because the manipulation was quite subtle, participants benefited from this contrast with the previous condition.

In sum, we have shown that directed deliberation can change the likelihood of selecting larger-later decisions, but we have not uncovered much about the mechanism for this change. It may be that this decision-making process changed the way participants made decisions, changed their discounting parameters, or it may have increased their cognitive control. Thus, in order to see whether directed deliberation can change people's decisions in larger stakes, whether it makes them more "rational" by increasing the effort exerted in forming their judgment, or whether it causes participants' discounting to become less *steep* or more exponential (less hyperbolic), another experiment was conducted.

2.4 Experiment 2

The first experiment showed that eliciting meta-preferences changes participants' choices. We saw, in a model-free way, participants increase their preference for larger-later options for decisions with small stakes. However, the first experiment did not allow us to determine the mechanism through which directed deliberation influences behaviour. In this experiment, we investigate these mechanisms, but we also test whether we can reproduce the change in preferences we observed for a different kind of directed deliberation—using evaluative judgment.

To identify how behaviour was influenced by prompting directed deliberation, we use a choice set taken from He et al. (2019). In this choice set, participants face choices where the smaller-sooner option arrives the same day (as in the first experiment), but also choices where the smaller-sooner option is in the future. This setup allows us to analyse whether participants become less likely to commit *common delay preference reversals* when directed deliberation is prompted. These preference reversals occur when participants choose the larger-later when both options are far away in the future, but as soon as the smaller-sooner nears immediacy, they change their preference to pick a smaller-sooner option (also referred to as *dynamic inconsistency* and the *common delay effect*.

The common delay effect is a phenomenon of interest because it reduces the overall welfare of the agent (see e.g., Ainslie, 1975; Loewenstein & Prelec, 1992; Thaler, 1981). The intuition for this is that if an agent structurally prefers A over B at time t but then chooses B over A at a later time τ, the agent is never able to achieve A (unless they can exclude B as an option). These preference reversals are a better metric for judging an improvement in decision-making following a manipulation, compared to an increase in larger-later choices, which may merely reflect a change in preferences (as there is no 'right' slope for a discount function—there are many good reasons to discount future outcomes heavily).

The choice set we use also contains choices with dominated options, where the later option has a smaller monetary reward, and thus is worse in both attributes. We take the number of *failures of dominance*—when participants choose this smaller-later option—as a tentative inverse measure of cognitive control. Lastly, this choice set is larger than the

one in the previous experiment (60 choices), allowing us to estimate discount functions and better assess differences in discounting parameters and model fits.

The choice set also enables us to analyse whether participants' preferences adhere to the discounted utility model. We take a delay discounting approach to investigating how directed deliberation changes people's choices. Specifically, we fit the participants' decisions to the discounted utility model (Samuelson, 1937):

$$f(D) = e^{-kD}, \tag{2.1}$$

where f is the discount factor with which future rewards are discounted, D is the delay and k is the discounting parameter.

People's choices are said to fit better, on aggregate, with hyperbolic discounting since people often exhibit present bias (see e.g., Loewenstein & Prelec, 1992; Thaler, 1981), so we also fit the hyperbolic discounting model to the data. This model is formalised as follows (Mazur, 1987):

$$f(D) = \frac{1}{1 + Dk}, \tag{2.2}$$

where D and k mean the same as in Equation 2.1. Aside from model-free estimates of larger-later choices, common-delay preference reversals and dominated choices, model comparisons and the parameter estimates for these two models will be our focus for the analysis.

Fitting the delay discounting models allows us to see whether participants' discounting parameters and model fits change when they are directed to use evaluative judgment compared to how they normally choose. Comparing these models will show, for instance, if participants' choices fit the discounted utility model better than the hyperbolic model when they are prompted to direct their deliberation, compared to what they normally do.

Thus, in this experiment we investigate whether eliciting evaluative judgment can (1) change discounting parameters, (2) decrease common delay preference reversals, (3) decrease failures of dominance. In addition, we see whether the manipulation can make participants' discounting fit better with the discounted utility model, as opposed to the hyperbolic discounting model. Lastly, we confirm whether this different type of

directed deliberation can also increase the number of larger-later options chosen, like in the previous experiment.

2.4.1 Methods

2.4.1.1 Design

We used a counterbalanced within-participant design with two conditions (Evaluative Judgment and Control), where all participants were subjected to two blocks of 60 intertemporal choice pairs, and the items in each block were identical. Participants were randomly selected into a group with one of the two possible orders of block presentation.

The 60 choice pairs were taken from (He et al., 2019). Each intertemporal choice consisted of two options that entail monetary rewards. The hypothetical payout of option 1 (t1) was either on the day of the experiment, in 3 months, in 6 months, or in 9 months. The delay of the other option (t2) would be t1 + 3 months, 6 months, or 9 months. The interest rates of the options are held constant over the various time differences between options. The maximum number of dominance failures a participant could commit per block was 12. See Table S2 for the full item set.

According to our simulation-based power calculations, 600 participants were required in order to attain 80% power for detecting a difference with an effect size of $d = .11$ at an α of .05, which is what we deemed to be the minimum theoretically significant effect size for a change in larger-later proportions for this choice set.

Participants were paid £1.75 for participating.

2.4.1.2 Participants

We used the Prolific participant recruitment platform to recruit subjects residing in the UK. They were screened for English fluency, no dyslexia, no dyscalculia, and normal or corrected-to-normal vision.

In accordance with the preregistration, we excluded participants if they did not complete the experiment or selected only one type of reward (only smaller-sooner or only larger-later). Similarly, participants who chose only larger-later options, but did not choose

the smaller-later choices in the trials with dominated options were excluded. This was done because modelling discount functions for data at the boundaries does not provide meaningful parameter estimates, and because it provides us with no information on the differences between the two conditions.

2.4.1.3 Procedure

Participants were directed from the Prolific platform to Qualtrics, where they first read through an information sheet, after which they signed consent. They were then prompted with an attention check that read "Which of the following is not a city in the US", in the title but the subtitle was clarified with "Regardless of the right answer, please select Chicago". Participants who did not answer "Chicago" were filtered out by Qualtrics.

The two experimental conditions had slightly different instructions, telling the participants that they were to make evaluative judgments or preferential choices. In addition, the three blocks prompted participants with different phrasing of the questions posed with the intertemporal choices. See Table 2.2 for the block instructions and choice prompts for each condition.

Table 2.2: The instructions and prompts by condition for the second experiment.

Condition	Evaluative Judgment	Control
Instructions	You will be presented with 60 questions that have two options each. The options are comprised of a money amount and a time when the reward would be paid out. You will be asked to judge which option is best for you, all things considered. We want you to imagine that you are actually presented these choices. An example of an option is: '£100 today'. If you indicate that this option is best, you mean to say that: it would be best for me, all things considered, to receive £100 today, rather than the other option.	You will be presented with 60 questions that have two options each. The options are comprised of a money amount and a time when the reward would be paid out. You will be asked to choose an option. We want you to imagine that you are actually presented these choices. An example of an option would be: '£100 today'. If you choose this option, you mean to say that: I choose to receive £100 today, rather than the other option.
Choice prompt	Which is best for you, all things considered?	Please choose one of the options

The questions within the blocks were displayed in random order, and the presentation of the options was randomised as well (i.e., smaller-sooner options could appear on the left or right). In this experiment, questions in both blocks were the same, so in an effort to clear participants' working memory of the choices they remembered from the first block, participants were asked to solve five simple arithmetic problems before the second block.

2.4.1.4 Analysis

The preregistration for this study can be found here. In the preregistration, we proposed to use the ANOVA framework for much of the analysis for this experiment, but considering the flexibility and robustness that generalised linear mixed models offer, we chose to use GLMMs throughout for the main analyses.

We tested whether our manipulation affected the likelihood of choosing a larger-later option using a binomial generalised linear mixed model. Only trials without dominated options were used for this analysis. The independent variables we included were the condition and the group, as well as an interaction term. We also included a random intercept for the participant. The dependent variable was whether the participant chose a larger-later (coded as 1), or a smaller-sooner (coded as 0). The condition and the group were included as fixed effects, with an interaction term. The participants and the question within the block were included as random intercepts.

To assess people's delay discounting rates, we fit the discounted utility model and the hyperbolic discounting model to participants' choices in each block. Specifically, we estimated the subjective value of a reward by estimating discounting parameters and multiplying the reward amount by that discount factor. We did this using Equation 2.1 and Equation 2.2, which gives:

$$V(R, D) = u(R) * f(D) \tag{2.3}$$

where *V(x)* is the subjective value of some reward x. R is the magnitude of the reward, and D is the delay, as before. For this study we assume *u(R)* to be the identity function, which is fairly standard (see e.g., Vincent & Stewart, 2020). To account for the probabilistic nature of human decision-making, we used a Fechner model (Becker et al.,

1963, 1964). Each discounted subjective value V was subsequently fed into a logistic choice rule:

$$\hat{y} = \frac{1}{1 + e^{\theta(V(ss) - V(ll))}} \tag{2.4}$$

where $\hat{y}$ is the choice prediction generated by the model; $\hat{y} = 0$ for $P(\text{ss}) = 1$ and $\hat{y} = 1$ for $P(\text{ll}) = 1$. Let ϑ be the temperature parameter, indicating how deterministic the participant is in their choices when deciding between options that have different discounted values. When ϑ is equal to 0, a participant is not influenced by differences in the subjective value of the options—often values of ϑ vary from 0-1. To compare each model's prediction, $\hat{y}$, against the actual choice the participant made, y, we calculated the log-likelihood for each choice:

$$LL(y) = ylog(\hat{y}) + (1 - y)log(1 - \hat{y}). \tag{2.5}$$

Subsequently, we grid-searched starting values for optimising the parameter estimates using maximum likelihood estimation for each model, and bounded all free parameters between 0 and 1, to find parameter values that minimise the summed negative log-likelihood:

$$nLL = -\sum_{n=1}^{N} LL_n, \tag{2.6}$$

where n is a trial in the complete set of trials N per condition per participant.

The discounting parameter k for the discounted utility model was then used as the dependent variable in a fractional logit mixed model (Brooks et al., 2017), with the condition as the independent variable, the question as a random slope and the participant as the random intercept.

We also fitted a fractional logit mixed model to analyse the proportion of possible common delay preference reversals a participant committed in each block of 60 choice pairs. Given the structure of the item set, participants could commit maximally 12 preference reversals, which we define as a reversal from preferring a smaller-sooner choice to a larger-later choice when a common delay is added to both options. For example, if a participant is to choose between \$100 today versus \$109 in three months, and chooses the smaller-sooner, but then chooses the larger-later when their choice is between \$100 in three months and \$109 in six months.

We analysed the number of failures of dominance with a generalised linear mixed model using the Poisson family and a logit link. For the dependent variable, we used only the dominated choices in the choice set, with the condition, group, age, and sex as the fixed effects, the participant as a random intercept, and the condition as a random slope.

For an exploratory analysis, we fit a linear mixed model to compare model fits between the hyperbolic discounting model and the discounted utility model per condition. For this analysis, the log-likelihood for each model fit to a participant's choices was the dependent variable, such that there were two observations for each participant per block. The predictors were the group and the condition, and we included the participant as a random intercept.

For completeness in the analysis of how evaluative judgment changes behaviour, we also analyse differences between conditions in the other estimated parameters. That is, we ran the same analysis as we did for discounted utility model parameter k_{exp} for the hyperbolic discounting model parameter k_{hyp}, and for ϑ in both the hyperbolic and exponential model as the dependent variables. We report these analyses in Appendix C.

As mentioned in the preregistration, the data for the 2.5% worst model fits were excluded from the main analyses.

2.4.2 Results

We recruited 630 participants and excluded those who chose only larger-later options (or only chose larger-later options, excluding the dominated trials) from the analysis for the non-dominated trials (n = 15). We then excluded participants with the 2.5% worst model fits for the discounted utility model (n = 15), leaving us with a final sample of 600 participants. The lower number of excluded participants compared to the first experiment reflects the difference in the number of choice pairs, the fact that the stakes are larger, and thus that this experiment captures a wider range of plausible time preferences.

Of these 600 participants, 376 (62.67%) were female with an average age of 36 (SD = 11.4), and 222 (37.0%) were male with an average age of 36.8 (SD = 12.0); 87.5% were United Kingdom nationals.

2.4.2.1 Main Analysis

We found insufficient evidence that participants selected more larger-later options in the Evaluative Judgment condition (M = .482, SD = .320) than in the Control condition (M = .476, SD = .308), Odds Ratio (OR) = 1.04, 95% CI [.996, 1.08], z = 1.774 p = .076. For reference, a coefficient of 1.04 means that the odds of choosing a larger later in the Evaluative Judgment condition versus in the Control condition were 1.04:1.

There was a significant difference in the number of common delay preference reversals reported in the Evaluative Judgment condition (Pr = .074) compared to the Control (Pr = .115). The odds ratio of participants reverting their preference after adding a common delay in the Evaluative Judgment condition versus the Control condition was .611:1, 95% CI [.409, .913], $t(1188) = 2.406$, p = .02. See Figure 2.2.

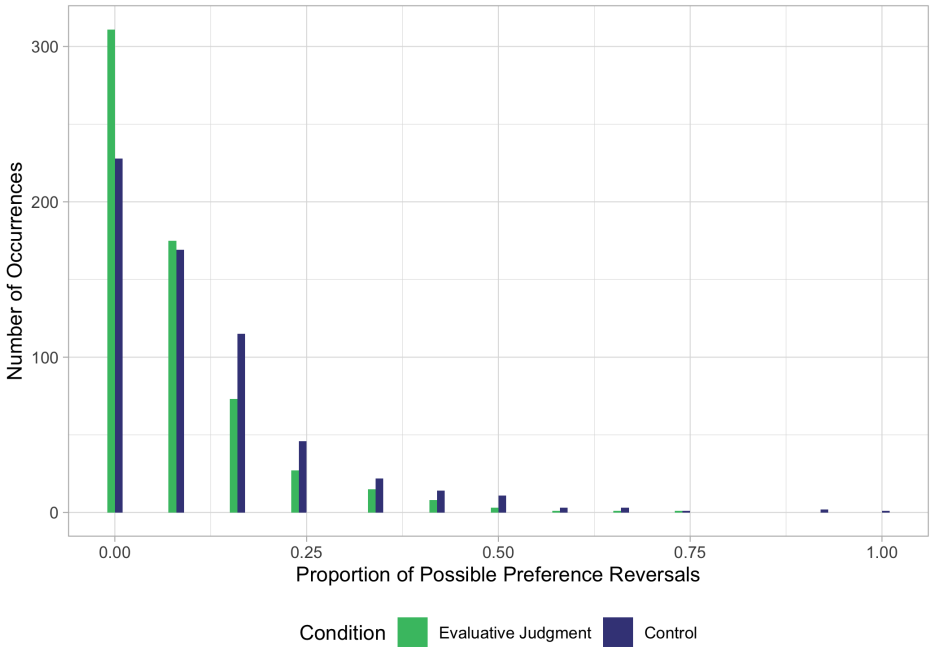

FIGURE 2.2: The proportion of possible preference reversals in each condition. The error bars represent 95% CI.

For the analysis on the likelihood of committing a failure of dominance, there was significant support for a model including the condition, $\chi^2(1) = 5.10$, p = .02. Participants committed more failures of dominance in the Evaluative Judgment condition (M = .062, SD = .441) than in the Control condition (M = .037, SD = .308), contrary to our predictions, OR = 2.175, 95% CI [1.063, 3.290], z = 2.175, p = .03. Showing that participants were twice as likely to commit a failure of dominance in the Evaluative

Judgment condition, though the base rate was very low—an average of .0374 failures of dominance per participant in the Control condition.

There was no evidence that k was lower in the Evaluative Judgment condition than in the Control condition, OR = .997, 95% CI [-0.460, .454], $t(1161)$ = -.013, p = .99. See Appendix C for tests of the effect of evaluative judgment on the other parameters.

2.4.2.2 Exploratory Analysis

More granular analysis of participants' behaviour shows that many participants in the control-first ordering moved towards a 'larger-later no matter what' approach in the Evaluative Judgment condition. This pattern was not visible in the opposite order. See Figure 2.3.

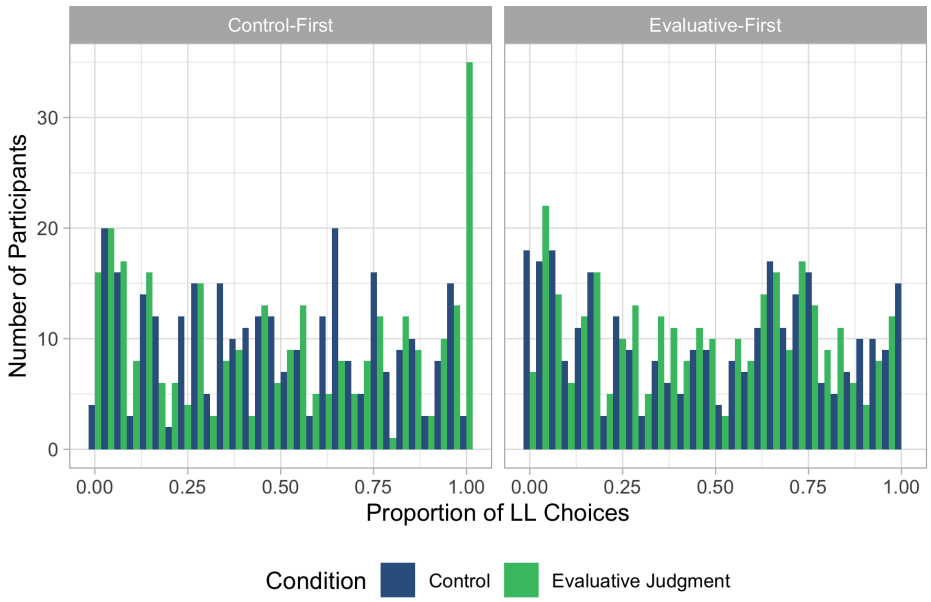

FIGURE 2.3: The proportion of larger-later choices per participant in each condition, partitioned by the ordering of the blocks

Different choices in the choice set were differentially affected by participants' evaluative judgment, mostly based on the reward magnitude. Namely, in the control-first ordering, participants tended to prefer larger-later options, especially in cases where the absolute and relative difference between the smaller-sooner and larger-later was less. See Appendix D for a figure depicting the effect of the manipulation on different choices in the choice set.

Further, participants' on aggregate seemed to be fit better by the discounted utility model (AIC = -37752), than the hyperbolic discounting model (AIC = -42940). We did not find any clear differences in model fits between the conditions (see Appendix E), $F = .67$, $p = .41$. This is surprising because fewer preference reversals ought to predict better fit of the discounted utility model for the Evaluative Judgment condition compared to the Control. This lack of correspondence between the results and the relative goodness-of-fit of the discounted utility model and the hyperbolic discounting model prompted us to conduct a parameter recoverability analysis (see e.g., Ballard et al., 2023).

For this analysis, we took participants' discounting parameters that we estimated from their choices in the experiment and used them to predict out-of-sample choices on the same item set. To do this, we estimate the discounted value of each option using Equation 2.1 and Equation 2.2, make a probabilistic choice prediction using Equation 2.4, where the binary choice probability is given by a Bernoulli trial.

Using these new fictitious choices, we estimated a new set of discounting parameters using the same procedure as before (the recovered parameters). Finally, we compare the recovered parameters from the fictitious choices to the parameters derived from the experimental observations and see how well they correspond, a measure of *parameter recoverability.*

We found that there was only a modest degree of correspondence between the recovered parameters and the initially estimated parameters. See Appendix F for a figure showing the recoverability of the k parameter for both delay discounting models.

2.4.3 Discussion

We aimed to determine whether and how directing deliberation toward evaluative judgment can change behaviour. The results from this experiment show that making decisions by focusing on evaluative judgment can decrease common delay preference reversals and, more tentatively, that people may make more future-oriented decisions. Although the latter difference was due to a small subset of individuals changing their preferences considerably.

We did not find the predicted attenuating effect of using evaluative judgment on the prevalence of failures of dominance. Rather, we observed an effect in the opposite

direction. However, the proportion of failures of dominance was low in both conditions, and the effect size was small—an increase from .037 to .062 failures of dominance per person out of a maximum of 12.

We also found no evidence that the discounted utility model fit participants' choices better in the Evaluative Judgment condition compared to the Control, nor that evaluative judgment affected estimated model parameters. We discuss these findings in more detail below.

2.5 General Discussion

We have tested whether specific ways of guiding the decision-making process can change people's intertemporal choices. We examined whether eliciting meta-preferences would make people more future-oriented in choices with small stakes, and whether this effect would carry over into subsequent intertemporal choices. In addition, we tested whether evaluative judgment could do the same for larger stakes, reduce the common delay effect and failures of dominance, as well as change discounting parameters and model fits. We found that when people were prompted to make decisions by invoking preferences about themselves, their decisions were more future-oriented, although not significantly so when they used evaluative judgment. We also found that people made significantly fewer common delay preference reversals when making their choices using evaluative judgment.

In the first study, participants selected more larger-later choices when they were instructed to make decisions with their meta-preferences in mind. This effect was not seen in the condition where participants were asked to *imagine* what they would do, suggesting that this effect was not due to abstract reasoning, nor due to a difference in incentives. This finding relates to Frankfurt (1988) his idea of second-order volitions, as we saw that people's preferences about their own actions in this context are different from what they actually do (which he suggests is an important part of being a person). It suggests that casting decisions with small stakes in a meta-preferential frame might increase the perceived significance of the reward. This interpretation is in line with findings that providing reasons for one's decision can attenuate the magnitude effect in

intertemporal choice through decreasing delay discounting for smaller rewards (Bulley & Schacter, 2020).

We found no evidence, however, that decisions by people with a high borderline personality trait score were affected differently by meta-preference elicitation than others in the sample. It may be that multiple factors are influencing these individuals in opposite directions: borderline personality disorder is characterised by a lack of trust (see e.g., Hanegraaf et al., 2021; Michael & van Baal, 2021), which has been known to cause people to discount the future more (Farah & Hook, 2017), but because of this, they should also suffer less from ceiling effects, giving them room to become more future-oriented compared to people who are not high in borderline personality disorder traits. It is worth noting that the manipulation used in this experiment resembles Acceptance and Commitment Therapy (ACT; Gloster et al., 2020; Hayes et al., 2006). In this treatment, acting in accordance with your values (which are conceptualised in a similar way as meta-preferences are here) is a core tenet. The evidence we found here, that people high in borderline personality disorder traits are amenable to changes in decision-making when attending to their meta-preferences, aligns with evidence that ACT may be an effective treatment for borderline personality disorder (Morgan & Aljabari, 2019). Assessing whether this type of directed deliberation is an important part of why ACT appears to be effective could be a fruitful avenue for future research.

In the second study, we found that participants were less likely to commit *common delay preference reversals* when their deliberation was directed towards evaluative judgment. That is, there was an attenuation of the common delay effect, showing that participants valued immediate smaller-sooner options less strongly over those same smaller-sooner options if they would arrive in the future. This is an important result because such preference reversals in real life are welfare-reducing (Loewenstein & Thaler, 1989; Thaler, 1981). Combined with the results from the first experiment, these are promising findings that highlight directed deliberation as a candidate for improving people's decision-making in contexts where welfare-reducing behaviour, like preference reversals, is common. As such, it is also evidence that directed deliberation can make people more economically rational in the intertemporal choice context.

We found an effect on dominated choices in the opposite direction of what we predicted. Namely, participants committed more failures of dominance (i.e., picked an option that

was worse in both value and time of receipt) when they were using evaluative judgment, although the effect size was very small and the prevalence in either condition was less than one per cent of the maximum number of failures of dominance. Therefore, it appears we may not have to be concerned by this finding.

We found insufficient evidence to conclude that participants' choices were more future-oriented when they relied on evaluative judgment than when they made choices as they normally would, although we observed a considerable quantitative difference. There are a number of ways to interpret this, one of which is that our manipulation was not effective (but see above on the preference reversal effect), another is that people generally do not become more future-oriented when they deliberate more because to them this does not appear to be 'wiser', or more adaptive. If this interpretation is right—that people do not become more future-oriented when they deliberate more—then it would put pressure on the notion that a lack of future-oriented time preferences is equivalent to or even indicative of impulsivity, as is often suggested in the literature. This is because impulsivity is characterised by a lack of deliberation, and so if more deliberation does not make one more future-oriented, then it must be that the degree of future-orientation observed in intertemporal choice is not measuring impulsivity. Such an interpretation is in agreement with our arguments in a previous work (van Baal et al., 2022), where we argue that time preferences should be dissociated from impulsivity and self-control.

Similarly, there was no evidence that there was an improvement in the fit of participants' choices with the discounted utility model (Samuelson, 1937) over the hyperbolic discounting model (Mazur, 1987) when they used evaluative judgment. Hyperbolic discounting is not utility maximising according to traditional economic theory (Ainslie, 1975), but more recently several researchers have suggested that hyperbolic discounting can be explained or even deemed rational if we allow and account for variable time preferences (He et al., 2019), uncertainty that increases linearly in time (Gabaix & Laibson, 2017), or uncertainty about ones own future discount rates (Farmer & Geanakoplos, 2009). That is not to say that hyperbolic discounting is the best model of participants' choices in either condition, but it might indicate that participants are not deviating from what looks like hyperbolic discounting when they rely on evaluative judgment because they already deem that they are acting adaptively.

It is, however, somewhat surprising that we see a sizeable reduction in preference reversals by eliciting evaluative judgment, but not a move away from hyperbolic discounting and towards the discounted utility model in relative model goodness-of-fit. One of the main reasons for developing the hyperbolic discounting model was that it could account for a 'behavioural regularity' that could not be explained by the discounted utility model—the common delay effect (Ainslie, 1975; Read, 2004)—so one would expect there to be correspondence between comparative goodness-of-fit between these models, and common delay preference reversals. However, as can be seen in Appendix F, the parameter estimates are not very reliable. In the parameter recovery analysis, we saw that if a decision-maker generates choices with a high k, it is unlikely we will be able to retrieve this information when we estimate the parameters using maximum likelihood estimation. Thus, the lack of any effects on parameter estimates or model fits is to be expected.

2.5.1 Limitations and Future Directions

The sample in our studies is not representative of the population in the United Kingdom; we over-sampled young people and did not screen for ethnicity. This means that the diversity of the sample likely does not reflect the diversity of the population, which impacts the external validity of our current experiment as there are many cultural aspects to decision-making (Yates & de Oliveira, 2016). The ways of deliberating decisions that are most helpful may also vary in different parts of the world. Therefore we have to be cautious in applying these findings in practice; more research on the effects of directed deliberation in various cultural contexts is necessary.

Another issue is that the effects we observed in the intertemporal choice context might have little bearing on control decisions people make in daily life. Even though domain-general time preferences are likely to play a role in many decisions, the complexity of people's lives usually dwarfs the predictive power of typical time preference measures obtained through intertemporal choice (Cohen et al., 2020). We view the results from this study as a promising sign for the future of deliberation research. Expanding research on the impact of different deliberation and contemplation processes to different domains, such as multi-attribute choice, risky choice, patch foraging, and especially to field research will elucidate how useful these decision-making methods are.

In these two experiments, we used two different ways to direct deliberation, which may have varying influences on decision-making. Therefore, the differences in results we saw between the two experiments may have been due to the conceptual differences between the two manipulations. Considering that the results of the experiments were promising in various aspects, we think a consolidated experiment to reproduce the findings is warranted and necessary.

Further, given the results of the parameter recovery analysis, it is clear that we were unable to retrieve reliable discounting parameters from this choice set. Therefore, future studies need to construct a choice set with recoverability in mind, and ideally conduct simulations beforehand (for a discussion of the importance of this, see Ballard et al., 2023).

Finally, it is difficult to control for the decision process people use in each of our conditions. It is possible, for example, that many people in the control group already use a meta-preferential mindset or assess the options akin to the process we prompt others to do in the evaluative judgment conditions. The fact that we observed effects of the manipulation indicates that this might not be too influential an issue. However, replicating the effects using the current approach supplemented with manipulation checks could further confirm the internal validity of the effect we observed. More generally, we believe further research on individual differences in deliberation and outcomes in decision-making may also be important to our understanding of human intertemporal choice processes.

2.6 Conclusion

We conclude that directed deliberation methods are able to influence intertemporal choices in different ways. This is a promising area for future research in answering questions of how to best make consequential decisions. Considering the success of these manipulations in changing various decision-making metrics, we believe a more systematic approach to research on guiding choice deliberation is warranted.

Determining reliable ways of improving people's decisions would greatly help the public, organisations, and policymakers. Our results suggest that using meta-preferences or

focusing on evaluative judgment when faced with an important intertemporal choice might be a good place to start.

2.7 Declarations

2.7.1 Funding Statement

This research was funded by the Monash Warwick Alliance Accelerator Fund. The funder was not involved in the research process.

2.7.2 Conflict of Interest

The authors declare no conflict of interest.

2.7.3 Ethics Statement

This research was approved by the Monash University Human Research Ethics Committee, ID: 23485.

2.8 References

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

Amasino, D. R., Sullivan, N. J., Kranton, R. E., & Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour*, *3*(4), 383–392. https://doi.org/10.1038/s41562-019-0537-2

Ballard, T., Luckman, A., & Konstantinidis, E. (2023). A systematic investigation into the reliability of inter-temporal choice model parameters. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-022-02241-7

Barker, V., Romaniuk, L., Cardinal, R., Pope, M., Nicol, K., & Hall, J. (2015). Impulsivity in borderline personality disorder. *Psychological medicine*, *45*(9), 1955–1964. https://doi.org/10.1017/S0033291714003079

Becker, G. M., Degroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, *8*(1), 41–55. https://doi.org/10.1002/bs.3830080106

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232. https://doi.org/10.1002/bs.3830090304

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, *9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

Bulley, A., & Schacter, D. L. (2020). Deliberating trade-offs with the future. *Nature Human Behaviour*, *4*(3), 238–247. https://doi.org/10.1038/s41562-020-0834-9

Cohen, J., Ericson, K. M., Laibson, D., & White, J. M. (2020). Measuring Time Preferences. *Journal of Economic Literature*, *58*(2), 299–347. https://doi.org/10.1257/jel.20191074

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, *143*(4), 1489. https://doi.org/10.1037/a0035976

DeShong, H. L., Mullins-Sweatt, S. N., Miller, J. D., Widiger, T. A., & Lynam, D. R. (2016). Development of a Short Form of the Five-Factor Borderline Inventory. *Assessment*, *23*(3), 342–352. https://doi.org/10.1177/1073191115581475

Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J. E., & Treibich, T. (2020). Rational Heuristics? Expectations and Behaviors in Evolving Economies with Heterogeneous Interacting Agents. *Economic Inquiry*, *58*(3), 1487–1516. https://doi.org/10.1111/ecin.12897

Ericson, K., White, J. M., Laibson, D., & Cohen, J. (2015). *Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better than Delay Discounting* (tech. rep. w20948). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w20948

Farah, M. J., & Hook, C. J. (2017). Trust and the poverty trap. *Proceedings of the National Academy of Sciences*, *114*(21), 5327–5329. https://doi.org/10.1073/pnas.1704798114

Farmer, J. D., & Geanakoplos, J. (2009). Hyperbolic discounting is rational: Valuing the far future with uncertain discount rates.

Frankfurt, H. G. (1988). Freedom of the Will and the Concept of a Person. In M. F. Goodman (Ed.), *What Is a Person?* (pp. 127–144). Humana Press. https://doi.org/10.1007/978-1-4612-3950-5_6

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, *40*(2), 351–401.

Gabaix, X., & Laibson, D. (2017). *Myopia and discounting* (Report). National bureau of economic research.

Gigerenzer, G. (2016). Towards a Rational Theory of Heuristics. In R. Frantz & L. Marsh (Eds.), *Minds, Models and Milieux: Commemorating the Centennial of the Birth of Herbert Simon* (pp. 34–59). Palgrave Macmillan UK. https://doi.org/10.1057/9781137442505_3

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gloster, A. T., Walder, N., Levin, M. E., Twohig, M. P., & Karekla, M. (2020). The empirical status of acceptance and commitment therapy: A review of meta-analyses. *Journal of Contextual Behavioral Science*, *18*, 181–192. https://doi.org/10.1016/j.jcbs.2020.09.009

Green, L., Fristoe, N., & Myerson, J. (1994). Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin & Review*, *1*(3), 383–389. https://doi.org/10.3758/BF03213979

Hanegraaf, L., van Baal, S. T., Hohwy, J., & Verdejo-Garcia, A. (2021). A systematic review and meta-analysis of 'Systems for Social Processes' in borderline personality and substance use disorders. *Neuroscience & Biobehavioral Reviews*, *127*, 572–592. https://doi.org/10.1016/j.neubiorev.2021.04.013

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006

He, L., Golman, R., & Bhatia, S. (2019). Variable time preference. *Cognitive psychology*, *111*, 53–79. https://doi.org/10.1016/j.cogpsych.2019.03.003

He, L., Zhao, W. J., & Bhatia, S. (2022). An ontology of decision models. *Psychological Review*, *129*(1), 49–72. https://doi.org/10.1037/rev0000231

Holton, R. (2009). *Willing, wanting, waiting*. Oxford University Press on Demand.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kirby, K. N., & Herrnstein, R. (1995). Preference Reversals Due to Myopic Discounting of Delayed Reward. *Psychological Science*, *6*(2), 83–89. https://doi.org/10.1111/j.1467-9280.1995.tb00311.x

Kraut, R. (2001). Aristotle's ethics.

Lawrence, K., Allen, J., & Chanen, A. (2010). Impulsivity in Borderline Personality Disorder: Reward-Based Decision-Making and its Relationship to Emotional Distress. *Journal of personality disorders*, *24*, 786–99. https://doi.org/10.1521/pedi.2010.24.6.785

Lempert, K. M., & Phelps, E. A. (2016). The Malleability of Intertemporal Choice. *Trends in Cognitive Sciences*, *20*(1), 64–74. https://doi.org/10.1016/j.tics.2015.09.005

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, *107*(2), 573–597. https://doi.org/10.2307/2118482

Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal choice. *Journal of Economic perspectives*, *3*(4), 181–193. https://doi.org/10.1257/jep.3.4.181

Magen, E., Dweck, C. S., & Gross, J. J. (2008). The hidden-zero effect: Representing a single choice as an extended sequence reduces impulsive choice. *Psychol Sci*, *19*(7), 648–9. https://doi.org/10.1111/j.1467-9280.2008.02137.x

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Quantitative analyses of behavior*, *5*, 55–73.

Michael, J., & van Baal, S. T. (2021). Key themes in recent research on interpersonal functioning in borderline personality disorder. *Current Opinion in Psychiatry*, *34*(1), 44–47. https://doi.org/10.1097/YCO.0000000000000649

Morgan, T. A., & Aljabari, R. (2019). Using Acceptance and Commitment Therapy in the Treatment of Borderline Personality Disorder: Rationale, Preliminary Evidence, and Future Directions. *Current Treatment Options in Psychiatry*, *6*(4), 271–283. https://doi.org/10.1007/s40501-019-00185-9

Naudé, G. P., Kaplan, B. A., Reed, D. D., Henley, A. J., & DiGennaro Reed, F. D. (2018). Temporal framing and the hidden-zero effect: Rate-dependent outcomes on delay discounting. *Journal of the experimental analysis of behavior*, *109*(3), 506–519.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of risk and uncertainty*, *23*(1), 5–32.

Read, D. (2004). Intertemporal choice. *Blackwell handbook of judgment and decision making*, 424–443.

Samuelson, P. A. (1937). A note on measurement of utility. *The review of economic studies*, *4*(2), 155–161.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). Afex: Analysis of factorial experiments. *R package version 0.13–145*.

Solnick, J. V., Kannenberg, C. H., Eckerman, D. A., & Waller, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation*, *11*(1), 61–77. https://doi.org/10.1016/0023-9690(80)90021-1

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The review of economic studies*, *23*(3), 165–180.

Thaler, R. H. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207. https://doi.org/10.1016/0165-1765(81)90067-7

van Baal, S. T., Walasek, L., Verdejo-Garcia, A., & Hohwy, J. (2022). Impulsivity and Self-Control as Timeless Concepts. https://doi.org/10.17605/OSF.IO/9EVFT

Vincent, B. T., & Stewart, N. (2020). The case of muddled units in temporal discounting. *Cognition*, *198*, 104203. https://doi.org/10.1016/j.cognition.2020.104203

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60*(2), 181–192. https://doi.org/10.1037/0022-3514.60.2.181

Yates, J. F., & de Oliveira, S. (2016). Culture and decision making. *Organizational Behavior and Human Decision Processes*, *136*, 106–118. https://doi.org/10.1016/j.obhdp.2016.05.003

2.9 Appendices

2.9.1 Appendix A

TABLE S1: The choice pairs for Experiment 1. Each smaller-sooner option (SS) is presented together with a larger-later option (LL), with a delay shown in days (t1).

Set	**SS**	**LL**	**t1**	**Daily Rate**
1	£4.00	£4.50	42	0.30%
2	£4.25	£4.75	42	0.28%
3	£4.50	£5.00	42	0.26%
2	£5.00	£5.50	35	0.29%
1	£5.25	£5.75	35	0.27%
3	£5.50	£6.00	35	0.26%
3	£6.00	£6.50	28	0.30%
2	£6.25	£6.75	28	0.29%
1	£6.50	£7.00	28	0.27%
3	£7.00	£7.50	21	0.34%
1	£7.25	£7.75	21	0.33%
2	£7.50	£8.00	21	0.32%
3	£3.00	£5.00	42	1.59%
2	£3.25	£5.25	42	1.47%
1	£3.50	£5.50	42	1.36%
3	£4.00	£6.00	35	1.43%
1	£4.25	£6.25	35	1.34%
2	£4.50	£6.50	35	1.27%
1	£5.00	£7.00	28	1.43%
2	£5.25	£7.25	28	1.36%
3	£5.50	£7.50	28	1.30%
2	£6.00	£8.00	21	1.59%
1	£6.25	£8.25	21	1.52%
3	£6.50	£8.50	21	1.47%
2	£2.00	£5.50	42	4.17%
1	£2.25	£5.75	42	3.70%
3	£2.50	£6.00	42	3.33%
1	£3.00	£6.50	35	3.33%
2	£3.25	£6.75	35	3.08%
3	£3.50	£7.00	35	2.86%
2	£4.00	£7.50	28	3.13%
3	£4.25	£7.75	28	2.94%
1	£4.50	£8.00	28	2.78%
1	£5.00	£8.50	21	3.33%
3	£5.25	£8.75	21	3.17%
2	£5.50	£9.00	21	3.03%

2.9.2 Appendix B

TABLE S2: The choice pairs for Experiment 1. Each set of 12 choices is assigned to a block, which is then randomly assigned a condition. Each smaller-sooner option (SS) is presented together with a larger-later option (LL), with delays (t0 and t1) shown in months.

SS	t0	LL	t1	Interest
100	0	93	3	−25%
100	0	106	3	25%
100	0	115	3	75%
100	0	122	3	125%
100	0	129	3	175%
100	0	87	6	25%
100	0	112	6	25%
100	0	132	6	75%
100	0	150	6	125%
100	0	166	6	175%
100	0	81	9	−25%
100	0	118	9	25%
100	0	152	9	75%
100	0	184	9	125%
100	0	214	9	175%
100	3	93	6	−25%
100	3	106	6	25%
100	3	115	6	75%
100	3	122	6	125%
100	3	129	6	175%
100	3	87	9	−25%
100	3	112	9	25%
100	3	132	9	75%
100	3	150	9	125%
100	3	166	9	175%
100	3	81	12	−25%
100	3	118	12	25%
100	3	152	12	75%
100	3	184	12	125%
100	3	214	12	175%
100	6	93	9	−25%
100	6	106	9	25%
100	6	115	9	75%
100	6	122	9	125%
100	6	129	9	175%
100	6	87	12	−25%
100	6	112	12	25%
100	6	132	12	75%
100	6	150	12	125%

Continued on next page

Table S2 – continued

SS	t0	LL	t1	Interest Rate
100	6	166	12	175%
100	6	81	15	−25%
100	6	118	15	25%
100	6	152	15	75%
100	6	184	15	125%
100	6	214	15	175%
100	9	93	12	−25%
100	9	106	12	25%
100	9	115	12	75%
100	9	122	12	125%
100	9	129	12	175%
100	9	87	15	−25%
100	9	112	15	25%
100	9	132	15	75%
100	9	150	15	125%
100	9	166	15	175%
100	9	81	18	−25%
100	9	118	18	25%
100	9	152	18	75%
100	9	184	18	125%
100	9	214	18	175%

2.9.3 Appendix C

Aside from testing for an effect of the manipulation in Experiment 2 on the discounted utility model parameter k, we report tests using the same fractional logit approach for the other parameters discussed in this paper, namely, hyperbolic discounting parameter k, and the temperature parameters for both models, ϑ.

There was no evidence that k_{hyp}, ϑ_{exp}, or ϑ_{hyp} were different in the Evaluative judgment condition compared to the Control condition, respectively, ORs: 1.05, 1.04, 1.01; 95% CIs [.682, 1.418], [.649, 1.68], [.71, 1.43]; ts(1195) = .278, .177, .046; all ps $> .75$. See Table S3 the means and standard deviations of the maximum likelihood estimates for each parameter.

Table S3: Means (SDs) of parameter estimates by condition for Experiment 2.

Parameter	Control	Evaluative Judgment
k_{exp}	0.068 (0.070)	0.068 (0.054)
ϑ_{exp}	0.060 (0.074)	0.062 (0.072)
k_{hyp}	0.115 (0.116)	0.121 (0.118)
ϑ_{hyp}	0.118 (0.074)	0.119 (0.072)

2.9.4 Appendix D

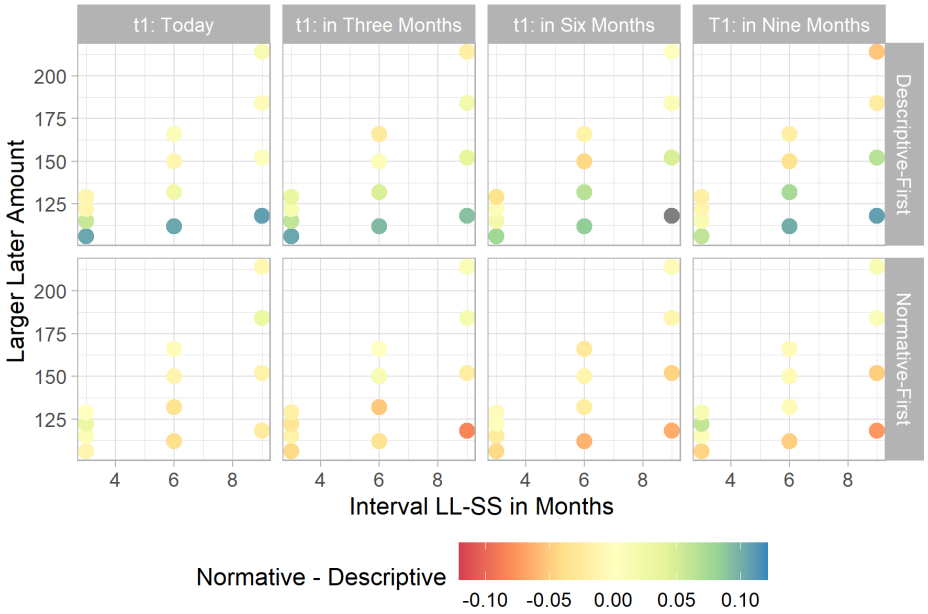

FIGURE S1: The proportion of larger-later choices per item in each condition, partitioned by the ordering of the blocks. Each panel shows a combination of the block ordering and the time at which the smaller-sooner reward arrives. The y-axis shows the larger-later reward amount, and the x-axis shows the interval between the smaller-sooner and larger-later. Points that are green (red) imply more (less) patient responding for that item in the evaluative judgment condition.

2.9.5 Appendix E

In comparing the model fits of the discounted utility model versus the hyperbolic discounting model, we found no discernible effect of condition on model fit. For a depiction of difference in model fit by condition, partitioned by group, see Figure S2.

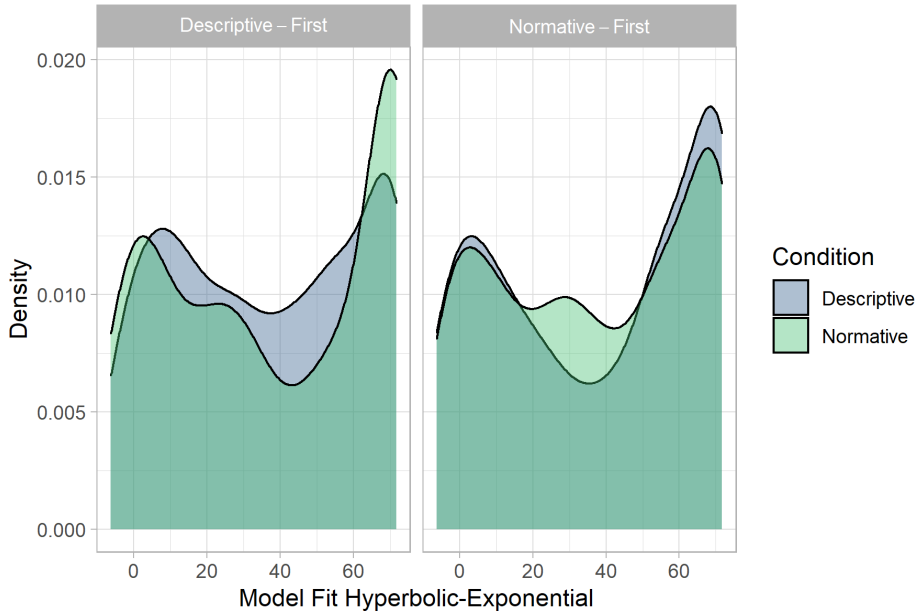

FIGURE S2: The difference in log-likelihood between the hyperbolic and exponential model. Higher values on the x-axis signify better hyperbolic model fits.

2.9.6 Appendix F

We performed a parameter recovery analysis to see how reliable the parameter estimates from our experiment were. We found that the parameters were not recovered well, especially for the hyperbolic model. See Figure S3.

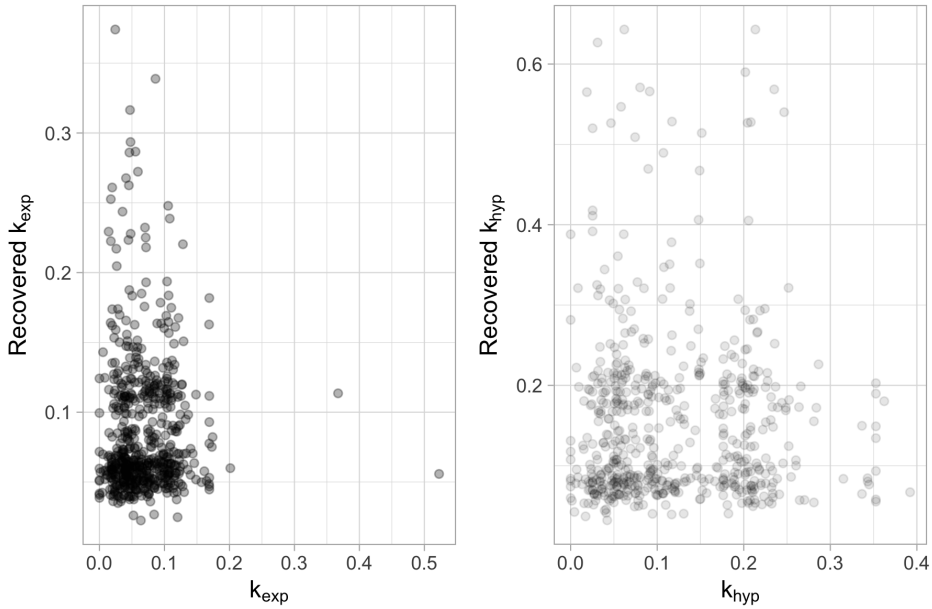

FIGURE S3: Recovered parameters plotted against generating parameters. For each panel, the x-axis shows the generating parameter, and the y-axis shows the recovered parameter

Chapter 3

The Relationship Between State Impulsivity and Urges

3.1 Linking Text for Chapter 3

Thus far, I have shown how impulsivity, self-control, and time preferences come apart and in what ways they influence each other. In addition, I have shown that deliberating about intertemporal choices in ways that emphasise higher-order preferences can systematically change how people construct preferences in ways that may be predictable. Specifically, when individuals are prompted to consider a decision in light of their evaluative judgment (i.e., what they think is best) or the preferences they have for their own behaviour (i.e., what they would like themselves to do), they may become more *future-oriented* for small stakes (i.e., more willing to wait) and they tend to become more *dynamically consistent* for large stakes (i.e., less likely to revert their preferences once the smaller-sooner becomes more immediate).

The result that, for large stakes, people do not seem to judge being more patient in intertemporal choice best, provides some initial evidence that people do not perceive their time preferences as being sub-optimal (i.e., there does not seem to be a difference between first-order and second-order preferences in this respect). This agrees with findings by Bulley et al. (2021) that people are not more confident that they made the right choice when they pick larger-later options. We could interpret this as evidence that it might not be a worthwhile general policy to nudge people towards more patient

behaviour. This is because proper nudging presupposes that the outcome of the intervention is implicitly preferred by the decision-maker (Thaler & Sunstein, 2009). In our study, we found no evidence that people deem it best to be more patient for decisions with larger stakes, at least in the choice set we used.

The results from Chapter 2 necessitate an update of the descriptive model of how impulsivity, self-control, and time preferences influence behaviour that was introduced in Chapter 1. Namely we saw that deliberation can change how rewards are valued, depending on characteristics of the reward (e.g., the magnitude). Though more research in this space is necessary, the most plausible mechanism for change is that deliberation highlights different "rewards" that can be achieved through taking one option or the other. See Figure 3.1 for a depiction of this updated model.

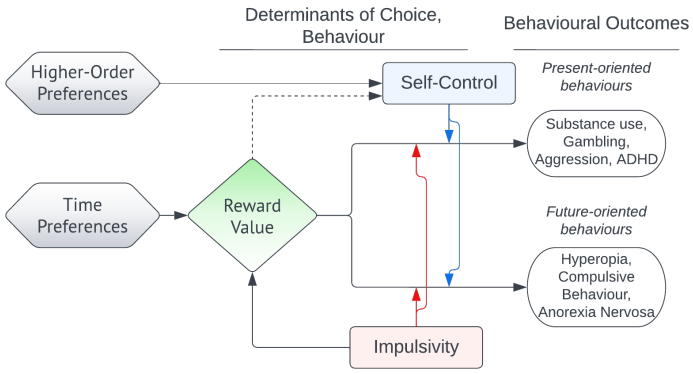

FIGURE 3.1: Updated conceptual model. In Chapter 2, I showed that deliberation can change the value of rewards. Hence I have now added a causal arrow to show that reward value can be affected by impulsivity (i.e., less deliberation).

Naturally, people still sometimes make decisions that they come to regret; humans are susceptible to state changes that cause them to make decisions that they would not otherwise make. However, much research is devoted to identifying impulsivity and self-control as *traits* and trying to remedy any perceived issues through education or therapy, essentially in an attempt to change the individual. In many cases, there is a more promising and efficient route to changing behaviour: taking advantage of or changing people's states. A key determinant of marketing success, for instance, is to get to a consumer at the right point in time or to get them to the right state such that they will buy a product (e.g., bliss points in the food industry; Moss, 2013). To understand better how we can help people deal with their motivational drives, it is important to see how people experience and deal with urges when they are in different mental states.

From this point onward in the thesis, I will be moving away from theoretical and fundamental decision-making research, and focusing on situational and applied research. I will start by examining how people's urges and self-control are affected by important common drivers of behaviour: hunger and state impulsivity. I will examine whether and how these states impact the intensity of different types of urges and our ability to control them.

This Chapter will contain an exploration of what makes people do things that have broader public health implications, and in the process, it will give us a new perspective on the manifestations of impulsivity and self-control—one that is much closer to participants' daily lives than an intertemporal choice between monetary amounts.

Note that the purpose of the research that follows is not to blame individuals for "failing" to control their urges, but merely to explore when these urges are strongest or most difficult to control, which is important information for effectively designing policies that protect public health (e.g., concerning substance use disorders, obesity), and individual finance (e.g., concerning gambling).

Addictive Behaviors 133 (2022) 107381

Contents lists available at ScienceDirect

Addictive Behaviors

journal homepage: www.elsevier.com/locate/addictbeh

State impulsivity amplifies urges without diminishing self-control

Simon Thomas van Baal [a,b,*], Neda Moskovsky [c], Jakob Hohwy [a,d], Antonio Verdejo-García [c]

[a] Monash University, Cognition and Philosophy Lab, Department of Philosophy, Australia
[b] University of Warwick, Department of Psychology, United Kingdom
[c] Monash University, Turner Institute for Brain and Mental Health, School of Psychological Sciences, Australia
[d] Monash Centre for Consciousness & Contemplative Studies, Australia

ARTICLE INFO

Keywords:
Impulsivity
Hunger
Ecological momentary assessment
Urges
Self-control
Craving

ABSTRACT

A disproportionate amount of research on impulsivity has focused on trait-related aspects rather than state fluctuations. As a result, the relationship between state impulsivity and moment-to-moment behaviour is unclear. Impulsivity is assumed to negatively affect self-control, but an alternative explanation, yet to be tested, could be that changes in state impulsivity and its homeostatic drivers influence the intensity of urges. We tested whether state impulsivity and hunger affected behaviour through a dual-process model, affecting both the experience of various urges, and self-control, using a smartphone-based experience sampling approach. We found that state impulsivity is associated with stronger urges, but we found no evidence of an association with diminished self-control. Being hungry amplifies urges across different types of urges, and both hunger and late hours are negatively related to the likelihood of controlling urges. These findings imply that the influence of hunger is not limited to the food domain, and provide new insight into the role of state impulsivity in daily life.

1. Introduction

Impulsivity, the predisposition to trigger rapid responses without sufficient forethought, is associated with numerous addictive behaviours (Berg et al., 2015; Verdejo-García et al., 2008). Impulsivity research has steadily grown in the last two decades (Sharma et al., 2014; Strickland & Johnson, 2021; Verdejo-García et al., 2008; Whiteside & Lynam, 2001), but advances have focused on trait characteristics and inter-individual differences, with comparatively much less progress on state impulsivity, state-related fluctuations such as those linked to energy input (hunger, satiety) and intra-individual variability.

Trait-focused research typically relies on 'distal' self-reports where individuals report what they typically do, which introduces uncertainty about how those traits are expressed in moment-to-moment behaviour. Understanding how the effects of impulsivity on behaviour unfold over time requires 'proximal' information on people's mental states when these behaviours occur (Curran & Bauer, 2011; Fisher et al., 2018; Kenrick & Funder, 1988).

It thus remains unclear how moment-to-moment impulsive drives influence behaviour. It is often posited that impulsivity increases the tendency toward maladaptive behaviours because it negatively affects self-control (the capacity to override or alter predominant response tendencies in support of the pursuit of long-term goals; Baumeister et al., 2007). This is evidenced, for example, by the inclusion of some version of 'lack of self-control' subscales in many impulsivity scale questionnaires (Gough, 2000; Parker & Bagby, 1997; Whiteside & Lynam, 2001; Wiers et al., 2010). In addition, evidence from cross-sectional studies suggests that people with stronger impulsivity traits experience stronger urges (or cravings; Doran et al., 2007; Papachristou et al., 2012; Yarmush et al., 2016), but little is known about the relationship between state impulsivity and the experience of urges.

To understand the relationship between state impulsivity and behaviour, we turn to dual-systems theories of behaviour. Dual-systems theories conceptualise human behaviour as a conflict between automatic and deliberative modes of behavioural control (Kahneman, 2011; Loewenstein, 1996; Metcalfe & Mischel, 1999). In support of this, both sensitisation of responses towards a range of incentives (i.e., urges), and low inhibitory control have been independently linked to vulnerability and escalation of addictive behaviours (Bechara, 2005; Feltenstein & See, 2008; Goldstein & Volkow, 2002, 2011; Potenza et al., 2003).

For this study, we conceptualise the intensity of urges as being regulated by a 'bottom-up' process, in contrast to self-control as a 'top-down' process, regulating behaviour. Thus, here we investigate whether state impulsivity affects behaviour through a dual-process model: a

* Corresponding author at: Cognition and Philosophy Lab, Department of Philosophy, 20 Chancellors Walk, Monash University, VIC 3800, Australia.
E-mail address: simon.vanbaal1@monash.edu (S.T. van Baal).

https://doi.org/10.1016/j.addbeh.2022.107381
Received 14 February 2022; Received in revised form 19 May 2022; Accepted 26 May 2022
Available online 28 May 2022

bottom-up influence, amplifying urges through up-regulating the stimulus–response system and a top-down influence on the deliberative system, down-regulating goal-directed behaviour, by affecting people's self-control. For example, an individual in an impulsive state might experience a stronger urge to gamble when they see an electronic gambling machine in a pub (bottom-up influence – upregulating the stimulus–response relationship). In addition, they may be less well-equipped than normal to curb urges of a fixed intensity, thus finding it harder to resist engaging in gambling (top-down influence – down-regulating self-control).

Like state impulsivity, hunger is an important common motivational drive and a state fluctuation that is implicated in many of the same behaviours. Hunger is an adaptive motivational state that drives us to eat, restoring homeostatic balance (Saper et al., 2002). There is extensive evidence that hunger enhances the valuation of food (Cameron et al., 2014; Gilbert et al., 2002) although it is unclear if this effect is domain-general, rather than just stimulus-specific, or whether it can trigger impulsive behaviours. More generally, hunger increases impulsive behaviour in non-human animals (Anderberg et al., 2016; Laude et al., 2012; Zheng et al., 2019). There is also emerging evidence in support of a domain-general effect of hunger in humans: people are more likely to choose an impulsive option when gambling hungry (Li et al., 2020), they are more likely to acquire non-food items (Xu et al., 2015), and the effect of hunger on delay discounting (i.e., that hungry people discount delays more) spills over into non-food domains, but the effect size is a quarter of that for food stimuli (Skrynka & Vincent, 2019). However, these experiments were mostly conducted in laboratory settings, and most used measures that do not provide much mechanistic information. In this study, we segregate the two pathways posited by dual system models, to assess how hunger influences the emergence of impulsive behaviour in real-world scenarios.

To test the dual-process models of state impulsivity and hunger, we use smartphone-based ecological momentary assessment (EMA), where we ask participants whether they had an urge to smoke, snack, consume alcohol, gamble, shop, or commit an act of aggression, and whether they were able to control these urges. Controlling such urges normally leads to better long-term goal performance (Finne et al., 2019; Mischel et al., 1989; Rawn & Vohs, 2006); these urges usually pertain to short-term gratification, and succumbing to such temptations can cause disturbances to people's lives in professional, health, and social domains (Bembenutty, 2011; Koomen et al., 2020; Schlam et al., 2013).

The EMA setup allows for unique insights into the influence of the time of day, state impulsivity, and hunger fluctuations on behaviours in different domains (Hofmann et al., 2012), which we can leverage to study the effect of the time of day on self-control—a new angle on the ongoing discussion about the evidence surrounding ego depletion (Friese et al., 2019). EMA studies diminish recall bias (Shiffman et al., 2008), are less laborious and artificial than laboratory studies, and do not rely on beliefs about the self ("What am I typically like?"). To maintain brevity of the surveys in the EMA procedure, we use the Momentary Impulsivity Scale (MIS; Tomko et al., 2014) to assess state impulsivity in each survey. This is a well-validated and reliable measure of state impulsivity.

The main question asked in this study is, therefore, whether state impulsivity and hunger influence behaviour according to a dual-process model, with a bottom-up influence that upregulates stimulus–response relationships, operationalised as the intensity of urges, and a top-down influence that downregulates self-control, operationalised as the likelihood of controlling urges. We will also investigate the differences between the various types of urges people face, and whether this interacts with the effect that hunger has on an urge (to test whether the effect of hunger is domain-general). In addition, we investigate whether there is an effect of the time of day on both dependent variables.

2. Methods

2.1. Participants

Efforts were undertaken to sample participants who would not normally take part in psychological research, to enhance the generalisability of findings. Specifically, we conducted a broad community-based recruitment strategy including flyers posted across different suburbs of Melbourne, Australia and different social media platforms, in addition to the more typical student-based recruitment systems (e.g., SONA).

This study is part of a larger project with a three-hour lab component (procedure here) in addition to the current EMA-based component. The EMA component always occurred after the lab study, so there was no overlap or influence from the lab study. Participants received $75 as remuneration for completing the entire study.

Based on the main aim (i.e., learning more about the context-evoked fluctuations in the intensity of urges) and the effect size of these fluctuations in previous data (Cohen's $d = 0.5$) (Verdejo-Garcia et al., 2015) we required 45 participants to test our hypothesis with 80% power and an alpha level of 0.05.

Inclusion criteria were based on requirements for the overarching project, in which participants completed, in addition to the EMA surveys, nutritional manipulations and a cognitive test battery. Thus, to be included, participants were required to have normal or corrected-to-normal vision, be fluent in English, have no food allergies or food intolerances that would impact the food provided in the satiety manipulation, no history of head trauma (e.g., traumatic brain injury), neurological (e.g., epilepsy, Parkinson's disease) or metabolic impairments (e.g., diabetes, indicated by blood glucose tests), and no current mental health conditions (i.e. psychosis, depression, substance use and eating disorders), indicated by screening interviews based on Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria (American Psychiatric Association, 2013). Exclusion criteria included cannulation contraindications (i.e., HIV, Hepatitis A, B or C diagnosis, low or high blood pressure, blood-thinning medication) because the experimental protocol for the other facets of the project included blood-drawing.

2.2. Apparatus

This study used an online EMA paradigm, administering the surveys through SEMA3 (Koval, 2019). SEMA3 uses an app that can be downloaded by participants on their phones. To complete their surveys, participants must respond to a notification on their phone, prompting them to open the app. Upon opening the app, SEMA3 would commence the survey and the participant would be able to respond to questions by using a horizontal slider and check-boxes.

2.3. Measures

We used purpose-built surveys to assess the presence, intensity and ability to control 6 different urges: snacking, drinking alcohol, gambling, shopping, smoking, and committing an act of aggression (operationalised as hitting something). Participants responded to the question of whether they were experiencing each type of urge listed above using a slider [0,10], where 0 indicated no urge, 1 indicated a very weak urge, and 10 indicated a very strong urge; we call the responses on this scale 'urge intensity'. After each question, participants were also asked whether they were hungry (Yes, No), and whether they were able to control the urge (Yes, No). The hunger question had a binary response mode because we wanted to make the measurements computationally easy for participants, and because questions about hunger are often posed in this binary manner colloquially. The urge control question also has a binary response mode because reporting partially controlling an urge, while possible, likely feels less natural than

indicating whether the urge was controlled or not.

In addition, participants completed the Momentary Impulsivity Scale (MIS) to measure state impulsivity (Tomko et al., 2014). The MIS is a well-validated measure of state-impulsivity for use in EMA setups, and correlates with the Barratt Impulsiveness Scale (BIS-11) total score (r = 0.44) and with two subscales of the Impulsive Behaviour Scale (UPPS) (Urgency, $r = 0.45$; Lack of Perseverance, $r = 0.35$). These subscales are particularly relevant in light of our interest in impulsive behaviour and self-control. Participants were asked to indicate how much each statement described their experience since the last completed prompt using a 5-point Likert scale (1 = very slightly or not at all; 2 = a little; 3 = moderately; 4 = quite a bit; 5 = extremely). The 4 statements were "I said things without thinking", "I spent more money than I meant to", "I have felt impatient", "I made a 'spur of the moment' decision".

2.4. Procedure

The experiment was conducted between 6 April 2018 to 18 December 2019, and the EMA protocol took 7 days to 10 days per participant, depending on their survey completion percentage. Informed consent was obtained from all participants and the project was approved by the Monash University Human Research Ethics Committee (Project ID: 11999).

To assess participants' eligibility, they were first sent a survey through Qualtrics, in which they entered basic demographic information, filled out several trait questionnaires, and completed an intertemporal choice task (which will not be used for this study). Afterwards, they were asked to come to the university twice where they would complete the Cognitive Impulsivity Suite (Verdejo-Garcia et al., 2021) in either a fasted or a sated state and in between these sessions they would complete the EMA experiment.

Participants received four daily surveys that were available for 120 min. The first survey would arrive between 9.30 and 10am, the second between 1 pm and 1.30 pm, the third between 4.30 pm and 5 pm, and the last between 8 pm and 8.30 pm. In these surveys, participants were asked questions about their urges and actions since the last survey. They were asked whether since the last survey they had felt an urge to snack, drink alcohol, gamble, shop, smoke, or commit an act of aggression (hit something).

2.5. Analysis

The main aims of the study were to investigate how state impulsivity and hunger influence the intensity of urges and self-control, and whether hunger affects urges in non-food domains. Accordingly, the main analysis was two-tiered: the first part of the analysis concerned the effects of hunger and state impulsivity on the intensity of urges, and the second part concerns the effect of state impulsivity and hunger on the likelihood of controlling an urge. We operationalise a bottom-up effect of hunger and state impulsivity as an effect on the intensity of urges, while we operationalise a top-down effect on self-control as an effect on the probability of controlling an urge. We also report analyses on the effects of the time of day on both dependent variables.

For completeness, we also provide tables comparing the models used in the current work with models where all interactions between variables of interest are included in the Supplementary Materials.

A linear mixed model was used for the analysis concerning perceived urge intensity during the EMA study. The predictors were the total score on the Momentary Impulsivity Scale (MIS; a continuous variable), the type of urge participants had, whether participants were hungry (indicated as 'yes' or 'no'), the time of day (rounded to the hour), and as control variables, participants' age, sex, and the day of the week. We included an interaction term for the type of urge and whether the participant was hungry (e.g., the urge to drink alcohol while hungry). The model also included the participant as a random intercept and the type of urge as a random slope.

For the analysis on the probability of controlling urges, we used a binomial generalised linear mixed model (GLMM) because the dependent variable is binary. The predictors were the responses on the Momentary Impulsivity Scale (MIS), the type of urge participants had, the intensity of the urge, whether participants were hungry (indicated as 'yes' or 'no'), the time of day, and as control variables, participants' age, sex, and the day of the week. The model included an interaction term between the type of urge and hunger. The model included the participant as a random intercept and the type of urge as a random slope.

Using the afex package (Singmann et al., 2015), the p-values for the variables in the LMM were calculated by means of F-tests using the Kenward-Roger approximation for degrees-of-freedom (because these are easier to interpret for factors than tests on regression coefficients), and the p-values for the GLMM were calculated by comparing the likelihoods of the full and restricted models using likelihood ratio tests (with corresponding Pearson's χ^2 statistics). We then investigated any significant effects of factors and their interactions in the models described above using the emmeans package (Lenth et al., 2018) to identify the differences between the estimated marginal means of each factor, where we report Bonferroni adjusted p-values. For significant effects of numeric predictors, we also report the model coefficient.

All data and code are available here.

3. Results

The final sample consisted of 47 participants and none who participated in this part of the project were excluded from the study. Of these, 28 (59.6%) were female $M_{age} = 24.7$ ($SD = 6.3$), 18 (38.3%) were male $M_{age} = 25.9$ ($SD = 6.4$), and one preferred not to reveal their sex. Participants missed 21.1% of prompts (350 out of 1666). Of the 1316 surveys returned, in 874 participants reported to have experienced an urge.

Some urges had a much higher base rate than others. For example, the urge to snack was reported 731 times, while the urge to gamble was only reported 111 times and the urge to smoke 109 times. The other most common urges were online shopping (n = 344) and drinking alcohol (n = 269). As is to be expected, there was large interpersonal variability in the occurrence of different types of urges. Every participant reported the urge to snack at least twice over the experiment (max. 39 times). In comparison, only 14 participants reported the urge to smoke, and 19 to commit an act of aggression, at least twice.

Participants' age did not significantly affect the strength of their urges [1, 10], $b = 0.016$, 95% CI [-0.280, 0.312], $F(1, 38) = 0.01$, $p = .93$, or the probability of controlling those urges, odds ratio (OR) = 0.902, 95% CI [0.646, 1.259], $\chi2(1) = 0.3$, p = .58. Sex also did not impact the strength of urges, $F(2, 19) = 2.01$, $p = .16$. Males (estimated marginal mean (EMM) = 3.64, 95% CI [3.02, 4.26]) did not report significantly different urge intensities than females (EMM = 3.15, 95% CI [2.62, 3.68]). The same held for the influence of sex on the probability of controlling urges: there was no significant difference in the probability of controlling urges for males versus females OR = 0.662, 95% CI [0.309, 1.419], $\chi2(2) = 1.93$, $p = .38$.

3.1. The influence of state impulsivity and hunger on the intensity of urges

Participants' responses on the MIS significantly predicted the strength of urges, $F(1, 1494) = 99.04$, $p < .001$. For each 1 SD increase on the MIS, participants reported a 0.622-point increase in the strength of the urge ($b = 0.622$, 95% CI [0.503, 0.741]). See Fig. 1 for a plot of the model output on the linear relationship between the intensity of urges and MIS scores.

Participants' hunger state was also a significant predictor for urge intensity $F(1, 1485) = 7.80$, $p < .01$. They reported more intense urges when they were hungry, $b = 0.385$, 95% CI [0.115, 0.656]. When participants indicated they were hungry, the intensity of urges went up by 0.385 points.

There was also a significant effect of the type of urge on the perceived

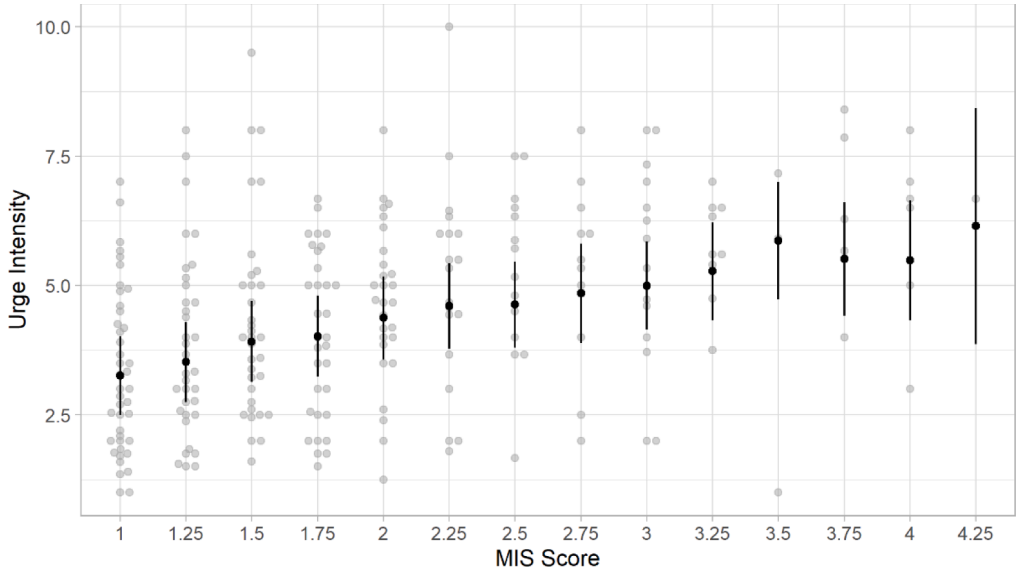

Fig. 1. The association between state-impulsivity (MIS) scores [1,5] (x-axis) and the urge intensity (y-axis) in the linear mixed model (MIS scores are presented as a factor here but is a continuous variable in the model). The error bars represent 95% CIs.

strength of urges, $F(5, 11) = 6.14$, $p < .01$. The urge to snack produced the strongest urges (estimated marginal mean (*EMM*) = 5.04, 95% CI [4.22, 5.85]), second strongest were urges to shop (*EMM* = 4.36, 95% CI [3.52, 5.20]), third were urges to drink alcohol (*EMM* = 4.13, 95% CI [3.28, 4.99]), fourth were urges to smoke (*EMM* = 3.38, 95% CI [2.21, 4.54]), followed by aggression-related urges (*EMM* = 3.76, 95% CI [2.68, 4.83]), and urges to gamble (*EMM* = 3.08, 95% CI [1.80, 4.36]).

There was also a significant interaction effect between the type of urge and hunger $F(5, 1079) = 2.72$, $p = .02$. Hungry participants reported stronger urges to snack (*EMM* = 5.46), $b = 0.846$, 95% CI [0.495, 1.196], $t(886) = 4.734$, $p < .0001$), but also to drink (*EMM* = 4.63), $b = 0.993$, 95% CI [0.488, 1.499], $t(1159) = 3.853$, $p < .001$) than participants who reported that they were not hungry (*EMM* = 4.61; EMM = 3.63, respectively). This means that the effect of hunger on the intensity of urges was stronger for urges to drink and snack than for other urges. See Fig. 2 for the relationship between the urge intensity of various urges and the hunger state of participants.

3.2. The influence of state impulsivity and hunger on the ability to control an urge

There was no significant effect of state impulsivity on the probability of controlling an urge OR = 0.871, 95% CI [0.739, 1.027], $\chi^2(1) = 2.72$, $p = .1$. This means that an increase of one standard deviation in a participants' response on the MIS would result in a (non-significant) change in the odds of controlling that urge between 0.739 and 1.027. Importantly, the 95% confidence interval includes 1 – the odds would stay the same.

Hunger significantly predicted whether participants controlled their urges, $\chi^2(1) = 4.26$, $p = .04$. A follow-up test showed that the odds ratio

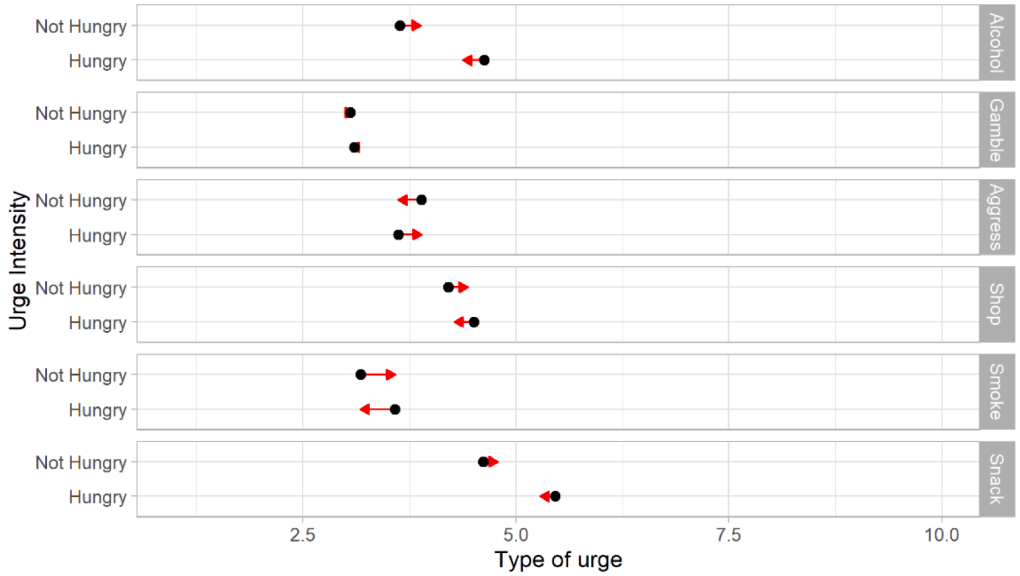

Fig. 2. The interaction effect between participants' hunger state (y-axis) and the different types of urge (panel labels) on urge intensity (x-axis). The comparison arrows represent the pairwise tests; arrows that do not overlap represent a significant difference at a bonferroni adjusted alpha level of 0.05.

of participants controlling an urge when participants were hungry, as opposed to when participants were not hungry was 0.644:1 (95% CI [0.426, 0.973].

The type of urge had a significant effect on the probability that participants controlled their urges, $\chi^2(5) = 28.38$, $p < .001$. Participants exhibited the lowest probability of controlling an urge to snack (*EMM* = 0.687, 95% CI [0.492, 0.833]). Second lowest probability of controlling an urge was for smoking (*EMM* = 0.758, 95% CI [0.461, 0.920]), third lowest was for drinking (*EMM* = 0.786, 95% CI [0.604, 0.898]). Participants were the most likely to control urges pertaining to gambling (*EMM* = 0.832, 95% CI [0.609, 0.940]), shopping (*EMM* = 0.866, 95% CI [0.725, 0.941]) and aggression (*EMM* = 0.949, 95% CI [0.833, 0.986]). This means that people had the lowest probability of controlling an urge to snack (holding all else equal), controlling on average 69/100 of these, whereas they had the highest probability of controlling an urge pertaining to aggression, controlling 95/100 of these. There was no significant interaction effect of hunger and the type of urge on the likelihood of controlling an urge $\chi^2(5) = 5.87$, $p = .32$. See Fig. 3 for the estimated marginal means of the probability of controlling an urge per urge type, partitioned by hunger state.

The perceived intensity of an urge significantly predicted the probability that a participant controlled that urge, $\chi^2(1) = 118.36$, $p < .001$. A more intense urge was associated with a lower probability of controlling the urge, OR = 0.694, 95% CI [0.647, 0.746]. This means that for every one-point increase on the 1–10 urge intensity scale, the odds of controlling the urge relative to the previous intensity level was 0.694. This roughly translates to the likelihood of controlling an urge with a '6' response on the urge intensity scale, being 0.694 as high as controlling an urge with a '5' response on the urge intensity scale. See Fig. 4 for the effect of the intensity of urges on the probability of controlling it.

3.3. The influence of the time of day on urge intensity and the likelihood of controlling an urge.

There was no effect of the time of day on the perceived intensity of urges, $b = 0.073$, 95% CI [-0.017, 0.163], $F(1, 1596) = 2.49$, $p = .12$.

The time of day had a significant effect on the probability that participants controlled their urges $\chi^2(1) = 12.39$, $p < .001$, as participants were less likely to control their urges later in the day, OR = 0.791, 95% CI [0.693, 0.902]. The coefficient in the model is in log odds, transforming this variable to odds would show that for each hour later in the day, the odds ratio would be 0.79:1 for controlling the urge compared to the previous hour. See Fig. 5 for the relationship between the time of day and the probability of controlling urges.

4. Discussion

In this study, we studied the effects of state impulsivity and hunger on the intensity of various types of urges and on self-control. We found that state impulsivity influences urge intensity, but not the probability of controlling an urge, suggesting a bottom-up influence on impulsive behaviour. Further, we found that hunger affects both urge intensity and the probability of controlling an urge, suggesting support for a dual-process model for the effect of hunger on behaviour. Lastly, we found that the likelihood of controlling an urge was lower when it was later in the day, whereas there was no such effect on urge intensity.

4.1. State impulsivity

Participants reported stronger urges when they were in an impulsive state. However, they did not report diminished self-control in conjunction with higher levels of state impulsivity. That is, while state impulsivity was indirectly associated with the probability of controlling urges through urge amplification, it did not have a direct effect on self-control. This evidence is in support of a bottom-up influence of state impulsivity, whereas we find no support for a top-down influence (i.e., state impulsivity appears to sensitise urges, while we found no evidence that it down-regulates inhibition). The lack of support for a top-down effect is surprising given the prevailing view that impulsivity is closely tied with a lack of self-control (Carver & White, 1994; Nigg, 2017; Patock-Peckham et al., 2001; Wiers et al., 2010).

Moreover, our findings concerning state impulsivity have implications to qualify the predictions of dual-systems models. Specifically, even though trait impulsivity ought to correlate with both the automatic and the deliberative systems, the two systems may not be equally affected by state impulsivity. This study provides evidence for a different effect of state impulsivity on the urge control process than is proposed about trait impulsivity, that is, it suggests that state impulsivity primarily affects people's capacity to resist predominant response tendencies (i.e., self-control) indirectly, by strengthening the emphasis on

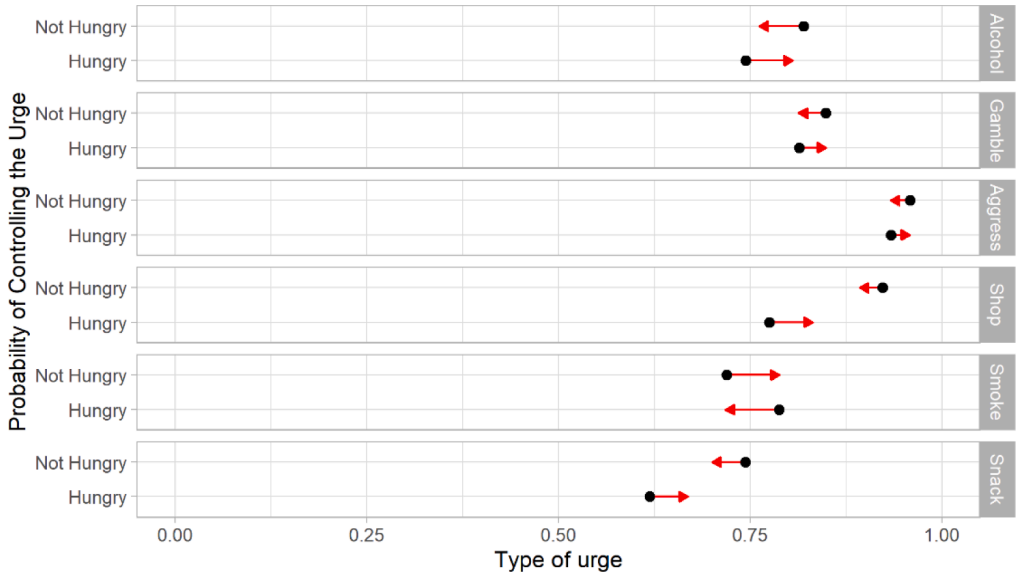

Fig. 3. The effect of hunger (y-axis) on the probability of controlling in urge (x-axis), partitioned by the type of urge (panel labels). The arrows represent pairwise comparisons; arrows that do not overlap represent a significant difference at a bonferroni adjusted alpha level of 0.05.

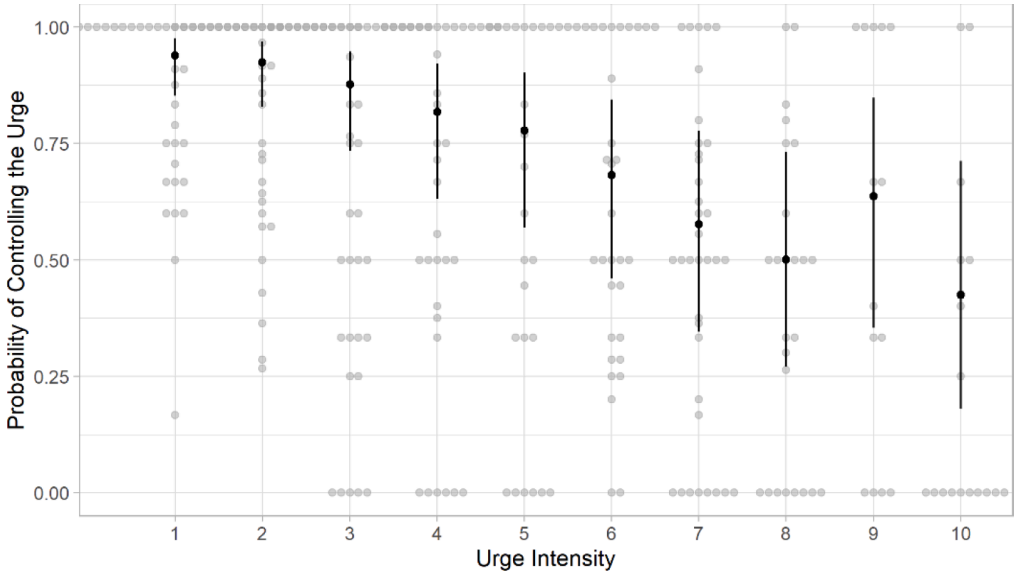

Fig. 4. The relationship between urge intensity (x-axis) and the probability of controlling an urge (y-axis). The grey dots represent data aggregated over participants in each urge intensity level. The error bars represent 95% CIs.

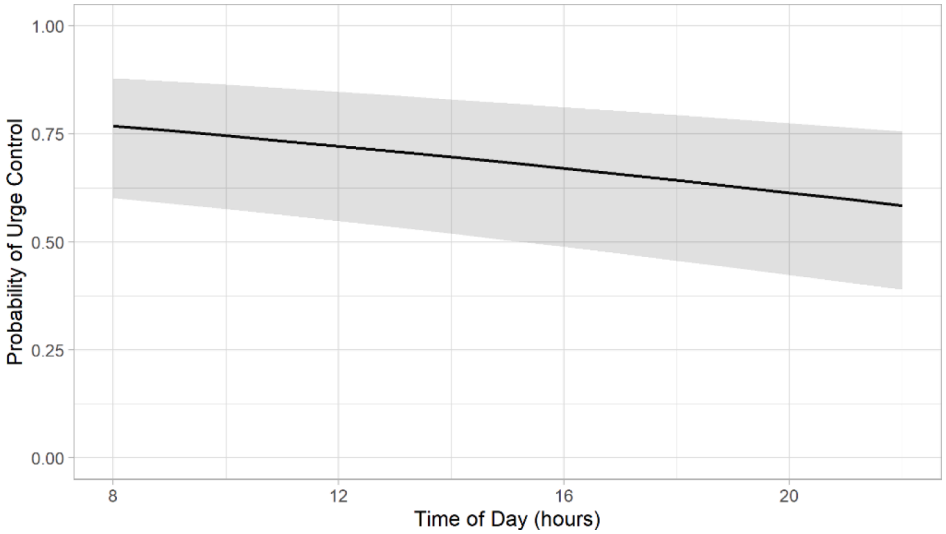

Fig. 5. The association between the time of day and the probability of controlling an urge. The grey band represents a 95% CI.

the impulse-response system—not by limiting people's ability to resist. More research needs to be done to determine whether state impulsivity is also directly associated with lesser self-control.

The implication is that state impulsivity impacts the salience of stimuli, which is significant because it has implications for clinical research on addiction. In people with drug addiction, drug-related stimuli are already more salient than other important stimuli (e.g., Lubman et al., 2008) and impulsivity is implicated in attentional bias toward drug-related stimuli (Coskunpinar & Cyders, 2013; Field & Cox, 2008), so fluctuations in state impulsivity might amplify this difference in salience and attention, temporarily making drug-related urges too strong to withstand.

Given the strong relationship between the intensity of urges and whether the urge is controlled, these results suggest that the effort exerted into improving the capacity to control urges when people are in an impulsive state could be better spent on other on-the-fly interventions that help reduce urges and cravings, or identifying and removing triggers for impulsive states (Hawker et al., 2021; Kakoschke et al., 2018).

4.2. Hunger

When participants were hungry, their urges were stronger, and they had a lower probability of controlling urges, even when controlling for urge intensity. More surprising, however, is that the effect of hunger on both urge intensity and the probability of controlling urges is not limited to eating-related stimuli (and thus is somewhat domain-general). That is, hunger amplifies the urges people experience, as well as decreasing the likelihood that they control an urge, holding fixed the intensity of the urge.

This effect corroborates the findings that hunger increases impulsive behaviour in non-human animals (Anderberg et al., 2016; Laude et al., 2012; Zheng et al., 2019), and some emerging evidence of the same effect in humans in the domains of gambling (Li et al., 2020) and delay

discounting (Skrynka & Vincent, 2019), which is implicated in various addictive behaviours (Amlung et al., 2017). It also increases the confidence in the external validity of the effect of hunger on impulsive behaviour because this effect was observed in participants' natural environments.

It initially seems that hunger and the time-of-day influence self-control (i.e., the *capacity* to control urges). However, we did not ask participants whether they resisted the urge, so it may be that they merely resisted their urges less when they were hungry, or when it was late (i.e., the *motivation* to control urges). It may still be that the time of day influences self-control, but it is likely that as it gets later in the day, people switch task priorities as they transition from pursuing 'have-to' goals, to 'want-to' goals (Inzlicht et al., 2014). This motivational effect has been proposed as an alternative to ego depletion, which is closely related to the time of day. However, to identify whether the effect is motivational or whether self-control is affected, more research is needed to assess whether participants are less likely to attempt to resist urges later in the day, and when hungry.

There was also an interaction effect of hunger with the type of urge; urges pertaining to drinking alcohol and snacking were more strongly amplified by hunger than the other types. It could be that hunger leads to stronger amplification of the urge to drink alcohol and snack because these both address the homeostatic need for energy. Indeed, there is some evidence that interoceptive capacity for differentiating hunger from thirst might be limited (Eiselt et al., 2021). Thus, trouble dissociating an urge to snack from hunger is to be expected, though specific food cravings often occur without the urge to snack (Massey & Hill, 2012), and trends in snack-food cravings and general hunger are dissociable (Reichenberger et al., 2018). But, this is a further leap for alcoholic drinks—there should be good dissociability between hunger and alcohol cravings. A conjecture for future research is that it is due to the association of alcoholic drinks with food (Escrivá-Martínez et al., 2020).

4.3. Methodological considerations

We were able to find these effects because of the smartphone-based EMA methodology used for the experiment, which allowed us to study the effects of hunger on impulsive behaviour in people's everyday environment. This is necessary to show that the effects of hunger on impulsive behaviour is present in the outside world, rather than just in the laboratory. Moreover, it allowed us to tease apart the effect of hunger on the urge, rather than just on the likelihood of controlling various types of urges without creating an unwieldy design with various experiments testing different types of urges.

Through its better ecological validity, EMA methodology has distinct advantages over other approaches, which bolsters confidence that findings in the study replicate in real life. However, researchers also to some extent must relinquish control over the environment. This means that researchers are unable to fully ascertain whether participants are paying attention and what other variables might be interfering with the experiment.

The lack of an effect for a top-down influence of impulsivity in this study may be due to the limited interoceptive abilities of participants. Participants might inflate the experienced intensity of urges in hindsight to justify their failure to control them. But, if this were the case, then we should not have found the top-down influence of hunger, as it would be unlikely that participants would not have the same interoceptive issue for the influence of hunger. Alternatively, the lack of a top-down influence in our sample could be caused by a lack of power for a smaller effect size, which could be addressed in future research.

There were some limitations to our design. In particular, the effect sizes of hunger on urge intensity and self-control in our study might be imprecise because we used a binary measure for measuring hunger; a continuous measure could provide more information on the effects of hunger on behaviour. In addition, our sample skewed towards female participants, and although there was no effect of sex on either of the dependent variables, this skew could affect the external validity of the results. As such, further research on sex differences in the experience and impact of state impulsivity is needed.

Future research could clarify the implications of impulsivity traits on daily life by correlating them with behaviour patterns in EMA experiments. Research in this area is sparse, but initial work has been completed on affective instability (Solhan et al., 2009). It is important to know how self-reported 'typical' behaviour or experiences, which is what trait measures usually rely on, translate to real-time behaviour. Especially because humans are prone to recall biases, and this information is often used to diagnose psychiatric disorders. Recall bias could cause misdiagnosis or overdiagnosis leading to overmedication, unnecessary treatment, and increased burden on the healthcare system (Bruchmüller et al., 2012). Shedding light on potential systematic errors in answering trait scales would be useful in clinical settings and could be utilised to improve trait scales.

The findings in the current work could aid clinical research, as impulsivity is a key symptom of many neurological and mental disorders, including brain injury, dementia, substance and behavioural addictions, bipolar, eating or personality disorders, and many psychopathological symptoms show variability (American Psychiatric Association, 2013; Ebner-Priemer et al., 2009; Moeller et al., 2001). A focus on the effects of state impulsivity and hunger on behaviour in clinical settings, using an EMA paradigm like the one presented in the current study, will be useful considering that these states are more proximal and thus highly relevant to behaviours of interest. Acquiring new insights on the influence of fluctuations in state impulsivity, hunger could aid in identifying better ways to maintain goal-directed behaviour and thus better treatment outcomes.

CRediT authorship contribution statement

Simon Thomas van Baal: Conceptualization, Methodology, Software, Data curation, Visualization, Validation, Writing – original draft, Writing – review & editing, Project administration. **Neda Moskovsky:** Resources, Methodology, Software, Data curation, Project administration. **Jakob Hohwy:** Writing – original draft, Writing – review & editing, Visualization, Supervision. **Antonio Verdejo-García:** Conceptualization, Methodology, Investigation, Resources, Funding acquisition, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to acknowledge Dr. Peter Koval and his team for providing access to the Smartphone Ecological Momentary Assessment (SEMA) app.

Funding statement

This study was funded by the Australian Research Council grant ARC DP180100245. The funding source was not involved in the research conducted for this project.

Ethics approval

Informed consent was obtained from all participants and the project was approved by the Monash University Human Research Ethics Committee (Project ID: 11999).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.addbeh.2022.107381.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th, Ed.).

Amlung, M., Vedelago, L., Acker, J., Balodis, I., & MacKillop, J. (2017). Steep delay discounting and addictive behavior: A meta-analysis of continuous associations. *Addiction, 112*(1), 51–62. https://doi.org/10.1111/add.13535

Anderberg, R. H., Hansson, C., Fenander, M., Richard, J. E., Dickson, S. L., Nissbrandt, H., … Skibicka, K. P. (2016). The stomach-derived hormone ghrelin increases impulsive behavior. *Neuropsychopharmacology, 41*(5), 1199–1209. https://doi.org/10.1038/npp.2015.297

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*(6), 351-355. 10.1111%2Fj.1467-8721.2007.00534.x.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience, 8*(11), 1458–1463. https://doi.org/10.1038/nn1584

Bembenutty, H. (2011). Academic delay of gratification and academic achievement. *Ed. Bembunutty, H. Self Regulated Learning: New Direction for Teaching and Learning. San Francisco. Wiley Periodicals, İnc*.

Berg, J. M., Latzman, R. D., Bliwise, N. G., & Lilienfeld, S. O. (2015). Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the UPPS for psychopathology. *Psychological assessment, 27*(4), 1129. https://doi.org/10.1037/pas0000111

Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *Journal of Consulting and Clinical Psychology, 80*(1), 128. https://doi.org/10.1037/a0026582

Cameron, J. D., Goldfield, G. S., Finlayson, G., Blundell, J. E., & Doucet, É. (2014). Fasting for 24 hours heightens reward from food and food-related cues. *PLoS One, 9*(1), e85970. https://doi.org/10.1371/journal.pone.0085970

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*(2), 319. https://doi.org/10.1037/0022-3514.67.2.319

Coskunpinar, A., & Cyders, M. A. (2013). Impulsivity and substance-related attentional bias: A meta-analytic review. *Drug and Alcohol Dependence, 133*(1), 1–14. https://doi.org/10.1016/j.drugalcdep.2013.05.008

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Doran, N., Spring, B., & McChargue, D. (2007). Effect of impulsivity on craving and behavioral reactivity to smoking cues. *Psychopharmacology, 194*(2), 279–288. https://doi.org/10.1007/s00213-007-0832-x

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *Journal of abnormal psychology, 118*(1), 195. https://doi.org/10.1037/a0014868

Eiselt, A.-K., Chen, S., Chen, J., Arnold, J., Kim, T., Pachitariu, M., & Sternson, S. M. (2021). Hunger or thirst state uncertainty is resolved by outcome evaluation in medial prefrontal cortex to guide decision-making. *Nature neuroscience, 1–6*. https://doi.org/10.1038/s41593-021-00850-4

Escrivá-Martínez, T., Herrero, R., Molinari, G., Rodríguez-Arias, M., Verdejo-García, A., & Baños, R. M. (2020). Binge Eating and Binge Drinking: A Two-Way Road? *An Integrative Review. Current pharmaceutical design, 26*(20), 2402–2415. https://doi.org/10.2174/1381612826666200316153317

Feltenstein, M., & See, R. (2008). The neurocircuitry of addiction: an overview. *British journal of pharmacology, 154*(2), 261-274. https://dx.doi.org/10.1038%2Fbjp.2008.51.

Field, M., & Cox, W. M. (2008, 2008/09/01/). Attentional bias in addictive behaviors: A review of its development, causes, and consequences. *Drug and Alcohol Dependence, 97*(1), 1-20. 10.1016/j.drugalcdep.2008.03.030.

Finne, E., Englert, C., & Jekauc, D. (2019). On the importance of self-control strength for regular physical activity. *Psychology of Sport and Exercise, 43*, 165–171. https://doi.org/10.1016/j.psychsport.2019.02.007

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences, 115*(27), E6106-E6115. 10.1073/pnas.1711978115.

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and social psychology review, 23*(2), 107-131. 10.1177%2F1088868318762183.

Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational behavior and human decision processes, 88*(1), 430–444. https://doi.org/10.1006/obhd.2001.2982

Goldstein, R. Z., & Volkow, N. D. (2002). Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. *American Journal of Psychiatry, 159*(10), 1642–1652. https://doi.org/10.1176/appi.ajp.159.10.1642

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews Neuroscience, 12*(11), 652–669. https://doi.org/10.1038/nrn3119

Gough, H. G. (2000). The California psychological inventory. *Testing and Assessment in Counseling Practice, 45*, 71.

Hawker, C. O., Merkouris, S. S., Youssef, G. J., & Dowling, N. A. (2021). A smartphone-delivered ecological momentary intervention for problem gambling (GamblingLess: Curb Your Urge): Single-arm acceptability and feasibility trial. *Journal of Medical Internet Research, 23*(3), e25786.

Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What people desire, feel conflicted about, and try to resist in everyday life. *Psychological science, 23*(6), 582-588. 10.1177%2F0956797612437426.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in cognitive sciences, 18*(3), 127–133. https://doi.org/10.1016/j.tics.2013.12.009

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kakoschke, N., Hawker, C., Castine, B., de Courten, B., & Verdejo-Garcia, A. (2018). Smartphone-based cognitive bias modification training improves healthy food choice in obesity: A pilot study. *European Eating Disorders Review, 26*(5), 526–532. https://doi.org/10.1002/erv.2622

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43*(1), 23. https://doi.org/10.1037/0003-066X.43.1.23

Koomen, R., Grueneisen, S., & Herrmann, E. (2020). Children delay gratification for cooperative ends. *Psychological science, 31*(2), 139-148. 10.1177%2F0956797619894205.

Koval, P. H., J; Dozo, N; Gleeson, J; Alvarez, M; Harrison, A; Vu, D; Susanto, R; Jayaputera, G; Sinnott, R. (2019). *SEMA3*. In https://sema3.com/.

Laude, J. R., Pattison, K. F., & Zentall, T. R. (2012). Hungry pigeons make suboptimal choices, less hungry pigeons do not. *Psychonomic Bulletin & Review, 19*(5), 884–891. https://doi.org/10.3758/s13423-012-0282-2

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version, 1*(1), 3.

Li, E., Hing, N., Russell, A. M., & Vitartas, P. (2020). Impulsive Sports Betting: The Effects of Food or Substance Consumption. *Journal of Gambling Studies, 36*(2), 539–554. https://doi.org/10.1007/s10899-020-09938-1

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65*(3), 272–292. https://doi.org/10.1006/obhd.1996.0028

Lubman, D. I., Allen, N. B., Peters, L. A., & Deakin, J. W. (2008). Electrophysiological evidence that drug cues have greater salience than other affective stimuli in opiate addiction. *Journal of psychopharmacology, 22*(8), 836-842. 10.1177%2F0269881107083846.

Massey, A., & Hill, A. J. (2012). Dieting and food craving. A descriptive, quasi-prospective study. *Appetite, 58*(3), 781–785. https://doi.org/10.1016/j.appet.2012.01.020

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3.

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science, 244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry, 158*(11), 1783–1793. https://doi.org/10.1176/appi.ajp.158.11.1783

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry, 58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Papachristou, H., Nederkoorn, C., Havermans, R., van der Horst, M., & Jansen, A. (2012). Can't stop the craving: The effect of impulsivity on cue-elicited craving for alcohol in heavy and light social drinkers. *Psychopharmacology, 219*(2), 511–518. https://doi.org/10.1007/s00213-011-2240-5

Parker, J. D., & Bagby, R. (1997). *Impulsivity in adults: A critical review* (pp. 142–157). Impulsivity: Theory, assessment, and treatment.

Patock-Peckham, J. A., Cheong, J., Balhorn, M. E., & Nagoshi, C. T. (2001). A social learning perspective: A model of parenting styles, self-regulation, perceived drinking control, and alcohol use and problems. *Alcoholism: Clinical and Experimental Research, 25*(9), 1284–1292. https://doi.org/10.1111/j.1530-0277.2001.tb02349.x

Potenza, M. N., Steinberg, M. A., Skudlarski, P., Fulbright, R. K., Lacadie, C. M., Wilber, M. K., … Wexler, B. E. (2003). Gambling urges in pathological gambling: A functional magnetic resonance imaging study. *Archives of General Psychiatry, 60*(8), 828–836. https://doi.org/10.1001/archpsyc.60.8.828

Rawn, C. D., & Vohs, K. D. (2006). The Importance of Self-Regulation for Interpersonal Functioning.

Reichenberger, J., Richard, A., Smyth, J. M., Fischer, D., Pollatos, O., & Blechert, J. (2018). It's craving time: Time of day effects on momentary hunger and food craving in daily life. *Nutrition, 55*, 15–20. https://doi.org/10.1016/j.nut.2018.03.048

Saper, C. B., Chou, T. C., & Elmquist, J. K. (2002). The need to feed: Homeostatic and hedonic control of eating. *Neuron, 36*(2), 199–211. https://doi.org/10.1016/S0896-6273(02)00969-8

Schlam, T. R., Wilson, N. L., Shoda, Y., Mischel, W., & Ayduk, O. (2013). Preschoolers' delay of gratification predicts their body mass 30 years later. *The Journal of Pediatrics, 162*(1), 90–93. https://doi.org/10.1016/j.jpeds.2012.06.049

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin, 140*(2), 374. https://doi.org/10.1037/a0034418

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol., 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). afex: Analysis of factorial experiments. *R package version*, 13–145.

Skrynka, J., & Vincent, B. T. (2019). Hunger increases delay discounting of food and non-food rewards. *Psychonomic Bulletin & Review, 26*(5), 1729–1737. https://doi.org/10.3758/s13423-019-01655-0

Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment, 21*(3), 425. https://doi.org/10.1037/a0016869

Strickland, J. C., & Johnson, M. W. (2021). Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. *Psychological Review, 128*(2), 336. https://doi.org/10.1037/rev0000263

Tomko, R. L., Solhan, M. B., Carpenter, R. W., Brown, W. C., Jahng, S., Wood, P. K., & Trull, T. J. (2014). Measuring impulsivity in daily life: The momentary impulsivity scale. *Psychological Assessment, 26*(2), 339. https://doi.org/10.1037/a0035083

Verdejo-García, A., Lawrence, A. J., & Clark, L. (2008). Impulsivity as a vulnerability marker for substance-use disorders: Review of findings from high-risk research, problem gamblers and genetic association studies. *Neuroscience and Biobehavioral Reviews., 32*(4), 777–810. https://doi.org/10.1016/j.neubiorev.2007.11.003

Verdejo-Garcia, A., Moreno-Padilla, M., Garcia-Rios, M. C., Lopez-Torrecillas, F., Delgado-Rico, E., Schmidt-Rio-Valle, J., & Fernandez-Serrano, M. J. (2015 Apr 21). Social stress increases cortisol and hampers attention in adolescents with excess weight. *PLoS One, 10*(4), Article e0123565. https://doi.org/10.1371/journal.pone.0123565

Verdejo-Garcia, A., Tiego, J., Kakoschke, N., Moskovsky, N., Voigt, K., Anderson, A., … Bellgrove, M. A. (2021). Nov). A unified online test battery for cognitive impulsivity reveals relationships with real-world impulsive behaviours. *Nat Hum Behav, 5*(11), 1562–1577. https://doi.org/10.1038/s41562-021-01127-3

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Wiers, R., Ames, S. L., Hofmann, W., Krank, M., & Stacy, A. (2010). Impulsivity, impulsive and reflective processes and the development of alcohol use and misuse in adolescents and young adults. *Frontiers in psychology, 1*, 144. https://dx.doi.org/10.3389%2Ffpsyg.2010.00144.

Xu, A. J., Schwarz, N., & Wyer, R. S. (2015). Hunger promotes acquisition of nonfood objects. *Proceedings of the National Academy of Sciences, 112*(9), 2688–2692. https://doi.org/10.1073/pnas.1417712112

Yarmush, D. E., Manchery, L., Luehring-Jones, P., & Erblich, J. (2016). Gender and impulsivity: Effects on cue-induced alcohol craving. *Alcoholism: Clinical and Experimental Research, 40*(5), 1052–1057. https://doi.org/10.1111/acer.13030

Zheng, H., Ton, H., Yang, L., Liufu, N., Dong, Y., Zhang, Y., & Xie, Z. (2019). Acute fasting does not induce cognitive impairment in mice. *Frontiers in Neuroscience, 13*, 896. https://doi.org/10.3389/fnins.2019.00896

3.7 Supplementary Materials

In the supplementary materials, we provide further details about the analysis. Not all interactions in both models were reported because they were not of interest for our research questions. However, the exclusion of these variables from the model could theoretically affect the coefficients and significance of the relationships between the other variables through omitted variable bias.

Thus, to ensure transparency of the relationships between the variables in a fully specified model, we report the effects of the other interactions in table format. See Table S1 and Table S2.

TABLE S1: A regression table, showing a comparison between a linear mixed model, with urge intensity as the dependent variable, that has maximally specified interactions between the predictors of interest (left) and the model used in the article (right).

	Dependent variable	
	Urge Intensity	
Independent variable	(1)	(2)
hungerState1	0.220***	0.193***
	(0.073)	(0.066)
typeOfUrge1	0.267*	0.175
	(0.14)	(0.182)
typeOfUrge2	-0.888**	-0.878***
	(0.369)	(0.274)
typeOfUrge3	-0.192	-0.2
	(0.225)	(0.216)
typeOfUrge4	0.441**	0.401**
	(0.194)	(0.188)
typeOfUrge5	-0.783**	-0.578*
	(0.337)	(0.339)
MIS	0.656***	0.622***
	(0.079)	(0.061)
hour	0.073	0.073
	(0.046)	(0.046)
age	-0.004	0.016
	(0.155)	(0.151)
sex1	-1.147***	-0.805**
	(0.324)	(0.342)
sex2	1.736***	1.121**
	(0.516)	(0.57)
weekDay1	0.134	0.128
	(0.11)	(0.109)
weekDay2	-0.047	-0.047
	(0.118)	(0.118)
weekDay3	-0.07	-0.099

Continued on next page

Table S1 – continued

	Urge Intensity	
	(1)	(2)
	(0.107)	(0.106)
weekDay4	0.006	-0.001
	(0.117)	(0.117)
weekDay5	-0.095	-0.055
	(0.108)	(0.107)
weekDay6	0.135	0.134
	(0.121)	(0.121)
hungerState1:typeOfUrge1	0.344***	0.304**
	(0.123)	(0.121)
hungerState1:typeOfUrge2	0.007	-0.169
	(0.217)	(0.174)
hungerState1:typeOfUrge3	-0.439**	-0.328**
	(0.171)	(0.158)
hungerState1:typeOfUrge4	-0.074	-0.045
	(0.118)	(0.112)
hungerState1:typeOfUrge5	-0.037	0.008
	(0.193)	(0.188)
hungerState1:MIS	0.051	
	(0.07)	
typeOfUrge1:MIS	-0.352***	
	(0.122)	
typeOfUrge2:MIS	-0.036	
	(0.198)	
typeOfUrge3:MIS	0.112	
	(0.181)	
typeOfUrge4:MIS	0.094	
	(0.12)	
typeOfUrge5:MIS	0.154	
	(0.227)	
hungerState1:typeOfUrge1:MIS	0.261**	
	(0.118)	
hungerState1:typeOfUrge2:MIS	-0.335*	
	(0.177)	
hungerState1:typeOfUrge3:MIS	0.165	
	(0.18)	
hungerState1:typeOfUrge4:MIS	-0.125	
	(0.111)	
hungerState1:typeOfUrge5:MIS	0.121	
	(0.21)	
Constant	4.187***	3.957***
	(0.298)	(0.311)
Observations	1,688	1,688
Log Likelihood	-3,533.75	-3,524.41
Akaike Inf. Crit.	7,179.50	7,138.83
Bayesian Inf. Crit.	7,483.66	7,383.24

Continued on next page

Table S1 – continued

	Urge Intensity	
	(1)	(2)
Note:	*p<0.1;**p<.05;***p<0.01	

Table S2: A regression table, showing a comparison between a binomial generalised linear mixed model, with self-control as the dependent variable, that has maximally specified interactions between the predictors of interest (left) and the model used in the article (right).

	Dependent variable	
	Urge Control	
Independent variable	(1)	(2)
hungerState1	-0.291**	-0.220**
	(0.118)	(0.105)
typeOfUrge1	-0.253	-0.303
	(0.227)	(0.208)
typeOfUrge2	-0.156	-0.005
	(0.452)	(0.412)
typeOfUrge3	1.360**	1.319***
	(0.583)	(0.501)
typeOfUrge4	0.361	0.268
	(0.265)	(0.239)
typeOfUrge5	-0.499	-0.463
	(0.441)	(0.41)
MIS	-0.135	-0.138*
	(0.129)	(0.084)
age	-0.078	-0.103
	(0.178)	(0.17)
sex1	-0.438	-0.452
	(0.427)	(0.423)
sex2	0.481	0.491
	(0.793)	(0.786)
hour	-0.240***	-0.236***
	(0.068)	(0.067)
weekDay1	0.179	0.186
	(0.164)	(0.164)
weekDay2	-0.251	-0.243
	(0.168)	(0.167)
weekDay3	-0.056	-0.047
	(0.155)	(0.154)
weekDay4	-0.152	-0.125
	(0.172)	(0.17)
weekDay5	0.136	0.121
	(0.155)	(0.154)
weekDay6	0.255	0.244
	(0.174)	(0.174)
urgeIntensity	-0.362***	-0.365***
	(0.037)	(0.036)

Continued on next page

Table S2 – continued

	Urge Control	
	(1)	(2)
hungerState1:typeOfUrge1	-0.006	-0.001
	(0.194)	(0.177)
hungerState1:typeOfUrge2	-0.002	0.092
	(0.319)	(0.253)
hungerState1:typeOfUrge3	-0.053	-0.028
	(0.349)	(0.301)
hungerState1:typeOfUrge4	-0.396**	-0.399**
	(0.199)	(0.181)
hungerState1:typeOfUrge5	0.494	0.406
	(0.307)	(0.289)
hungerState1:MIS	0.089	
	(0.117)	
typeOfUrge1:MIS	0.324	
	(0.205)	
typeOfUrge2:MIS	0.111	
	(0.274)	
typeOfUrge3:MIS	-0.127	
	(0.368)	
typeOfUrge4:MIS	-0.412**	
	(0.197)	
typeOfUrge5:MIS	0.015	
	(0.393)	
hungerState1:typeOfUrge1:MIS	-0.239	
	(0.194)	
hungerState1:typeOfUrge2:MIS	0.05	
	(0.246)	
hungerState1:typeOfUrge3:MIS	0.139	
	(0.354)	
hungerState1:typeOfUrge4:MIS	0.201	
	(0.179)	
hungerState1:typeOfUrge5:MIS	0.046	
	(0.339)	
Constant	1.623***	1.602***
	(0.44)	(0.428)
Observations	1,687	1,687
Log Likelihood	-790.357	-795.875
Akaike Inf. Crit.	1,692.71	1,681.75
Bayesian Inf. Crit.	1,996.83	1,926.13
Note:	*p<.1;**p<.05;***p<0.01	

We also provide the variable coding for the different factor levels for the day of the week and the type of urge. This can be used to interpret the effects of the different days and urges on the dependent variables in Table S1 and Table S2. Deviance (or sum contrast) coding is different from dummy (or treatment) coding because it allows the model to

center the effects over the grand mean—instead of, say, at one particular reference level urge, and reference level weekday.

TABLE S3: Contrasts for the sex factor in both models. Contrasts are deviance (sum contrast) coded, which means that all the contrasts sum to 1. In this case, we would have to hold both sex coefficients at -1 to see the effect of being male compared to the grand mean.

	sex1	sex2
Female	1	0
NA	0	1
Male	-1	-1

TABLE S4: Contrasts for the weekday factor in the model.

	Weekday1	Weekday2	Weekday3	Weekday4	Weekday5	Weekday6
Monday	1	0	0	0	0	0
Tuesday	0	1	0	0	0	0
Wednesday	0	0	1	0	0	0
Thursday	0	0	0	1	0	0
Friday	0	0	0	0	1	0
Saturday	0	0	0	0	0	1
Sunday	-1	-1	-1	-1	-1	-1

TABLE S5: Contrasts for the type of urge factor in the model.

	typeOfUrge1	typeOfUrge2	typeOfUrge3	typeOfUrge4	typeOfUrge5
Alcohol	1	0	0	0	0
Gamble	0	1	0	0	0
Aggression	0	0	1	0	0
Shop	0	0	0	1	0
Smoke	0	0	0	0	1
Snack	-1	-1	-1	-1	-1

In addition, we provide the results of F-tests and likelihood ratio tests comparing the maximally specified models to the models used in the article, to show the difference in goodness-of-fit between different model specification. See Table S6 and Table S7 for these results.

TABLE S6: An ANOVA table, showing conditional F-tests for the predictors in the maximally specified model with urgen intensity as the dependent variable. Of particular interest are the bottom three rows because these interactions are not in the model used for article.

Effect	df	F	p.value
hungerState	1, 1502.42	8.68 **	0.003
typeOfUrge	5, 15.03	6.69 **	0.002
MIS	1, 504.56	59.01 ***	<.001
hour	1, 1591.08	2.43	0.119
age	1, 39.04	0	0.984
sex	2, 13.24	4.26 *	0.037
weekDay	6, 1582.48	0.63	0.709
hungerState:typeOfUrge	5, 1144.44	2.90 *	0.013
hungerState:MIS	1, 880.25	0.45	0.501
typeOfUrge:MIS	5, 329.08	1.49	0.191
hungerState:typeOfUrge:MIS	5, 872.88	1.98 +	0.08

TABLE S7: An ANOVA table, showing likelihood ratio tests for the predictors in the maximally specified model with whether the participant controlled the urge as the dependent variable. Of particular interest are the bottom three rows because these interactions are not in the model used for the article.

Effect	df	Chisq	p.value
hungerState	1	5.70 *	0.017
typeOfUrge	5	29.03 ***	<.001
MIS	1	1.11	0.292
age	1	0.23	0.634
sex	2	1.73	0.421
hour	1	12.59 ***	<.001
weekDay	6	6.64	0.355
urgeIntensity	1	114.94 ***	<.001
hungerState:typeOfUrge	5	5.79	0.327
hungerState:MIS	1	0.59	0.441
typeOfUrge:MIS	5	8.3	0.141
hungerState:typeOfUrge:MIS	5	5.06	0.409

Chapter 4

Curbing Urges with Compassion and Future Thinking

4.1 Linking Text for Chapter 4

In the previous Chapter, I showed that state changes are important for how people experience urges and the concomitant behavioural outcomes; in situations where people are hungry or when they feel impulsive, their urges become stronger.

The changeability of the intensity of urges undoubtedly sometimes leads to sub-optimal situations from a public health point of view—people become more vulnerable to their baser impulses in situations that are often predictable. While normally effective solutions to individual fallibilities can be found in system-level change (Chater & Loewenstein, 2022; Singer, 2022), sometimes systemic change is not feasible for a variety of reasons. This Chapter will explore people's experiences and ability to cope with urges in such a context.

Much of the research for the current PhD thesis was conducted during the COVID-19 pandemic. When the pandemic struck, it paralysed many nations across the globe. Policymakers struggled to find effective policies to combat a quickly developing virus with little evidence of how it spread or even how dangerous it would be if it would be allowed to run its course through society. One of the problems in combating COVID-19 was that the time frame in which the situation developed precluded any large-scale system change.

For example, it turned out that most viral transmission occurred indoors (though there was much uncertainty about this at first), and that one of the most effective remedies for this would be to have proper ventilation (see e.g., Berry et al., 2022). However, updating building standards for public spaces to ensure proper ventilation and applying those standards to older buildings would have taken decades to complete. Instead, public health officials advised people and businesses to open doors and windows—something that is particularly unpleasant in places where winters are cold or summers oppressive.

The lack of effective system change policy solutions as well as the lack of evidence on viral transmission pathways thus led to a need for behavioural interventions that were aimed at changing how people went about their day-to-day business (Bavel et al., 2020). This led to difficult public health trade-offs because such interventions were aimed mostly at reducing close contact between people; in most countries the public was asked to physically distance, keeping 1.5m between each other at all times, and more strongly, many countries enforced societal lockdowns where people were only allowed to leave their homes for specific reasons. Such interventions had large implications for mental health and physical health (Morina et al., 2021).

Nonetheless, many public health officials decided that reducing the transmission of COVID-19 was more pressing because public health systems were not prepared for such high and urgent disease burdens. In hospitals, there were shortages of many key resources for providing proper healthcare. Countries across the globe were scrambling for personal protective equipment, there were too few intensive care beds, and there were staff shortages (many of the support staff, nurses, and doctors experienced burn out, or got COVID-19 themselves, after which they needed to quarantine). It was thus important for reducing the burden of COVID-19 on the healthcare system and the general population that adherence to public health guidelines was adequate. Very few of the required *protective behaviours* (i.e., protective against viral transmission) could be stimulated through incentives, much less effectively policed. Instead, political leaders often opted for frequent press conferences in which they gave updates on the situation in their area and pleaded for compliance with public health guidelines.

On a personal level, I was surprised by how radically life changed in such a short time period. I felt that if I had trouble adapting to this new way of life, others would too. After contemplating what this meant for my PhD programme and my research in general,

my main supervisor at Monash University, Jakob Hohwy, and I discussed whether we would be able to contribute anything meaningful to the management of the pandemic scientifically. The next three Chapters are part of this effort. This COVID-19 research prompted interesting new research collaborations, other projects (e.g., chatbot research, see van Baal, Le, Fatehi, Hohwy, et al., 2022; van Baal, Le, Fatehi, Verdejo-Garcia, et al., 2022), and a deeper appreciation on my part of how intertwined behavioural science and public health are.

In the next two Chapters, I investigate the effectiveness of simple communication-based manipulations to increase people's compliance with public health guidelines. The current Chapter will focus on investigating experimentally whether two particular messages are effective at bringing about increases in individual-level behaviours that diminish the chance of viral transmission (i.e., "protective behaviours"). Participants receive prompts designed to stimulate compassion and contemplation about the future.

Similar to Chapter 3, I measured the intensity of participants' urges and the probability that they control those urges, but this time the urges we measured pertain to protective behaviours (e.g., the urge to leave home). In addition, for this Chapter, we are not only interested in how people experience and deal with urges, but we are also interested in the mechanisms through which our prompts can influence behaviour. In terms of the model developed in Chapter 1, we are interested in whether, if these prompts work, they desensitise urges through changing their potency (i.e., an attenuation of the relationship between the reward valuation and behaviour), or whether they increase self-control.

This Chapter has been published.

Episodic future thinking and compassion reduce public health guideline noncompliance urges: A randomised controlled trial

Simon T. van Baal, Antonio Verdejo-García, Jakob Hohwy

Abstract

Background

People often feel urges to engage in activities that violate pandemic public health guidelines. Research on these urges has been reliant on measures of typical behaviour, which fail to capture these urges as they unfold. Guideline adherence could be improved through interventions, but few methods allow for ecologically valid observation of the range of behaviours that pandemic guidelines prescribe.

Methods

In this preregistered parallel randomised trial, 95 participants aged 18-65 from the UK were assigned to three groups using blinded block randomisation, and engaged in episodic future thinking (n = 33), compassion exercises (n = 31), or a control procedure (n = 31). Following an ecological momentary assessment procedure, participants report on the intensity of their occurrent urges (min. 1, max. 10) and their ability to control them. The study further investigates whether, and through which mechanism, state impulsivity and vaccine attitudes affect guideline adherence.

Results

Episodic future thinking (b = -1.80) and compassion exercises (b = -1.45) reduced the intensity of urges. State impulsivity is associated with stronger urges, but we found no evidence that vaccine hesitancy predicts lesser self-control.

Conclusions

We conclude that episodic future thinking exercises and compassion training may be used to decrease non-compliance urges of individuals who are an acute public health risk for the community, such as those in voluntary isolation.

keywords: impulsivity, episodic future thinking, compassion, self-control, urges, pandemic

4.2 Introduction

It is challenging to study predictors and methods for improving pandemic public health guideline adherence because such behaviour is not readily observed in laboratory settings, nor easy to reveal with self-report cross-sectional surveys. Using ecological momentary assessment rather than one-shot surveys, the focus of this study is to find whether episodic future thinking and compassion exercises could contribute to increasing adherence to public health guidelines for preventing COVID-19 spread. We also investigate whether state impulsivity and vaccine attitudes predict guideline adherence, while assessing through which mechanism these predictors affect behaviour.

Much research focuses on designing public health communication to achieve optimal public health guideline adherence (Bavel et al., 2020). Protective behaviours such as staying home during a lockdown can have immediate adverse impact on people's financial situation (U.S. Department of Labor, 2021), mental health (Barari et al., 2020; Niedzwiedz et al., 2021; Pieh et al., 2021), and physical health (Constandt et al., 2020; Robinson et al., 2021). In contrast, the effects of non-adherence are often less immediate: it may take time before symptoms from infection and accompanying negative consequences are experienced; there may be subsequent effects on others rather than oneself, such as by infecting loved ones or causing outbreaks in the community. Decisions on adopting protective behaviours therefore constitute a dilemma between choosing the long-term greater good versus the short-term individual gain. Here we test if increasing people's future-orientedness and compassion can stimulate the adoption of protective behaviours during a pandemic.

To increase future-orientedness, we use episodic future thinking (EFT): imagining or simulating experiences that might occur in one's future. EFT decreases the degree to which rewards are devalued if they are received further in the future, known as delay discounting (Benoit et al., 2011; Peters & Büchel, 2010), which implies that the perceived value of immediate rewards will be diminished relative to future rewards. This means that EFT likely affects the intensity of urges, though the effects of EFT in various

domains suggest that EFT might also impact self-control independently of the strength of urges (Snider et al., 2016; Stein et al., 2017; Stein et al., 2016).

Adopting protective behaviours is ultimately also prosocial, and prosocial behaviour can be enhanced by stimulating compassion, the feeling that arises in witnessing another's suffering and that motivates a subsequent desire to help (Goetz et al., 2010). Compassion training has a valuation element in addition to a behavioural element, which means compassion training could affect both the intensity of urges and self-control (Omoto et al., 2009; Weng et al., 2013), both of which this study will investigate, in the context of urges of non-adherence to protective behaviours in a pandemic.

In addition, vaccine hesitancy, attitudes on the effectiveness of these vaccines, and predictions about how soon the pandemic will end could factor into people's behaviour. These attitudes are usually linked to other attitudes and behaviours relevant to pandemic behaviour such as lesser social distancing and mask-wearing (Latkin et al., 2021; Romer & Jamieson, 2020). We therefore also investigate how such predictors of guideline adherence influence moment-to-moment behaviour.

Impulsivity, the tendency to make rapid responses for short-term gratification and with insufficient regard for negative consequences (Moeller et al., 2001), is negatively correlated to public health guidelines adherence (Kooistra & van Rooij, 2020; Wismans et al., 2021). The steep delay discounting characteristic of high impulsivity can be influenced by fluctuations in internal states (Odum & Baumann, 2010). To understand how impulsivity affects moment-to-moment behaviour, it is important to gauge people's mental state when behaviours occur (Curran & Bauer, 2011; Fisher et al., 2018). Thus, we study these behaviours in an ecologically valid manner, where proximal information on state impulsivity is obtained.

This study seeks to avoid the distortions that often afflict self-report measures about typical behaviour (Althubaiti, 2016). To gain insights into moment-to-moment protective behaviour in the 'wild' and real-time changes following behaviour change interventions, we employ an ecological momentary assessment (EMA; or experience sampling) paradigm.

In our preregistered analyses, we predicted that both the compassion intervention and the episodic future thinking intervention would increase the likelihood of controlling

urges, where we analyse their effect on self-control and the intensity of urges. Furthermore, we predicted that state impulsivity is associated with stronger urges and fewer attempts to resist urges, and that vaccine hesitancy and shorter predicted back-to-normal time frames are negatively correlated with the likelihood of controlling an urge through diminishing self-control.

4.3 Methods

4.3.1 Design

This study used an individually randomised parallel group trial with three groups: EFT, Compassion, and Control. In a single-blind procedure, participants were assigned to a group after they completed an eligibility assessment, which was also when baseline data was collected. By responding to survey prompts on their mobile phones, participants provided up to 5 repeated measurements throughout the day – maximally 35 surveys completed throughout the one-week long experiment. This study was preregistered on 21/03/2021 on the Open Science Framework and registered as a clinical trial at clinicaltrials.gov on 02/09/2021, clinical trial ID: NCT05031559.

4.3.2 Participants

The final sample contained 95 UK residents, recruited using volunteer sampling through the Prolific participant platform. Participants were 18-65 years of age, and within that age group, we created a representative sample based on sex and age (2x4), with age strata 18-29, 30-41, 42-53, 54-65, (e.g., 18–29-year-old females).

Participants (n = 293) completed an eligibility survey prior to the experiment, which was used to create a representative sample. Participants were added to the EMA software in two rounds in order to cope with varying drop-out across the eight demographic strata, since there was more dropout than anticipated during the EMA app download-phase.

Participants were asked whether they or one of their family members were part of a group that is vulnerable to COVID-19. Participants also answered questions on their willingness to take a COVID-19 vaccine and their beliefs about vaccine efficacy. We

also elicited predictions of when people would be able to resume on-site work (insofar as that will ultimately be the case), when people would be able to go on holiday, and when life would go back to 'normal'. These predictions and vaccination attitudes were then combined into three scores, varying from 0 to 10.

Notably, the UK government announced on 29 March that 6 people from 6 different households would be allowed to meet outside. This, together with Easter weekend, produced a situation wherein people were likely to have non-adherence urges.

4.3.3 Randomisation and Masking

Groups were assigned according to block randomisation (8 strata, 3 groups; sequence obtained from sealedenvelope.com), see Supplementary Materials for the distribution of age and sex per group. Participants were unaware of their condition assignment, but the experimenter was (considering that they had to be added to the appropriate group in the software). Participants were either assigned to the EFT condition, the Compassion condition, or the Control condition.

4.3.4 Procedure

Each morning at 7.30am (expiry time 10am), participants would be asked to do either an EFT exercise, a compassion exercise, or reflect on a recent news story related to COVID-19. For the EFT exercise, participants were asked to imagine themselves in a future without COVID-19, for example, where they were travelling without restrictions. For the compassion exercise, participants were asked to imagine the suffering of individuals who were badly affected by COVID-19, for example, through the loss of family members. The news stories of COVID-19 concerned the negative impacts of COVID-19 on society (e.g., public health, business). All prompts are included in the Supplementary Materials (Table S1). After each group-dependent prompt, participants would be prompted with "Remember that your behaviour has an effect on the COVID-19 situation". Videos of the user interface are available.

Each day, after the morning survey, participants would receive 5 surveys that were available for 1 hour. In randomised order, they were asked whether since the last survey they had felt an urge to *not* wash their hands, *not* cover their mouths when coughing or

sneezing, not socially distance (e.g., to hug, shake hands), *not* leave details for contact tracing, or whether they had felt an urge to leave their house, touch their face, or avoid getting tested when it would have been better to do the opposite (from a COVID-19 standpoint). Participants responded using a slider [0,10], where 0 indicated no urge, 1 indicated a very weak urge, and 10 indicated a very strong urge. We then administered the Momentary Impulsivity Scale (Tomko et al., 2014).

4.3.5 Analysis

To determine sample size, we estimated an effect size of a 5-percentage point increase in the probability to control an urge in the EFT and Compassion groups, and assumed that participants would indicate they had an urge 3 times a day. We identified that 95% power, under these assumptions, could be achieved by collecting data from 90 participants. The power analysis is publicly available.

The intensity of urges was modelled using a cumulative link mixed model (CLMM) with a logit link, using the `ordinal` R package (Christensen, 2018). The intensity of the urge was entered directly into the model—no averaging was conducted—and we included the following predictors: the group, the type of urge (and interaction between those), state impulsivity, with age, sex, the time of day and the day of the week as control variables, and the participant as the random intercept.

The type of urge was not included as a predictor in the main preregistered models, but it was specified in the exploratory analyses section and because the effects of the intervention might differ across domains, we decided to include it in the main analyses. In both analyses, we decided to deviate from the preregistration by including interaction terms between the type of urge and the intervention because we deemed it likely that the intervention might affect some urges more than others. The preregistration also specified the use of a linear mixed model, but due to the ordinal nature of the response variable, we deviated from this plan and conducted the analysis using a CLMM.

We used a binomial generalised linear mixed model (GLMM) to conduct the self-control analysis. In addition to the variables in the model above, vaccine hesitancy, vaccine

effectiveness beliefs, back-to-normal timeline predictions, and whether participants attempted to resist the urge were included. The analysis was conducted using the `afex` package (Singmann et al., 2015).

A false discovery rate adjusted alpha of .05 was used to determine whether the effects based on the CLMM or on the GLMM were significantly different from those expected if the null hypothesis were correct. The `emmeans` package (Lenth et al., 2018) was used to conduct pairwise tests between factor levels; differences noted for the CLMM are on the latent scale (where the scale and location are arbitrary), while the differences for the GLMM are odds ratios. For numeric predictors, we zero-centered predictors to facilitate the interpretation of coefficients and differences between factor levels, and we used model coefficients to assess significance. For categorical predictors, we used sum contrast coding to compare the effects of independent variables against the grand mean. Effect sizes are reported as odds ratios. Effect sizes are reported as odds ratios.

Our analyses were preregistered, data and analysis are available (van Baal et al., 2021).

4.4 Results

The experiment took place from 29 March to 4 April 2021. The UK was in a state of lockdown then, but most regions in the UK were in the early phases of reopening. In total, 200 participants were added to the EMA software, 112 of whom downloaded the app, and 97 of those completed more than 50% of the EMA surveys. As indicated in our preregistration, participants who completed less than 50% of EMA surveys were excluded. Finally, two participants never reported having urges of non-adherence, and thus these were excluded from the final sample. In the final sample, there were 40 (42.1%) males, with a mean age of 41.0 (SD = 14.0), and 55 females (57.9%) with a mean age of 41.0 (SD = 12.5). See Table S1 for demographic information per group; see Figure S2 for a CONSORT diagram describing the flow of the recruitment process.

Table S1: Demographic information of the sample, per group.

Condition	Age: Mean (SD)	Male	Vaccinated	Covid
Compassion	41.8 (13.5)	35.5%	41.9%	6.5%
Control	39.8 (13.2)	41.9%	45.2%	3.2%
EFT	41.8 (13)	48.5%	39.4%	6.1%

Further, 40 participants (42.1%) reported they had received a COVID-19 vaccine, and 5 reported that they had been diagnosed with COVID-19 at some point, with one participant reporting they had experienced both of these events. Even though the individual risk of COVID-19 is mitigated for vaccinated individuals and those who previously contracted COVID-19, they were still required to comply with the guidelines for various reasons. Therefore, vaccinated individuals were not excluded.

Missing data in EMA study designs usually occurs survey-wise (Sun et al., 2021), which is also the case in our study—8.7% of the data collection surveys were missing in the final sample, with only 0.9% of the data collection surveys missing individual values. Many participants missed surveys on Monday because they were still familiarising with the software. Missing data for the data collection surveys was disproportionately concentrated in the responses of female participants (10.2% missing data for females; 7.3% for males). Moreover, participants with missing data tended to be slightly older (M_{age} = 43.7 for missing data entries; M_{age} = 40.2 for non-missing data entries). See Table S2 for a table describing the structure of the data, and further details on the missing data.

Without the access to other momentary information that correlates with the outcome variables, this pattern of missingness largely prohibits the use of modern data methods to eliminate bias or improve statistical power. We did not find evidence the day of the week, the sex or age of participants impacted the outcome variables, so there appears to be no need to assume the pattern of missingness in this study gives reason to be concerned about biased estimates. Therefore, we conclude that the data is missing at random (MAR), and these variables are included in the statistical models so the risk of biased estimates is limited (McLean et al., 2017). The other 1.1% of surveys containing missing data only missed values on the MIS, most likely due to a software error. We found no observable patterns in the other variables for these missing values, so here too we will assume these data are MAR.

Different types of urges occurred at different rates: over the one-week-long experiment, 83 out of the 95 participants reported the urge to leave the house at least once, and did so 6.80 times on average (SD = 6.09), 79 reported the urge to touch their face (M = 7.91, SD = 8.14), 73 Participants reported the urge to disregard social distancing guidelines (M = 4.66, SD = 4.37), while 60 reported the urge to not wash their hands (M = 5.03, SD = 6.11). Only 33 participants reported the urge to not cover their mouth

($M = 2.82$, $SD = 4.39$), 21 participants reported the urge to not leave their contact details ($M = 3.86$, $SD = 7.18$), and 17 participants reported the urge to avoid getting tested ($M = 4.82$, $SD = 7.77$).

There was high variance in the number of urges people experienced, and the average number of urges experienced was similar over the different groups: in the EFT group, people had 20.0 urges on average ($SD = 30.2$) in the Compassion group, people had 23.1 urges on average ($SD = 16.3$) and 23.0 ($SD = 26.7$) in the Control group.

The various types of urges were different in their intensity and in their controllability. We report these differences partitioned by group in Table S2, but we also report general differences in the Supplementary Materials.

TABLE S2: The intensity of urges and the probability that participants controlled them, by allocation group.

Type of urge	Condition	Urge intensity	Prob. control
Not to cover mouth/nose	EFT	2.11 (0.92)	0.77 (0.92)
	Compassion	2.72 (1.3)	0.7 (1.3)
	Control	1.85 (0.43)	0.78 (0.43)
To leave the house	EFT	4.13 (2.18)	0.45 (2.18)
	Compassion	4.25 (1.89)	0.44 (1.89)
	Control	5.06 (1.55)	0.43 (1.55)
Not to provide contact tracing details	EFT	1.31 (0.66)	0.8 (0.66)
	Compassion	1.78 (0.62)	0.52 (0.62)
	Control	4 (0.94)	0.5 (0.94)
Not to physically distance	EFT	4.34 (1.86)	0.57 (1.86)
	Compassion	3.6 (1.63)	0.67 (1.63)
	Control	3.46 (1.58)	0.74 (1.58)
Not to get tested	EFT	1.58 (1.28)	0.21 (1.28)
	Compassion	2.02 (1.09)	0.2 (1.09)
	Control	5.83 (0.38)	0.12 (0.38)
To touch one's face	EFT	2.93 (1.55)	0.27 (1.55)
	Compassion	3.27 (1.43)	0.39 (1.43)
	Control	3.45 (1.59)	0.51 (1.59)
Not to wash hands	EFT	2.74 (1.46)	0.77 (1.46)
	Compassion	2.81 (1.43)	0.75 (1.43)
	Control	3.22 (1.3)	0.72 (1.3)

4.4.1 Predictors for Urge Intensity

Participants experienced weaker urges in the EFT group, b = -1.798, 95% CI [-2.923, -.672], z = -3.824, p <.001, and in the Compassion group, b = -1.449, 95% CI [-2.580, -.317], z = -3.064, p <.01, than in the Control group. This means that, for instance, the average predicted odds of reporting a stronger urge in the Control condition than in the EFT condition across all urges, time points, and days was 1.798:1. Urge intensity was not significantly different in the EFT group from the Compassion group (z = .742, p = .46). See Figure S1.

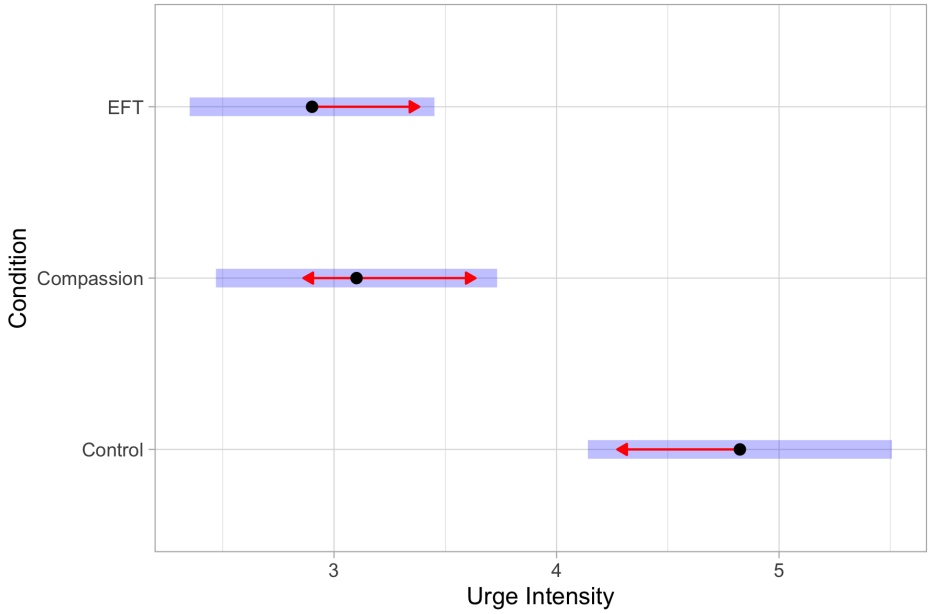

FIGURE S1: The effects of the between-participants conditions: Episodic Future Thinking manipulation (EFT; top of figure), the Compassion manipulation (middle), as compared to the Control condition (bottom), on the perceived intensity of urges (x-axis, location and scale are arbitrary). The points are estimated marginal means, the red arrows are comparison arrows reflecting the pairwise tests, and the error bars are 95% CIs.

There were also interactions between group and type of urge: urges to avoid leaving details for contact tracing were weaker in the EFT group, b = -5.737, 95% CI [-7.598, -3.876]; z = 7.025, p < .0001, and in the Compassion group, b = -4.871, 95% CI [-6.776, -2.965]; z = 6.161, p < .0001, than in the Control group, but were not significantly different from each other, z = 1.021, p = .31. Further, urges to avoid getting tested were also weaker in the EFT group, b = -6.833, 95% CI [-8.879, -4.787]; z = 7.995, p < .0001, and in the Compassion group, b = -5.384, 95% CI [-7.396, -3.372]; z = 6.407, p

$< .0001$, than in the Control group, but were not significantly different from each other, $b = 1.449$, 95% CI [-.403, 3.302], $z = 1.873$, $p = .06$. See Figure S2.

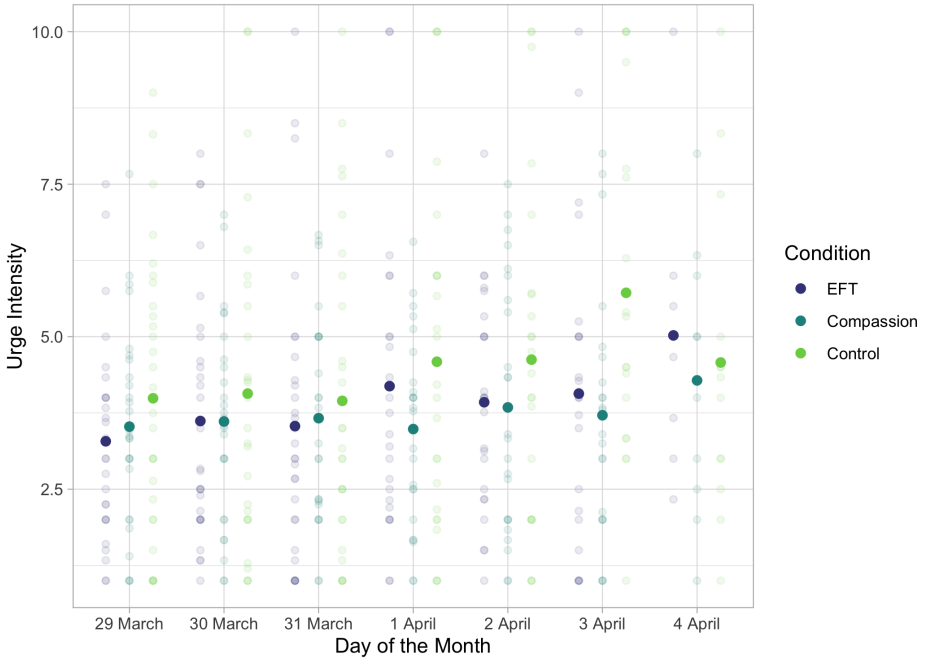

FIGURE S2: The intensity of urges (y-axis), partitioned by day (x-axis), and group (colour). The experiment was partially conducted over a public holiday, at which time stronger urges of non-adherence would be expected. 2 April (third line from the right) was Good Friday, and 4 April (the right-most line) was Easter Sunday. The coloured points represent the estimated marginal means, and the error bars are 95% CIs. The grey data points each represent the aggregated data of one participant.

State impulsivity had a significant effect on the intensity of urges, $b = .362$ 95% CI [0.207, .518], $z = 4.580$, $p < .001$. See Figure S3 for a depiction of the relationship between state impulsivity and the intensity of urges.

4.4.2 Predictors for Self-Control

There were no significant differences in self-control between the groups, all $ps > .1$.

There were also interactions between the type of urge and the group: people were significantly more likely suppress the urge to not cover their mouth when coughing or sneezing in the Control group than in the Compassion group, OR = 16.613, 95% CI [1.56, 177.58], $z = 2.832$, $p = .01$, but not than in the EFT group, OR = 3.403, 95% CI [.24, 47.57], $z = 1.087$, $p = .27$. People were also more likely to suppress the urge to touch their face in the Control group than in the Compassion group, OR = 4.085, 95%

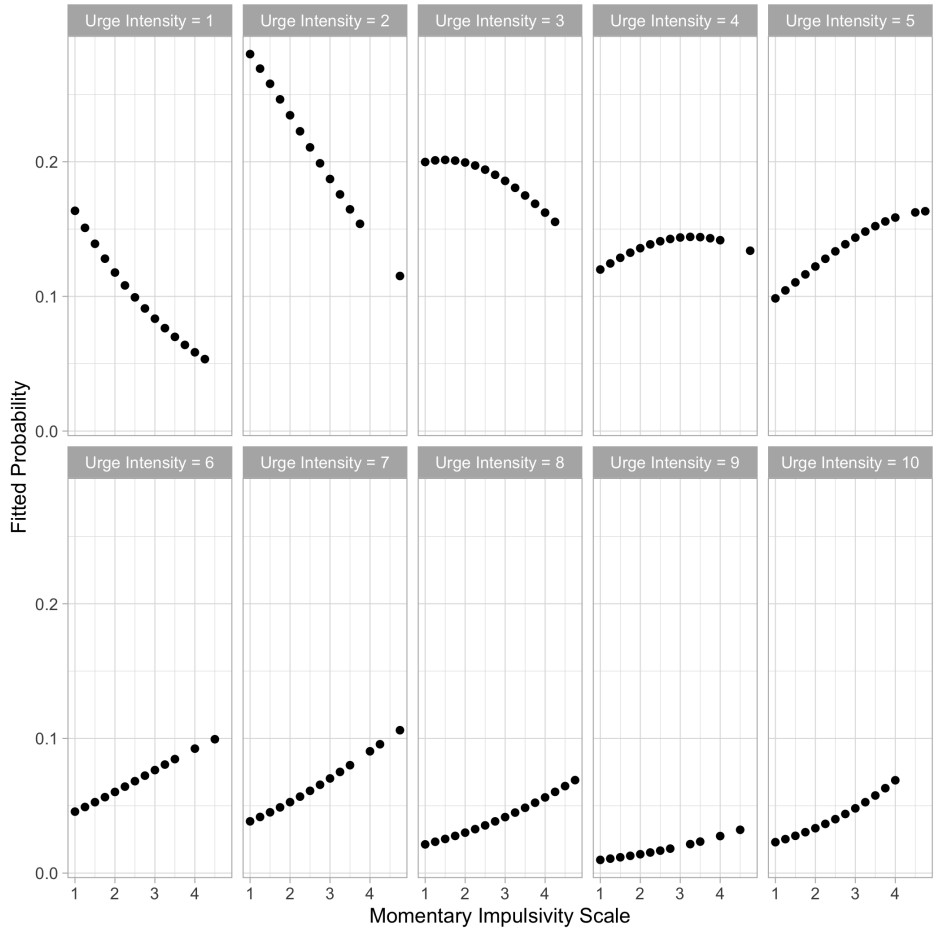

FIGURE S3: The differences in the intensity of urges (left), and the ability to control them (right) between the types of urges measured (y-axis). From top to bottom, the urges are: to not wash one's hands, to touch one's face, to avoid getting tested, to not socially distance oneself, to skip leaving contact details when entering an establishment, to leave the house, and to sneeze/cough without covering one's nose/mouth. The points are estimated marginal means, the red arrows are comparison arrows reflecting the pairwise tests, and the error bars are 95% CIs.

CI [1.47, 11.39], $z = 3.286$, $p < .01$, and in the EFT group, OR = 3.198, 95% CI [1.12, 9.15], $z = 2.648$, $p = .01$.

State impulsivity did not have a significant effect on self-control, b = .045, 95% CI [-.293, .203], $z = .0356$, $p = .72$. Furthermore, state impulsivity did not have a significant effect on the probability of attempting to resist an urge b = -.103, 95% CI [-.320, .114], z -.928, $p = .35$

Neither vaccine hesitancy nor judgments about vaccine effectiveness significantly predicted self-control, b = -.098, 95% CI [-.262, .066], z = -1.173, $p = .24$; b = .038, 95% CI [-.213, .229], z = .070, $p = .94$. Neither did predictions about when life would go back to 'normal' after the pandemic, b = -.141, 95% CI [-.726, .483]; z = -.393, $p = .69$.

4.5 Conclusion

This study recorded the intensity and controllability of various types of urges pertinent to pandemic management. We were able to measure people's urges of non-adherence to protective behaviours during a pandemic without risking recall bias, by using ecological momentary assessment. Our findings show that episodic future thinking and compassion exercises reduced the intensity of urges to avoid protective behaviours, but did not affect self-control. We also found that different urges occur at widely varying rates within and between participants, which is an important consideration when assessing the relative impact of these urges.

Our findings show that episodic future thinking and compassion exercises reduced the intensity of certain urges, but we found no evidence that it affected self-control in our sample. This finding broadly aligns with the evidence that EFT can enhance future-oriented decision making in various contexts (Snider et al., 2016; Stein et al., 2017; Stein et al., 2016). Given that urges usually pertain to immediate rewards, this reduction in the strength of urges after an EFT exercise is most likely because EFT reduces the relative value of immediate rewards compared to future rewards (Benoit et al., 2011; Peters & Büchel, 2010).

The mechanisms through which EFT and compassion exercises affect behaviour differ: EFT enhances future-oriented decision making in various contexts (Snider et al., 2016; Stein et al., 2017; Stein et al., 2016), supposedly by decreasing delay discounting (Benoit et al., 2011); compassion exercises can increase prosociality (Goetz et al., 2010), and there is also some evidence it might also increase future-oriented decision making (DeSteno et al., 2014). These mechanisms both likely promote the salience of the potential negative consequences to people's actions, which may be the reason for their effectiveness in this context. Alternative explanations include that compassion exercises can lead to increased positive affect and motivation (Dahl et al., 2016), and can be helpful to deal with daily stressors (Reddy et al., 2013), which may also decrease the perceived intensity of urges. These results suggest that an invitation to engage in EFT and compassion-inducing talking points could be incorporated into press conferences and some public announcements to decrease urges of noncompliance during public health crises. Additionally, people who pose a specific risk to the community (e.g.,

those in voluntary self-isolation after travelling abroad) could be invited to periodically perform such a task.

We found no evidence that vaccine attitudes or predictions of back-to-normal timelines were associated with self-control in our sample. Other studies show that vaccine hesitancy is correlated with lesser social distancing and mask wearing (Latkin et al., 2021), but most studies reporting these relationships rely on judgements about typical behaviour, or intentions to comply with guidelines. The lack of evidence for a significant relationship between vaccine attitudes and guideline adherence in this study suggests that more research is needed to understand how these attitudes affect moment-to-moment behaviour.

Different urges occurred at varying rates within and between participants, which is an important consideration when assessing the relative impact of these urges. If an urge is relatively rare, but difficult to control, then it may not be as relevant for policymakers and other key stakeholders to see if the probability to control this urge can be increased. Urges to avoid getting tested, or to not leave contact details were relatively infrequent, but the fact that around 20% of the sample reported one of these urges at least once is worrying given their importance to pandemic management (Stuart et al., 2021).

State impulsivity was related to stronger urges, but not to diminished self-control. This evidence suggests that state fluctuations in impulsivity play an important, but poorly understood role in determining public health guideline adherence during pandemics. A recent study has found this 'bottom-up' effect of state impulsivity for a different, more general domain of urges (van Baal, Moskovsky, et al., 2022). It also suggests that interventions targeting the internal state of the individual, and impulsivity in particular, might be effective at ameliorating their guideline adherence. Future research could investigate whether state impulsivity, as well as other internal states, can be targeted to improve public health guideline adherence.

4.5.1 Limitations and Future Directions

The main limitations of this study were that the heterogeneity of the experience of certain types of urges rendered the sample size too small to draw accurate inferences in some domains. Only around 20 individuals, spread over three groups, reported having

urges to avoid getting tested or to avoid leaving details for contact tracing at least once. Furthermore, the lack of evidence that the manipulations affected self-control could be due to a lack of power, rather than the absence of a meaningful effect. The assumptions we made for our power analysis were optimistic, especially because we did not account for participant clustering of responses.

The lack of power is also visible in the effect size estimates of the compassion manipulation's influence on the likelihood of covering one's nose and mouth while coughing or sneezing, where the spread of the confidence interval suggests near empty strata. Hence there remains considerable uncertainty about the effectiveness of the intervention in these domains.

This speaks to the strengths and weaknesses of the ecological momentary assessment paradigm because, on the one hand, it is a powerful paradigm for events that occur often and to a wide range of people (such as the urge to abandon social distancing), and it can provide insight into behaviour 'in the wild'. On the other hand, for events that only happen for a narrow subset of people, or that happen infrequently, ecological momentary assessment needs to be applied to that particular subset, or another approach should be considered. A further limitation concerns the extent to which the findings apply to different populations. Ecological momentary assessment is known to generate missing data, which creates uncertainty about effect size estimates and reduces overall data quality (Sun et al., 2021). In our study, we mitigated this cost of ecological approaches by incentivising high compliance and implementing a compliance threshold. Nonetheless, the effect size estimates we report should be considered in light of the uncertainty associated with missing data.

Another limitation of the current work is that we did not administer any baseline scales to investigate whether individual differences, such as impulsivity or compassion traits, predict responsiveness to the interventions. The lack of evidence for a relationship between the baseline variables we collected and self-control also need not generalise to the population because of our sample size limitations.

In the current work, participants were asked to engage with others' suffering and prompted to consider that their actions influence the situation they were asked to picture. We did not explore exactly how participants' states changed after the manipulations to avoid demand characteristics, so we cannot be sure about the exact mechanisms that caused

the observed effects. Future research could, for example, investigate whether only empathising with the suffering would be sufficient, or whether stronger language linking one's own behaviour to others' suffering would be more effective.

Future research could focus on the role of behavioural science interventions, including, but not limited to, episodic future thinking and compassion-inducing exercises, in producing desired behaviour during public health crises. In this line of research, it may be promising to also consider the combination of compassion and episodic future thinking, to see if the effects might interact positively. Further research could also address whether individual differences can predict responsiveness to this type of intervention. We did not administer any baseline scales to investigate whether individual differences, such as impulsivity or compassion traits, predict responsiveness to the interventions.

We also deem it important that more research is devoted to uncovering the factors predicting moment-to-moment decision-making in people's daily lives, where ecological momentary assessment and GPS data (Gollwitzer et al., 2020) could play a critical role. These methods can provide a more comprehensive understanding of the effects of behavioural interventions, shedding light on their longevity and externalities.

4.6 Declarations

4.6.1 Ethics Approval and Consent to Participate

We confirm that all experimental protocols were approved by the Monash University Human Research Ethics Committee (approval number 26253). We confirm that informed consent was obtained from all subjects. All methods were carried out in accordance with relevant guidelines and regulations.

4.6.2 Consent for Publication

All participants gave consent to the publication of the study.

4.6.3 Availability of Data and Materials

The datasets generated and/or analysed during the current study are available in the OSF repository.

4.6.4 Competing Interests

Authors declare no competing interests.

4.6.5 Funding

This work is supported by a Monash-Warwick Alliance Accelerator grant and by the Three Springs Foundation. The funding sources were not involved in the research conducted.

4.6.6 Acknowledgements

The authors would like to thank Dr. Peter Koval for the use of the SEMA3 platform.

4.7 References

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, *9*, 211–217. https://doi.org/10.2147/JMDH.S104807

Barari, S., Caria, S., Davola, A., Falco, P., Fetzer, T., Fiorin, S., Hensel, L., Ivchenko, A., Jachimowicz, J., King, G., Kraft-Todd, G., Ledda, A., MacLennan, M., Mutoi, L., Pagani, C., Reutskaja, E., & Slepoi, F. R. (2020). Evaluating COVID-19 Public Health Messaging in Italy: Self-Reported Compliance and Growing Mental Health Concerns. *medRxiv*, 2020.03.27.20042820. https://doi.org/10.1101/2020.03.27.20042820

Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., ... Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), 460–471. https://doi.org/10.1038/s41562-020-0884-z

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *Journal of Neuroscience*, *31*(18), 6771–6779.

Berry, G., Parsons, A., Morgan, M., Rickert, J., & Cho, H. (2022). A review of methods to reduce the probability of the airborne spread of COVID-19 in ventilation systems and enclosed spaces. *Environmental Research*, *203*, 111765. https://doi.org/10.1016/j.envres.2021.111765

Chater, N., & Loewenstein, G. (2022). The i-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray. https://doi.org/10.2139/ssrn.4046264

Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the R package ordinal. *Submitted in J. Stat. Software*.

Constandt, B., Thibaut, E., De Bosscher, V., Scheerder, J., Ricour, M., & Willem, A. (2020). Exercising in times of lockdown: An analysis of the impact of COVID-19 on levels and patterns of exercise among adults in Belgium. *International journal of environmental research and public health*, *17*(11), 4144. https://doi.org/10.3390/ijerph17114144

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Dahl, C. J., Lutz, A., & Davidson, R. J. (2016). Cognitive processes are central in compassion meditation. *Trends in Cognitive Sciences*, *20*(3), 161–162. https://doi.org/10.1016/j.tics.2015.12.005

DeSteno, D., Li, Y., Dickens, L., & Lerner, J. S. (2014). Gratitude: A Tool for Reducing Economic Impatience. *Psychological Science*, *25*(6), 1262–1267. https://doi.org/10.1177/0956797614529979

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological bulletin*, *136*(3), 351. https://doi.org/10.1037/a0018807

Gollwitzer, A., Martel, C., Marshall, J., Höhs, J. M., & Bargh, J. A. (2020). Connecting self-reported social distancing to real-world behavior at the individual and us state level. https://doi.org/10.31234/osf.io/kvnwp

Kooistra, E. B., & van Rooij, B. (2020). Pandemic Compliance: A systematic review of influences on social distancing behaviour during the first wave of the COVID-19 outbreak. *Available at SSRN 3738047*.

Latkin, C. A., Dayton, L., Yi, G., Colon, B., & Kong, X. (2021). Mask usage, social distancing, racial, and gender correlates of COVID-19 vaccine intentions among adults in the US. *PloS one*, *16*(2), e0246970. https://doi.org/10.1371/journal.pone.0246970

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.

McLean, D. C., Nakamura, J., & Csikszentmihalyi, M. (2017). Explaining System Missing: Missing Data and Experience Sampling Method. *Social Psychological and Personality Science*, *8*(4), 434–441. https://doi.org/10.1177/1948550617708015

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American journal of psychiatry*, *158*(11), 1783–1793. https://doi.org/10.1176/appi.ajp.158.11.1783

Morina, N., Kip, A., Hoppen, T. H., Priebe, S., & Meyer, T. (2021). Potential impact of physical distancing on physical and mental health: A rapid narrative umbrella review of meta-analyses on the link between social connection and health - PMC. *BMJ Open*, *11*(3). https://doi.org/10.1136/bmjopen-2020-042335

Niedzwiedz, C. L., Green, M. J., Benzeval, M., Campbell, D., Craig, P., Demou, E., Leyland, A., Pearce, A., Thomson, R., & Whitley, E. (2021). Mental health and health behaviours before and during the initial phase of the COVID-19 lockdown:

Longitudinal analyses of the UK Household Longitudinal Study. *J Epidemiol Community Health*, *75*(3), 224–231. https://doi.org/10.1136/jech-2020-215060

Odum, A. L., & Baumann, A. A. (2010). Delay discounting: State and trait variable. In *Impulsivity: The behavioral and neurological science of discounting.* (pp. 39–65). American Psychological Association.

Omoto, A. M., Malsch, A. M., & Barraza, J. A. (2009). Compassionate acts: Motivations for and correlates of volunteerism among older adults. *The science of compassionate love: Theory, research, and applications*, 257–282.

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, *66*(1), 138–148. https://doi.org/10.1016/j.neuron.2010.03.026

Pieh, C., Budimir, S., Delgadillo, J., Barkham, M., Fontaine, J. R., & Probst, T. (2021). Mental health during COVID-19 lockdown in the United Kingdom. *Psychosomatic medicine*, *83*(4), 328–337. https://doi.org/10.1097/PSY.0000000000000871

Reddy, S. D., Negi, L. T., Dodson-Lavelle, B., Ozawa-de Silva, B., Pace, T. W., Cole, S. P., Raison, C. L., & Craighead, L. W. (2013). Cognitive-based compassion training: A promising prevention strategy for at-risk adolescents. *Journal of Child and Family Studies*, *22*(2), 219–230. https://doi.org/10.1007/s10826-012-9571-7

Robinson, E., Boyland, E., Chisholm, A., Harrold, J., Maloney, N. G., Marty, L., Mead, B. R., Noonan, R., & Hardman, C. A. (2021). Obesity, eating behavior and physical activity during COVID-19 lockdown: A study of UK adults. *Appetite*, *156*, 104853. https://doi.org/10.1016/j.appet.2020.104853

Romer, D., & Jamieson, K. H. (2020). Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. *Social Science & Medicine*, *263*, 113356. https://doi.org/10.1016/j.socscimed.2020.113356

Singer, J. (2022). *There are no accidents: The deadly rise of injury and disaster—who profits and who pays the price*. Simon; Schuster.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). Afex: Analysis of factorial experiments. *R package version 0.13–145*.

Snider, S. E., LaConte, S. M., & Bickel, W. K. (2016). Episodic future thinking: Expansion of the temporal window in individuals with alcohol dependence. *Alcoholism:*

clinical and experimental research, *40*(7), 1558–1566. https://doi.org/10.1111/acer.13112

Stein, J. S., Sze, Y. Y., Athamneh, L., Koffarnus, M. N., Epstein, L. H., & Bickel, W. K. (2017). Think fast: Rapid assessment of the effects of episodic future thinking on delay discounting in overweight/obese participants. *J Behav Med*, *40*(5), 832–838. https://doi.org/10.1007/s10865-017-9857-8

Stein, J. S., Wilson, A. G., Koffarnus, M. N., Daniel, T. O., Epstein, L. H., & Bickel, W. K. (2016). Unstuck in time: Episodic future thinking reduces delay discounting and cigarette smoking. *Psychopharmacology*, *233*(21), 3771–3778. https://doi.org/10.1007/s00213-016-4410-y

Stuart, R. M., Abeysuriya, R. G., Kerr, C. C., Mistry, D., Klein, D. J., Gray, R. T., Hellard, M., & Scott, N. (2021). Role of masks, testing and contact tracing in preventing COVID-19 resurgences: A case study from New South Wales, Australia. *BMJ open*, *11*(4), e045941. https://doi.org/10.1136/bmjopen-2020-045941

Sun, J., Rhemtulla, M., & Vazire, S. (2021). Eavesdropping on Missing Data: What Are University Students Doing When They Miss Experience Sampling Reports? *Personality and Social Psychology Bulletin*, *47*(11), 1535–1549. https://doi.org/10.1177/0146167220964639

THE EMPLOYMENT SITUATION — MAY 2021 (tech. rep. USDL-21-0980). (2021). U.S. Department of Labor.

Tomko, R. L., Solhan, M. B., Carpenter, R. W., Brown, W. C., Jahng, S., Wood, P. K., & Trull, T. J. (2014). Measuring impulsivity in daily life: The momentary impulsivity scale. *Psychological assessment*, *26*(2), 339. https://doi.org/10.1037/a0035083

van Baal, S. T., Hohwy, J., & Verdejo-Garcia, A. (2021). Resisting urges of non-compliance during lockdown in the pandemic: Data & Code. https://doi.org/10.17605/OSF.IO/PJ9FT

van Baal, S. T., Le, S. T. T., Fatehi, F., Hohwy, J., & Verdejo-Garcia, A. (2022). Cory COVID-Bot: An Evidence-Based Behavior Change Chatbot for COVID-19. *Informatics and Technology in Clinical Care and Public Health*, 422–425. https://doi.org/10.3233/SHTI210948

van Baal, S. T., Le, S. T. T., Fatehi, F., Verdejo-Garcia, A., & Hohwy, J. (2022). Effecting behaviour change using an artificial intelligence chatbot: A pilot randomised controlled study. https://doi.org/10.31234/osf.io/2xuat

van Baal, S. T., Moskovsky, N., Hohwy, J., & Verdejo-García, A. (2022). State impulsivity amplifies urges without diminishing self-control. *Addictive Behaviors*, *133*, 107381. https://doi.org/10.1016/j.addbeh.2022.107381

Weng, H. Y., Fox, A. S., Shackman, A. J., Stodola, D. E., Caldwell, J. Z., Olson, M. C., Rogers, G. M., & Davidson, R. J. (2013). Compassion training alters altruism and neural responses to suffering. *Psychological science*, *24*(7), 1171–1180. https://doi.org/10.1177/0956797612469537

Wismans, A., Letina, S., Wennberg, K., Thurik, R., Baptista, R., Burke, A., Dejardin, M., Janssen, F., Santarelli, E., & Torrès, O. (2021). The role of impulsivity and delay discounting in student compliance with COVID-19 protective measures. *Personality and Individual Differences*, 110925. https://doi.org/10.1016/j.paid.2021.110925

4.8 Supplementary Materials

Intensity and controllability of the types of urges Reported urges, in order of intensity, were: leaving the house (M = 4.46, SD = 1.86), not socially distancing (M = 3.79, SD = 1.69), touching one's face (M = 3.22, SD = 1.52), avoiding getting tested (M = 2.74, SD = .93), avoiding washing hands (M = 2.92, SD = 1.39), not leaving details for contact tracing (M = 2.60, SD = .83), and, weakest, not covering mouth and nose when coughing/sneezing (M = 2.23, SD = .93).

Reported success at controlling urges, in descending order of likelihood, was: not washing hands (M = .748, SD = .074); not covering their mouths (M = .747, SD = .057); not socially distancing (M = .662, SD = .076); not leaving contact details for contact tracing purposes (M = .581, SD = .008); leaving the house (M = .441, SD = .102); touching one's face (M = .390, SD = .095); avoiding getting tested (M = .186, SD = .038). It is important to note that these probabilities must be judged together with the relative frequencies of each urge. For instance, the probability of controlling an urge to avoid getting tested is extremely low, but these urges were infrequently experienced.

Differing dropout across groups after the second round of participant entries led to an imbalance in the female/male sample division. Here we display the age and sex distribution between groups (Figure S1).

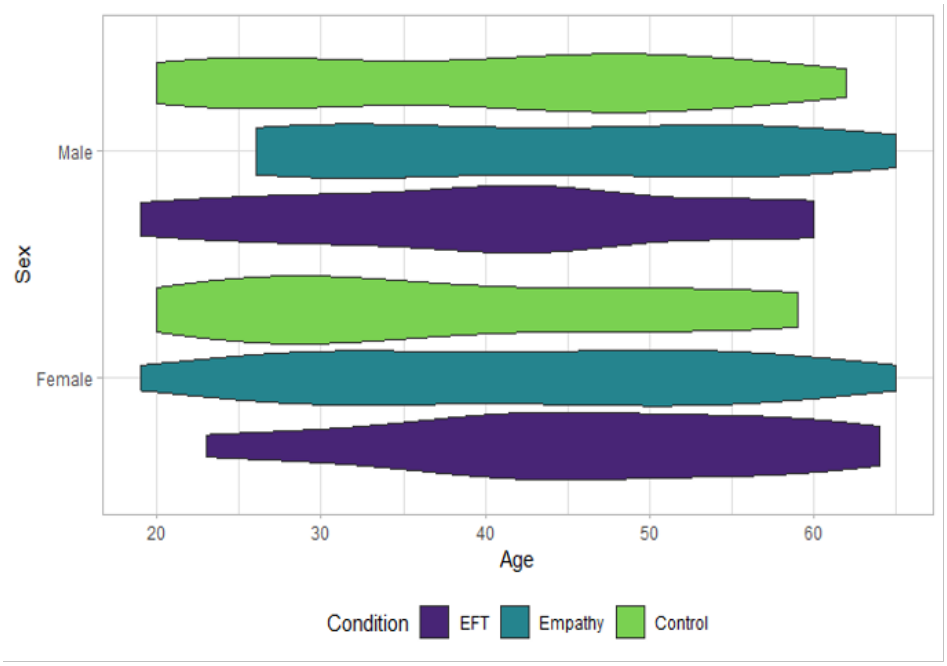

FIGURE S1: The age and sex distribution of the sample.

Below we also provide an adjusted CONSORT diagram to show the experimental procedure and the exclusions, see Figure S2. This diagram deviates from the standard CONSORT flow diagram because the design we used for the study differs substantially from the standard design in the medical sciences.

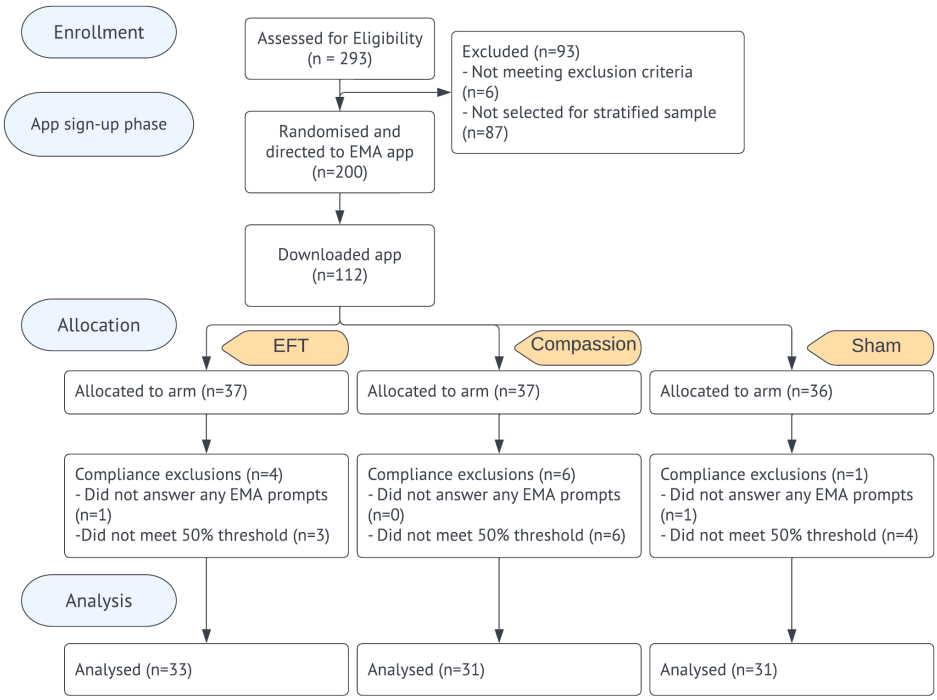

FIGURE S2: A consort diagram of the experimental procedure.

TABLE S1: The instructional prompts participants received each morning in their 7.30am survey.

Episodic Future Thinking	**Compassion**	**Control**
Please imagine yourself after lockdowns and restrictions are over. You are on holiday, at your favourite destination. Try to imagine how you feel, and picture your surroundings (think about smell and sounds too).	Please imagine yourself as a vulnerable person. You have severe asthma and will have significant trouble breathing if you catch COVID-19. Try to feel the emotions that might be going through you.	Please reflect on the news that shutdowns of the Ferguson shipyard during the Covid pandemic have added an extra £4.3m to the cost of two over-budget and delayed CalMac ferries.
Please imagine yourself in the future when lockdowns and restrictions are over. You are travelling to a place you always wanted to go. Imagine how you feel and picture your surroundings (think about smells and sound too).	Please imagine yourself as a family member of a person in hospital due to COVID-19. You want to be with them, but you are not allowed to because you might catch it yourself. Try to feel the emotions that might be going through you.	Please reflect on the news that the COVID-19 restrictions in Scotland were relaxed on 13 March. Up to four people from two different households can now gather outside.

Continued on next page

Table S1 – continued from previous page

Episodic Future Thinking	Compassion	Control
Please imagine yourself in the future when lockdowns and restrictions are over, and picture that you are doing something you love. Imagine how you feel and picture your surroundings (think about smell and sound too).	Please imagine yourself as a vulnerable person. You are an elderly person in a retirement home. You have not been able to receive any visitors for months, and your interactions with the other residents are restricted. Try to feel the emotions that might be going through you.	Please reflect on the fact that on 5 March, COVID-19 infections were rising in Europe again.
Please imagine yourself after lockdowns and restrictions are over. You are doing your favourite activity that became possible after lockdowns lifted. Imagine how you feel, and picture your surroundings (think about smell and sound too).	Please imagine yourself as a healthcare worker in the emergency room. You have not been able to get proper sleep and have had to keep families separated because it was too dangerous to see their loved ones. Try to feel the emotions that might be going through you.	Please reflect on the reporting from 14 March that many people in Lincolnshire are not turning up for their vaccinations.
		Please reflect on the news that on 14 March people were fined 800 pounds each for having a party in Gloucester.
		Please reflect on the news that 48,000 businesses have signed up for workplace testing.
		Please reflect on the news that 28% of people in need of social care have seen their health decline during COVID-19.

TABLE S2: Data structure information. This table shows the number of observations of variables segmented by whether an urge was reported by the participant or not. It also provides information about missing values. When a variable is numeric, the information provided in each cell is: mean (standard deviation); when it is a factor the information provided is: count (percentage).

Label	Total N	Missing N	Levels	No Urge	Urge reported	(Missing)	Total
Total N (%)				17770 (81.7)	2092 (9.6)	1894 (8.7)	21756
Urge magnitude	2092 (10.5)	17770	Mean (SD)	NA	4.2 (2.7)	NA	4.2 (2.7)
Urge control	2087 (10.5)	17775	Urge controlled	0 (0.0)	967 (46.2)	0 (0.0)	967 (4.4)
			Urge not controlled	0 (0.0)	1120 (53.5)	0 (0.0)	1120 (5.1)

Continued on next page

Table S2 – continued from previous page

Label	Total N	Missing N	Levels	No Urge	Urge reported	(Missing)	Total
			(Missing)	17770 (100.0)	5 (0.2)	1894 (100.0)	19669 (90.4)
MIS	19683 (99.1)	179	Mean (SD)	1.5 (0.6)	1.8 (0.7)	1.2 (NA)	1.5 (0.7)
Type of urge	19862 (100.0)	0	Cover mouth/nose	2746 (15.5)	93 (4.4)	269 (14.2)	3108 (14.3)
			Leave your house	2271 (12.8)	565 (27.0)	272 (14.4)	78 (14.3)
			No contact tracing details	2753 (15.5)	84 (4.0)	271 (14.3)	3108 (14.3)
			Not physical distancing	2496 (14.0)	340 (16.3)	272 (14.4)	3108 (14.3)
			Not getting tested	2754 (15.5)	83 (4.0)	271 (14.3)	3108 (14.3)
			Touch mouth or nose	2215 (12.5)	625 (29.9)	268 (14.1)	3108 (14.3)
			Not washing hands	2535 (14.3)	302 (14.4)	271 (14.3)	3108 (14.3)
Time of day	19862 (100.0)	0	Mean (SD)	15.1 (3.5)	14.8 (3.4)	15.0 (3.8)	15.0 (3.5)
Weekday	19862 (100.0)	0	Mon	1093 (6.2)	151 (7.2)	16 (0.8)	1260 (5.8)
			Tue	2707 (15.2)	431 (20.6)	12 (0.6)	3150 (14.5)
			Wed	2717 (15.3)	344 (16.4)	26 (1.4)	3087 (14.2)
			Thu	2804 (15.8)	331 (15.8)	22 (1.2)	3157 (14.5)
			Fri	2800 (15.8)	305 (14.6)	17 (0.9)	3122 (14.4)
			Sat	2813 (15.8)	265 (12.7)	23 (1.2)	3101 (14.3)
			Sun	2836 (16.0)	265 (12.7)	7 (0.4)	3108 (14.3)
			(Missing)	0 (0.0)	0 (0.0)	1771 (93.5)	1771 (8.1)
Sex	19862 (100.0)	0	Female	9990 (56.2)	128 (53.9)	1230 (64.9)	12348 (56.8)
			Male	7780 (43.8)	964 (46.1)	664 (35.1)	9408 (43.2)
Age	19862 (100.0)	0	Mean (SD)	41.2 (12.9)	37.2 (13.9)	43.7 (12.2)	41.1 (13.0)

Chapter 5

Balancing the Impulse to Leave Home against Public Health

5.1 Linking Text for Chapter 5

In the last two Chapters, we saw that state changes affect how impulsivity and self-control manifest in people's daily lives, and how these behaviours may be influenced by simple prompts aimed at influencing people's mental state. Specifically, I found that state impulsivity and hunger can both sensitise urges, and that hunger seems to decrease people's ability to control urges. I also showed how typically selfish urges in the pandemic can be reduced by thinking about possibilities of a positive future and one's role in accomplishing this future, or by considering the situations of others compassionately and linking these outcomes to one's own actions.

The two Chapters brought us closer to understanding how impulsive behaviours and self-control manifest in daily life because they sent short surveys to participants' mobile phones to gain insight into their urges and struggles in their lives. The findings emphasise that changing one's state merely by eating, thinking of the future, or thinking compassionately about those who are hard done by can make a considerable difference to the experience of urges. The difference in the experience of urges due to variations in impulsive states and the treatment administered in Chapter 4 provides further support to the model introduced in Chapter 1. That model suggests that state impulsivity affects the drive to engage in behaviours that are of value to the agent. In Chapter 3 we

observed that this is the case, whereas in Chapter 4 we intervened with a contemplative manipulation. In both cases, state impulsivity sensitised urges but did not impair self-control for urges of similar intensity.

The previous two Chapters, however, mostly pertain to the direct experience of motivational drives and resulting behaviours, without taking into consideration the beliefs and attitudes that underlie those behaviours: how badly did they want to avoid those behaviours? What are the predictors for this desire? And why?

The next two Chapters concern behaviours for which impulsivity and self-control play an important role, namely prosocial behaviours. In the current Chapter, I will explore the topic of prosocial behaviour further, but change the focus from people's urges and their ability to control them, to their attitudes and beliefs about prosocial behaviour. The subject of this Chapter is the perceived acceptability, in the broad sense, of leaving one's house while in lockdown during the COVID-19 pandemic. This is an interesting context for the current thesis because it allows for a nuanced analysis of how people balance the *collective prosocial*—staying home to limit viral transmission in the community—with selfish or *private prosocial* needs and wants. For example, balancing collective public health by staying home and avoiding viral transmission against going out to take care of one's physical and mental health.

Casting this in the terms of the conceptualisation and model developed in Chapter 1, this Chapter takes the reward, or the impulse, in the model to be the desire to leave the house, and sheds light on what shapes the potential intra-personal conflict that occurs when lower-order preferences (e.g., to leave the house to see one's family) do not match with higher-order preferences (e.g., I *want to want* to protect the vulnerable because I would like to be a good community member).

It is worth noting, however, that due to design and time constraints, no explicit psychological measures of impulsivity or self-control were administered in the following two Chapters. As such, the examination of impulsivity and self-control in the following two Chapters are less direct. Nonetheless, in both Chapters, much like in Chapter 4, I pit against each other the drive for private gains and the superordinate goals of protecting public health.

This Chapter, thus, will provide insight into the beliefs behind the motivational drives for impulsive behaviours and self-control in scenarios that are crucial for public health in disease outbreaks. Moreover, it will explore how different messaging strategies affect attitudes regarding these behaviours, providing insight into how public health officials might intervene depending on the stage of the outbreak.

Received: 27 May 2021 | Accepted: 21 December 2021

DOI: 10.1111/bjop.12554

the british psychological society
promoting excellence in psychology

ORIGINAL ARTICLE

Risk perception, illusory superiority and personal responsibility during COVID-19: An experimental study of attitudes to staying home

Simon T. van Baal[1,2] | Lukasz Walasek[2] | Daniela Karanfilovska[3] | Allen C. Cheng[3,4] | Jakob Hohwy[1]

[1]Cognition & Philosophy Lab, Department of Philosophy, Monash University, Melbourne, Victoria, Australia

[2]Department of Psychology, University of Warwick, Warwick, UK

[3]Infection Prevention and Healthcare Epidemiology Unit, Alfred Health, Melbourne, Victoria, Australia

[4]School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

Correspondence
Simon T. van Baal, Cognition & Philosophy Laboratory, Department of Philosophy, Monash University, Melbourne, VIC, Australia.
Email: Simon.vanBaal1@monash.edu

Funding information
Monash-Warwick Alliance Accelerator

Abstract

Little is known about how different government communication strategies may systematically affect people's attitudes to staying home or going out during the COVID-19 pandemic, nor how people perceive and process the risk of viral transmission in different scenarios. In this study, we report results from two experiments that examine the degree to which people's attitudes regarding the permissibility of leaving one's home are (1) sensitive to different levels of risk of viral transmission in specific scenarios, (2) sensitive to communication framings that are either imperative or that emphasize personal responsibility, or (3) creating 'loopholes' for themselves, enabling a more permissive approach to their own compliance. We find that the level of risk influences attitudes to going out, and that participants report less permissive attitudes to going out when prompted with messages framed in imperative terms, rather than messages emphasizing personal responsibility; for self-loopholes, we find no evidence that participants' attitudes towards going out in specific scenarios are more permissive for themselves than for others. However, participants report they are more rigorous in staying home than others, which may cause moral licensing. Additionally, we find that age is negatively associated with permissive attitudes, and that male participants are more permissive to going out. Thus, during phases where it is important to promote staying home for all scenarios,

including those perceived to be low-risk, imperative communication may be best suited to increase compliance.

KEYWORDS

COVID-19, illusory superiority, imperatives, personal responsibility, risk perception, social distancing

INTRODUCTION

COVID-19 is a medical condition, but achieving compliance with public health measures is a behavioural, socioeconomic and ethical matter, which severely tests the patience, resources and decision-making of any population. To enable efficient suppression (or mitigation or containment) of the pandemic, future variants and other disease outbreaks, it is therefore critical to explore the conditions under which the public complies optimally with requests to stay home and to socially distance themselves (Anderson et al., 2020; Bavel et al., 2020; Lunn et al., 2020; Zhang et al., 2020).

Governments focus on effective communication to the public to enhance compliance while matching the language to the severity of the situation. This is prudent because message framing impacts physical distancing intentions and attitudes (Everett et al., 2020). For example, in the UK, communication changed from a strong imperative 'Stay Home' to language that invites personal responsibility: 'Stay Alert'. This was mirrored in Australia, where the motto communicated to the public was initially 'if you can stay home, you must stay home', changing to 'stay safe', when the danger to the health care system decreased. It is unclear, however, how people's attitudes towards going out change when different communication styles are used to guide compliance. In particular, we are interested in whether imperative-based instrumental language (Renn, 2008) and personal responsibility-based communication that seeks to increase civic engagement (Head, 2011) affect risk perception and attitudes to behaviour for different levels of risk. Further, we are interested in whether such communication styles modulate attitudes to behaviour differentially when people are considering their own case or that of others.

COVID-19 risk perception has been studied in light of political orientation (Barrios & Hochberg, 2020), personal experience with the virus, individualistic and prosocial values, trust in science (Dryhurst et al., 2020) and perceived probability of getting infected (Wise et al., 2020). However, less is known about the attitudes towards going out for activities with different levels of transmission risk associated with them. There has been some interest in 'marginal' cases where decisions are particularly difficult (Lunn et al., 2020), but what constitutes a 'marginal' case differs for each person and depends on the context. The starting point for this study is therefore whether people's attitudes to going out in specific scenarios correspond to the actual transmission risk (i.e. people ought to be more stringent about going out in high-risk scenarios).

As mentioned, there are prominent examples where communications regarding desired behaviour during the COVID-19 pandemic largely fall into one of two classes: imperative messaging, and communication that invites personal responsibility and reasoning from the public. The communications that invoke personal responsibility are often part of a strategy to create or maintain civic engagement, whereas the imperative simplifies the task at hand, a strategy that appears to follow the (often criticized) deficit model. The deficit model homogenizes the audience and communicates information that serves to fill a deficit in knowledge (Meyer, 2016). Such an approach is often not beneficial for civic engagement, but when risks are high and without ambiguity, instrumental discourse can work well (Renn, 2008). In contrast, communication that encourages civic engagement (or community engagement) can enhance the effectiveness of containment measures in public health emergencies (Renn, 2008, 2020) and increase the likelihood of cooperation by the public (Head, 2011). Thus, which mode of communication is most successful at swaying attitudes on going out is yet to be determined and this knowledge could be of assistance for public health responses. This study thus analyses how the different modes of

communication (imperative and responsibility-based) influence behavioural attitudes across high, low and minimal risk levels.

People will display responsible decision-making only if they do not create 'self-loopholes' for their actions. That is, if they believe an action is impermissible in general (or for others), then they should not believe it is permissible for themselves. In contrast to many of our everyday moral decisions in normal times (e.g., giving to charity or not), a wrong decision during a pandemic can have disastrous consequences (e.g., joining a large gathering or travelling to an unaffected area while infected) (for modelling that highlights this moral aspect, see Donnarumma & Pezzulo, 2020). Loophole reasoning has been evident in some widely publicized cases of politicians and public officials publicly espousing general physical distancing edicts and yet privately finding exceptions for themselves. Here, we therefore investigate if attitudes to scenarios at different risk levels become laxer when they are presented in the first person compared to the third person.

An efficient government communication strategy for managing the COVID-19 crisis thus requires clarity on the role of several psychological factors, which may interact in critical ways, for responsible decision-making about whether to stay home or go out: the perception of different risk levels of everyday scenarios, whether instructions to stay home are imperative or invoke personal responsibility, and how various communication strategies affect beliefs about one's own actions and those of others.

Accordingly, the two experiments reported in the current study measure how permissive participants' attitudes to going out of the house are for a set of everyday scenarios. In the first experiment, scenarios presented to the participants vary in risk level, are presented in the first or third person, and with either imperative-based, personal responsibility-based, or no messaging. We made the following predictions: (1) attitudes to going out will be least permissive for high-risk scenarios, and increasingly permissive for low and minimal-risk scenarios. (2) There will be less permissive attitudes to going out when communication is framed as an imperative and less permissiveness when personal responsibility is emphasized, especially for higher risk levels, compared to the control condition. For loophole reasoning, a more exploratory approach seems appropriate: we are interested in whether permissiveness to going out is higher for cases pertaining to other people than those pertaining to oneself. The second experiment aims to replicate some of the findings in experiment 1 and further explores the possibility of loophole reasoning.

EXPERIMENT 1

Methods

Participants

The data collection was completed between 8 and 16 June 2020 via the Prolific online research participant database, sampling residents in the UK. At this time, the effects of the pandemic in the UK had marginally improved because the country started to get a grip on the first wave of COVID-19 through physical distancing measures, and lockdown restrictions were eased significantly. The importance of pandemic response measures had likely become evident to the population at that time.

Data from 607 participants, using volunteer sampling (through Prolific), were collected, and those who did not report that they resided in the UK during the survey were excluded, leaving 581 participants. The participants were paid £1.25 for their participation. All participants provided informed consent as approved by the Monash University Human Research Ethics Committee (Project ID: 24774).

The sample consisted of 39.2% males with a mean age of 40.1 (SD = 13.2), and 60.8% females with a mean age of 38.2 (SD = 12.3). Participants were required to be fluent in English (92.4% were native speakers and 7.1% of the participants indicated that English was their second language). Participants were excluded before completion of the experiment if they were diagnosed with dyslexia, dyspraxia,

attention deficit hyperactivity disorder or if they had trouble reading for any reason (including uncorrected abnormal vision).

Apparatus

Participants were each presented with 30 scenarios, and they evaluated these scenarios by indicating to what degree they were certain that it was alright to leave the house under the circumstances described in the scenario (we called this 'self-isolation' in the experiment, and it was made clear that this referred to staying home under stay-at-home orders). Participants indicated their certainty on a visual analogue scale (VAS) ranging left-to-right from 0 to 1, without them being able to see this number, with 0 signifying 'completely certain it is not alright to go out' and 1 signifying 'completely certain it is alright to go out'. The framing 'alright to go out' was chosen for the dependent variable as a succinct, jargon-free, colloquial and normatively neutral indicator of the participant's attitude to leaving the home in a given hypothetical scenario.

The VAS anchor was made to be invisible until the participants would press on the bar with their mouse cursor. The VAS was displayed on the same page as the scenario and question prompt, and the length of the VAS was adjusted in proportion to the dimensions of the viewing screen, as determined by Qualtrics programming.

Procedure

After signing consent, participants were required to pass an attention check, in which they received a question asking them to indicate which one of four options was not a city in the US, with the instructions reading 'Regardless of the right answer, please select Chicago'. If they did not complete this question correctly, the participants were not allowed to complete the survey and no data were recorded. Afterwards, the participants completed basic demographics questions and they were prompted with instructions:

'This study is about self-isolation decision making during the Covid-19 crisis. Imagine that the country you live in is in lockdown during the first wave of the disease, while the number of daily new cases is starting to come down. Everybody except essential workers is required to self-isolate in order to stop the spread of Covid-19 through the community. We will prompt you with 30 different scenarios concerning self-isolation. Please take your time as you go through these. For each scenario, please indicate to what extent you deem it alright for people to go out of the house. We will start with an example, so you can familiarize with the task'. See Appendix for additional instructions presented in conjunction with a practice trial.

The scenarios were divided into three levels, presenting a classification of the risk of transmission of the virus. This within-participants Risk Level factor thus has three levels: minimal-risk, low-risk and high-risk. The scenarios were categorized into risk levels by consensus of two experts in infectious diseases, with extensive experience in COVID-19 public health (AC, an infectious diseases physician and former Deputy Chief Health Officer to the Australian state of Victoria; and DL, an experienced clinical nurse consultant in infection prevention). We deemed the risk profile of the scenarios used to be fairly generalizable across populations, even though risk and public health messaging may have differed in some respects between the UK and Australia at various points of the pandemic. Participants were not shown the risk levels and were not explicitly asked about them. The order of presentation of the scenarios was randomized (for a complete list of the scenarios, see Table A1). Within each of the 30 trials, the participants were reminded of the overall situation with the words 'The country is in lockdown, people are self-isolating' at the top of the screen above the scenario prompt.

As a between-participant manipulation, the scenarios were either presented referring to the participant themselves (i.e. referring to 'you'), or using an indefinite pronoun (i.e. referring to 'someone'). For

example, 'You want to go for a vigorous walk at sunrise...' versus 'Someone wants to go for a vigorous walk at sunrise…'. This Perspective factor thus has two levels: Self and Other.

As a further between-participant manipulation, each scenario was followed by communication on how to handle decisions on staying home or going out. After reading the scenario, either they read the sentence 'people should stay home, if they can stay home' (Imperative condition), 'people should consider whether they will get close to other people before they go out' (Personal Responsibility condition) or no further communication (Control condition). This is then the Communication factor, with three levels. These conditions were presented between participants to help ensure participants stayed naive to the purpose of the experiment.

Participants were randomly assigned to one of the six treatment arms, parsed by the Communication factor, and the Perspective factor. This was done by using the randomized block assignment option with quotas in Qualtrics. In the final sample 192 participants were assigned to the Control condition (97 in the Self condition, 95 in the Other condition), 197 participants were assigned the Imperative condition (Self, n = 98; Other, n = 99) and 192 participants were assigned the Personal Responsibility condition (Self, n = 96; Other, n = 96). The experiment then has a mixed 3 x 2 x 3 design, with three independent variables (Risk Level, Perspective and Communication), and one dependent variable, permissiveness (i.e. indicating certainty about whether it is alright to go out or not on the VAS). There were two between participant factors, Communication and Perspective, there were thus six groups with 95–99 participants in each.

Analysis

Participants' permissiveness to going out is operationalized as the scores participants reported on the VAS. The independent variables tested for the main analysis are Risk Level, with three levels (in ascending order: minimal-risk, low-risk and high-risk), perspective, with two levels (Self and Other) and communication, with three levels (Control, Imperative and Personal Responsibility). See Table 1 for descriptive statistics partitioned by the three factors.

Given the non-normal, continuous, ordinal nature of the data, we conducted a continuous ordinal regression (Manuguerra & Heller, 2010; Manuguerra et al., 2020), with permissiveness as the dependent variable, including each of these factors as independent variables together with the participant's age and sex. In addition, the participant was included as a random intercept in the model and an interaction between Communication and Risk Level was included. Given the risk and communication factors consisted of three levels and main effects of factors cannot be judged from model coefficients if factors consist of more than two levels, likelihood ratio tests were used to assess the presence of main effects and interaction effects. The other effects were assessed through t-tests on the model coefficients.

TABLE 1 The age and sex distribution of participants in the 6 groups, partitioned by the Communication factor (3 levels), and the Perspective factor (2 levels).

	Communication					
	Imperative		**Personal Responsibility**		**Control**	
	Mean Age (*SD*)	***N***	**Mean Age (*SD*)**	***N***	**Mean Age (*SD*)**	***N***
Self						
Female	37.8 (12.1)	61 (62.24%)	36.59 (11.84)	61 (63.54%)	37.22 (11.46)	59 (60.82%)
Male	39.97 (13.59)	37 (37.76%)	38.63 (12.23)	35 (36.46%)	40.39 (12.22)	38 (39.18%)
Other						
Female	37.9 (12.32)	58 (58.59%)	40.45 (13.4)	60 (62.5%)	39.26 (12.32)	54 (56.84%)
Male	43.98 (13.21)	41 (41.41%)	37.64 (12.4)	36 (37.5%)	39.41 (14.41)	41 (43.16%)

Any effects were considered significant if they fell below the significance criterion (a = .05). Differences between factor levels were considered significant based on a false discovery rate (FDR) corrected alpha-level of .05. We also obtained a Bayes Factor for inclusion of the Perspective effect to assess the evidence for the null hypothesis using the bayesTestR package (Makowski et al., 2019), based on the approach by Wagenmakers (2007). In addition, an exploratory analysis was conducted by analysing the effects of age and sex on the VAS responses.

We were unable to get a reasonable a priori estimate of the effect sizes this study would be likely to have. A G*Power (Faul et al., 2009) estimate for 90% power at an effect size of d = .1 produced a sample size of 432 participants with our mixed design. To mitigate the ambiguity of the effect sizes a priori, the target sample size was set to 140% of this estimate.

Data are available here: https://osf.io/dbqa4/?view_only=65b6efa2b9944b9fbb2d0b46abe95878

Results

The sample was well balanced across conditions in terms of age and sex, although female participants were-overrepresented in all conditions (see Table 1).

There was a main effect of the Risk Level factor $\chi^2(13.53) = 4082.1$, $p < .0001$. Model coefficients show that the participants indicated lower permissiveness to going out in the scenarios identified as high-risk scenarios (M = .202, SD = .283) than in the minimal-risk scenarios (M = .509, SD = .362), OR = 8.637, 95% CI [8.045, 9.272], $t(572) = 59.50$, $p < .0001$, and than in the low-risk scenarios (M = .333, SD = 31.9), OR = 2.872, 95% CI [2.693, 3.062], $t(572) = 32.13$, $p < .0001$. Participants also reported significantly lower permissiveness in the low-risk scenarios compared to the minimal-risk scenarios, OR = 3.007, 95% CI [2.784, 3.248], $t(572) = 27.95$, $p < .0001$. The odds ratios refer to the chances of obtaining a lower value on the VAS scale, so the odds of getting a lower permissiveness value in a high-risk scenario compared to the low-risk scenario ~ 2.872:1 (see Figure 1a).

There was also a main effect of the Communication factor (Imperative, Personal Responsibility and Control) on permissiveness, $\chi^2(4.16) = 79.8$, $p < .0001$. The imperative phrasing reduced permissiveness (M = .277, SD = .323), OR = 1.486, 95% CI [1.137, 1.940], $t(572) = 2.90$, $p = .01$ compared to the control (M = .324, SD = .348), but there was no evidence that the personal responsibility phrasing did the same (M = .312, SD = .335), OR = 1.117, 95% CI [.833, 1.401], $t(572) = .82$, $p = .41$. After the FDR correction, participants' judgements in the imperative condition were not significantly less permissive than

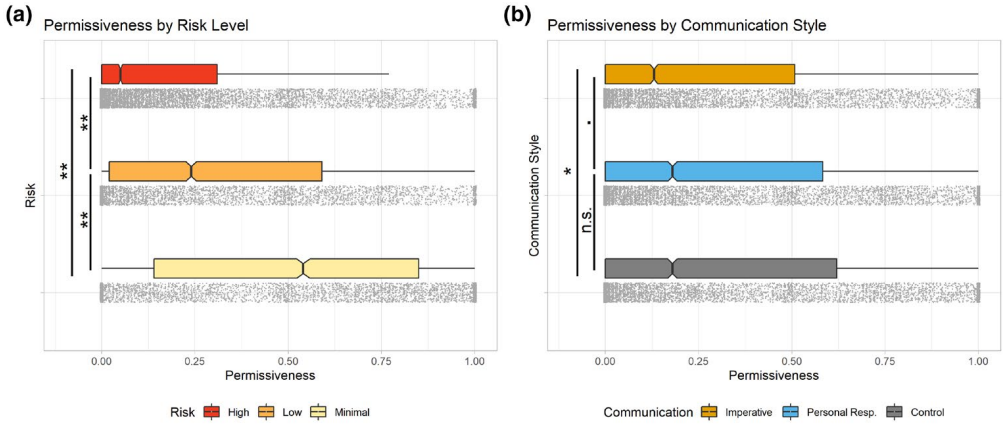

FIGURE 1 Main effects of Risk Level factor and Communication factor. The left panel (a) indicates participants' visual analogue scale (VAS) responses to the three Risk levels (respectively: high-risk, low-risk and minimal-risk), while the right panel (b) pertains to the three Communication conditions (respectively: Imperative, Personal Responsibility and Control). Higher VAS responses correspond to more permissive attitudes to leaving home. **: $p < .001$; *: $p < .05$;.: $p < .1$; n.s.: $p > .1$

the personal responsibility condition, OR = 1.327, 95% CI [1.016, 1.733], $t(572) = 2.07$, $p = .057$. This means that the imperative framing caused less permissive attitudes towards going out of the house than the control, while we found no evidence that personal responsibility communication had this effect (see Figure 1b).

There was a significant interaction effect of the Communication factor and the Risk Level factor on permissiveness, $\chi^2(4.32) = 78.7$, $p < .0001$. Where the difference between the Imperative condition ($M = .456$, $SD = .356$) and the Control condition ($M = .555$, $SD = .364$) was greatest in the minimal-risk scenarios, OR = 1.895, 95% CI [1.415, 2.536], $t(572) = 4.29$, $p < .001$. The difference ($M = .294$, $SD = .310$ v $M = .371$, $SD = .327$) was attenuated (but still present) for the low-risk scenarios, OR = 1.677, 95% CI [1.257, 2.237], $t(572) = 3.52$, $p < .001$, and attenuated further for the high-risk scenarios, where the effect of the imperative condition was not significant ($M = .191$, $SD = .276$ v $M = .203$, $SD = .288$), OR = 1.032, 95% CI [.079, 1.354], $t(572) = .23$, $p = .92$. The personal responsibility condition displayed the same trend, but the effect was not significant in any of the conditions (minimal-risk: $M = .518$, $SD = .359$; low-risk: $M = .335$, $SD = .315$; high-risk: $M = .213$, $SD = .286$) ORs: 1.270, 1.250, .884; $ps > .1$. See Figure 2 for a depiction of the differences in attitudes for each risk level per communication condition.

Participants' reported permissiveness to going out when scenarios were presented referring to someone ($M = .299$, $SD = .334$) as compared to when presented as pertaining to oneself ($M = .309$, $SD = .338$), was not significantly higher, OR = 1.021, 95% CI [.910, 1.145], $t(572) = .18$, $p = .85$. To assess support for the null hypothesis, that there was no effect of the perspective factor, we assessed the evidence for excluding the effect from the model, $BF_{01} = 1.62$. This evidence in favour of excluding the effect from the model shows anecdotal support for the absence of a perspective effect.

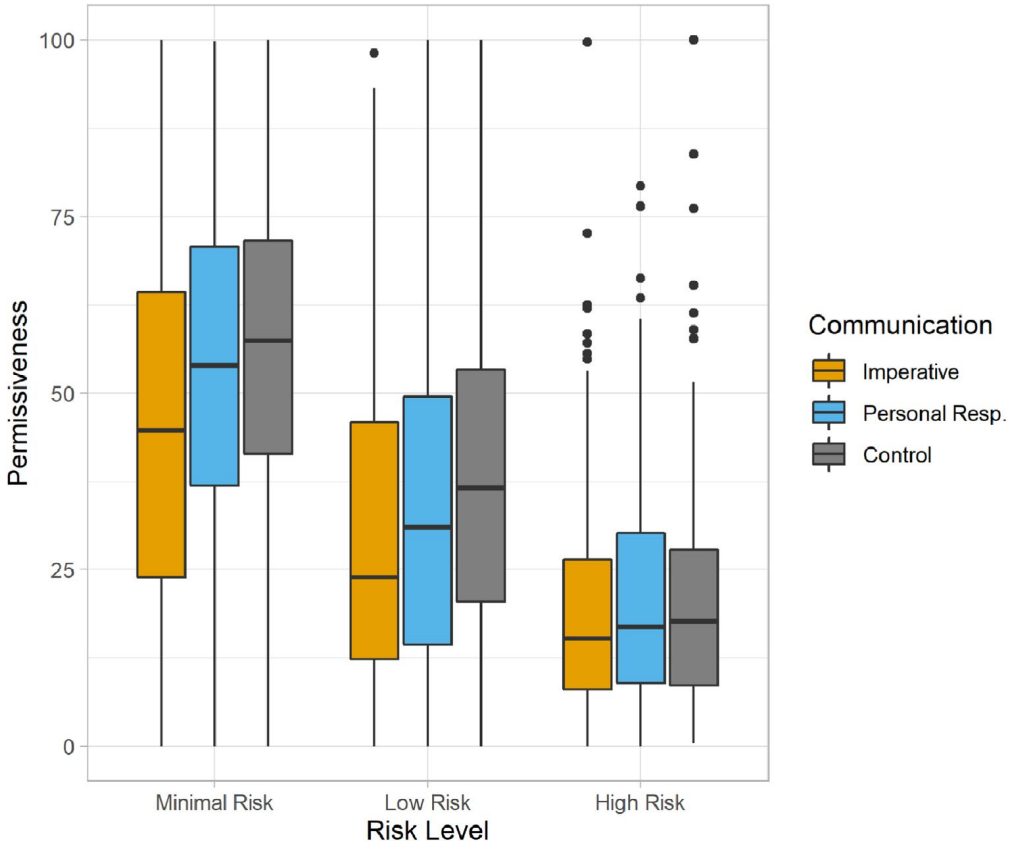

FIGURE 2 Interaction of the risk level of a scenario and the communication style. The orange boxplot indicates the responses in the Imperative condition, the blue indicates the responses in the Personal Responsibility condition, and the grey indicates the responses in the Control condition. The data are represented on the measurement scale.

Subsequent exploratory analysis showed that age and sex affected permissiveness. Younger people were more likely to report permissive attitudes to going out, OR = 1.035, 95% CI [1.026, 1.044], $t(572) = 7.79$, $p < .0001$, and male participants were more likely to report permissive attitudes than females ($M = .335$, $SD = .346$; $M = .284$, $SD = .327$), OR = 1.481, 95% CI [1.325, 1.657], $t(572) = 3.448$, $p < .0001$. See Figure 3 for a depiction of the relationship between permissiveness, age and sex.

Discussion of Experiment 1

The findings from the first experiment are that (1) participants' attitudes become less permissive to going out with increasing risk level of transmission; (2) imperative communication decreases participants' permissiveness, especially in low and minimal-risk scenarios. There was no self-loophole effect. An exploratory analysis of the demographic variables revealed that younger participants were more permissive to going out than older participants, and males were more permissive than females.

This experiment has some limitations, which provided reasons for conducting a second experiment. First, this experiment used a VAS with no banding or marking beyond the labels for the endpoints. This choice of scale could have had an impact on results; floor and ceiling effects were evident in this experiment, which might conceal further effects of interest. It is possible that a banded VAS design, or giving dynamic feedback upon placing the marker, would modulate the results (Matejka et al., 2016), and there are multiple factors to consider when deciding to employ a midpoint or not (Chyung et al., 2017). Another experiment was needed to examine if the inclusion of a midpoint would lessen the floor/ceiling effects evident in this study (as well as in other studies, such as Barari et al., 2020). Thus, the second experiment incorporated a between-participants factor assessing the effects of a midpoint marker on the VAS versus no midpoint maker on the VAS (this factor will be referred to as VAS Marker). To not dramatically increase the required n to obtain reliable results, it was necessary to drop the Communication factor.

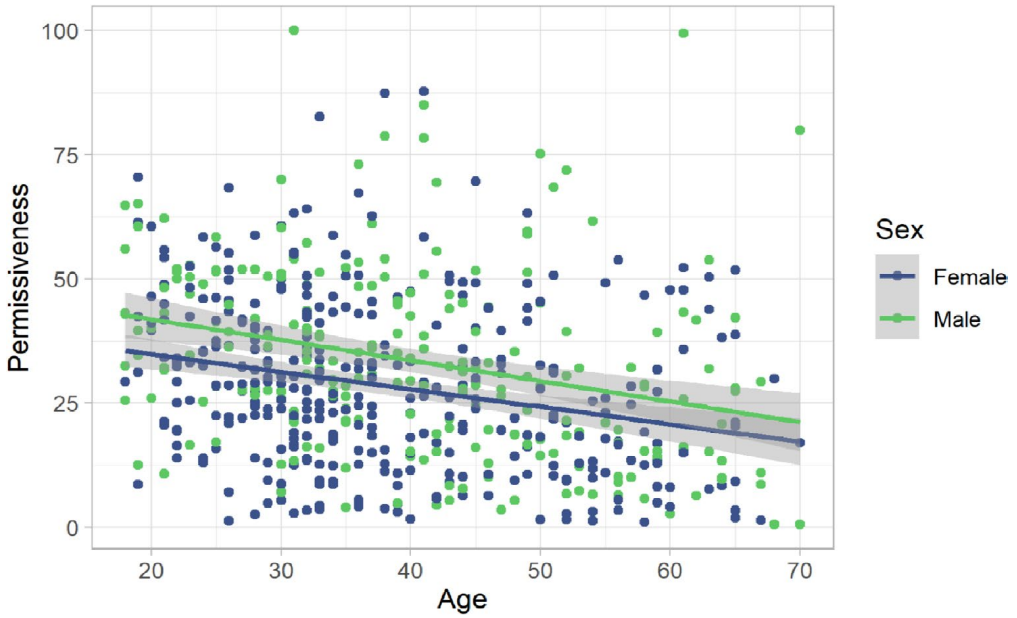

FIGURE 3 Exploratory visualization of the relationship between participants' permissiveness attitudes and demographic variables in experiment 1. The blue points are average visual analogue scale (VAS) responses for females, and the green points are average VAS responses for males. The blue and green regression lines show the effect of age on permissiveness, partitioned by sex.

Second, the initial experiment used an indefinite pronoun (i.e. 'someone') when describing scenarios in the Other condition. This could have weakened our perspective manipulation because 'someone' can be used to describe oneself. Thus, the second experiment was designed to address this issue by referring to 'a person in your neighbourhood'.

In addition, we investigated another possible way that people may arrive at loophole reasoning. People have a general tendency to view their own actions more favourably, which is called self-serving bias (Mezulis et al., 2004), and to view themselves as better than average, illusory superiority (Zell et al., 2020). This could also be the case for pandemic response behaviour, such that people will falsely believe their physical distancing behaviour is more rigorous than that of others. This is potentially harmful because if people believe that they have acted morally (because they think they are comparatively rigorously staying home), they will be more likely to behave immorally later (Blanken et al., 2015), which could be a reason for suboptimal compliance. This counterintuitive behaviour pattern is called moral licensing. Exploring this other form of self-loopholes is important because the hypothetical scenarios in this study are unlikely to capture such effects, given that these scenarios are supposed to be viewed in isolation, and not in context of one's past actions. The second experiment investigates that question by asking participants to rank their compliance as compared to others in their community, and uses the results of that to infer whether they exhibit illusory superiority.

The effects of sex and age were discovered due to exploratory analysis. Therefore, we deemed it necessary to replicate these results and strengthen our confidence in the evidence that these effects exist. Given these three limitations, we decided to conduct another experiment.

EXPERIMENT 2

The second experiment was conducted to replicate the effects of the first experiment holding fixed the communication messaging, to validate the methodology of our VAS design, to verify the sex and age effects, and to further investigate self-loopholes and beliefs about one's own physical distancing behaviour compared to that of others.

Thus, we made the following predictions for the second experiment: (1) participants are sensitive to levels of risk of viral transmission, and as such display decreasing permissiveness to going out over increasing levels of risk. (2)(a) Younger people and (b) males are more permissive to going out. (3) Introducing a midpoint marker on the VAS will alleviate the ceiling and floor effects that were evident in the first experiment. (4) Participants make self-loopholes when the phrasing of a scenario is explicitly allocentric. (5) Participants display illusory superiority when they compare their compliance to that of others in their community.

It is noteworthy that the second experiment was conducted 2 months after the first, in which the situation was marginally different. For instance, restrictions were in place, but lockdowns had long been lifted at that point because the UK was seemingly able to keep infection rates low. Schools were planned to reopen in the next month, and the R number was said to be above 1 only in three relatively densely populated areas of England (London, the North-West, the South-West, see e.g. https://www.itv.com/news/2020-08-07/coronavirus-r-number-could-be-above-one-in-london-south-west-and-north-west).

Methods

Participants

The data collection were completed on 7 August 2020 via the Prolific online research participant database. Data from 412 participants were collected, and three participants whose responses were all at either bound of the scale were excluded from analysis, as well as one participant who answered 'Other' on the sex question was excluded from analysis due to the low number of observations in this category,

leaving 408 participants. Participants were paid £1.25 for their participation. All participants provided informed consent as approved by the Monash University Human Research Ethics Committee (Project ID: 24774).

The sample consisted of 33.1% males with a mean age of 37.2 (*SD* = 11.7) and 66.9% females with a mean age of 33.1 (*SD* = 10.6). Participants were required to be fluent in English (89.0% were native speakers and 10.8% of the participants indicated that English was their second language). Participants were excluded before completion of the experiment if they had participated in the initial experiment, were diagnosed with dyslexia, dyspraxia, attention deficit hyperactivity disorder, or if they had trouble reading for any reason (including uncorrected abnormal vision).

Apparatus

The VAS was largely the same as in Experiment 1, but as a between-participant manipulation, the VAS was either identical to the one in the initial study (No Marker condition), or it contained a visible midpoint on the scale at 'uncertain whether it is alright or not' (Midpoint condition). In both conditions, we also included a label underneath the midpoint: 'Uncertain if it is alright or not' (see Figure A2).

After the 30 scenarios were presented, all participants were prompted with a question in which they were instructed to indicate where they ranked in the population in terms of compliance. The scale ranged from 0 signifying 'I self-isolated the least', 50 signifying '50% isolated more, and 50% isolated less than me', and 100 signifying 'I self-isolated the most'. The VAS for this question showed a number as the participants hovered over the slider with their cursor, so they could provide a more precise ranking. The question was phrased as:

> We want you to think about when the country was in lockdown. Consider how much you were self-isolating, and how much others were self-isolating. Where would you rank yourself in terms of self-isolation compared to others in your community?

Procedure

While in Experiment 1 participants were presented with self or other-framings of scenarios, this experiment also referred to the participant themselves (i.e. identical to the self-framing), but referred more explicitly to another person (i.e. referring to 'a person in your neighbourhood'). Further, given the prevalence of imperative framing in government communication during the response to COVID-19 in various nations, all scenarios were framed with imperative messaging, phrased identically to the manipulation that was administered in Experiment 1.

Participants were randomly assigned to the four treatment arms for the two between-participant factors, perspective and VAS marker. There were 202 participants in the Midpoint condition (Self, n = 100; Other, n = 102) and 206 in the No Marker condition (Self, n = 103; Other, n = 103). This experiment had a mixed 3 × 2 × 2 design, with three independent variables (Risk Level, Perspective, VAS Marker), and two dependent variables, permissiveness (i.e. indicating certainty about whether it is alright to go out or not on the VAS), and self-adherence ranking. There were two between participant factors, VAS Marker and Perspective.

Analysis

The main analysis for this experiment was analogous to the analysis in experiment 1; we conducted a continuous ordinal regression (Manuguerra & Heller, 2010; Manuguerra et al., 2020), with permissiveness as the dependent variable, including each of the three factors as independent variables, together

with demographic variables age and sex, and the participant as a random intercept. Considering the risk factor consisted of three levels and main effects of factors with more than two levels cannot be judged from model coefficients, likelihood ratio tests were used to assess the presence of main effects and interaction effects involving the risk factor. The other effects were assessed through t-tests on the model coefficients. Any effects were considered significant if they fell below the significance criterion ($\alpha = .05$). Differences between factor levels were considered significant based on a FDR corrected alpha-level.

We also obtained a Bayes Factor for inclusion of the VAS Marker and the perspective effect to assess the evidence for the null hypothesis using the bayesTestR package (Makowski et al., 2019), based on the approach by Wagenmakers (2007).

To assess the support for the null hypothesis that VAS Marker does not affect permissiveness, we obtained the Bayes Factor for inclusion of the Perspective and VAS Marker factors to assess the evidence for the null hypotheses using approach mentioned before.

In addition, we tested whether floor and ceiling effects were alleviated by transforming the VAS responses: the dependent variable was a dichotomous variable representing whether the response fell between [0,.1] or [.9, 1] (i.e. at the bottom or top 10% of the scale). Then, a likelihood ratio test of binomial generalized linear mixed models was used to identify whether a model including the VAS Marker factor fit the data significantly better than a model without that factor.

We then conducted a Wilcoxon Rank Sum test to assess whether the median of participants' self-reported self-adherence rankings was greater than 50. We also assessed the evidence on whether sampling at different timepoints affected participants' permissiveness by collating the data from both experiments, and running a continuous ordinal regression including the Communication, Perspective, Risk Level, VAS Marker, age and sex variables, as well as a variable for the timepoint at which data were collected. The coefficient for the timepoint was used to infer whether participants attitudes were different in Experiment 1 than in Experiment 2.

Results

The age and sex distribution per group was less well-balanced in this experiment compared to the first (see Table 2), but given that participants' characteristics were included in the model, this modest imbalance should not negatively impact the results.

Permissiveness in the No Marker condition ($M = .302$, $SD = .310$) was not significantly different from the Midpoint condition ($M = .315$, $SD = .317$), OR = 1.162, 95% CI [.891, 1.516], $t(408) = 1.11$, $p = .27$. There was also no significant difference in permissiveness between the Self condition ($M = .306$, $SD = .319$) and Other condition ($M = .311$, $SD = .307$), OR = 1.008, 95% CI [.763, 1.332], $t(408) = .06$, $p = .95$. We obtained Bayes Factors to assess support for the null hypotheses (i.e. no main effect of perspective or the VAS marker), resulting in $BF_{01} = 1.57$ for perspective and $BF_{01} = 1.36$ for the VAS marker. This indicates anecdotal evidence for the absence of an effect for both factors. For density plots of the responses partitioned by risk level (see Figure A3). The binomial GLMM showed that the VAS Marker was not a significant predictor for the likelihood of responding at the extremes of the scale [0 −.1;.9 – 10], $\chi^2(8) = .00$, $p = .98$.

Like in the previous experiment, there was a main effect of risk on permissiveness, $\chi^2(6.08) = 2214.9$, $p < .0001$. Participants reported less permissiveness of going out in high-risk scenarios ($M = .223$, $SD = .334$) than in the low-risk scenarios ($M = .333$, $SD = .301$), OR = 2.402, 95% CI [2.226, 2.592], $t(408) = 22.60$, $p < .0001$, and than in the minimal-risk scenarios ($M = .480$, $SD = .334$), OR = 6.628, 95% CI [6.100, 7.204], $t(408)$, $p < .0001$. They also reported less permissiveness in the low-risk scenarios compared to the minimal-risk scenarios, OR = 2.759, 95% CI [2.518, 3.024], $t(408) = 21.71$, $p < .0001$. This replicates the effect of experiment 1, which showed that participants reported less permissive attitudes (by reporting a lower score on the VAS) to going out for increasing risk levels (see Figure 4).

Older people reported less permissive attitudes to going out than younger people, OR = 1.042, 95% CI [1.029, 1.055], $t(408) = 6.63$ $p < .0001$ and female participants ($M = .300$, $SD = .310$) reported less

TABLE 2 Descriptive statistics of age and sex distribution in the 4 conditions that consist of the VAS Marker factor (2 levels), and the Perspective factor (2 levels). The N column indicates the number of participants in each cell, as a percentage of the number of participants in that group

	VAS marker			
	Midpoint marker		No midpoint marker	
	Mean age (*SD*)	*N*	Mean age (*SD*)	*N*
Self				
Female	31.97 (9.84)	61 (59.8%)	36.74 (10.72)	68 (66.02%)
Male	39.03 (10.72)	39 (38.24%)	36.94 (11.77)	35 (33.98%)
Other				
Female	32.83 (11.12)	71 (71%)	30.93 (9.51)	73 (70.87%)
Male	35.03 (11.46)	31 (31%)	37.5 (12.47)	30 (29.13%)

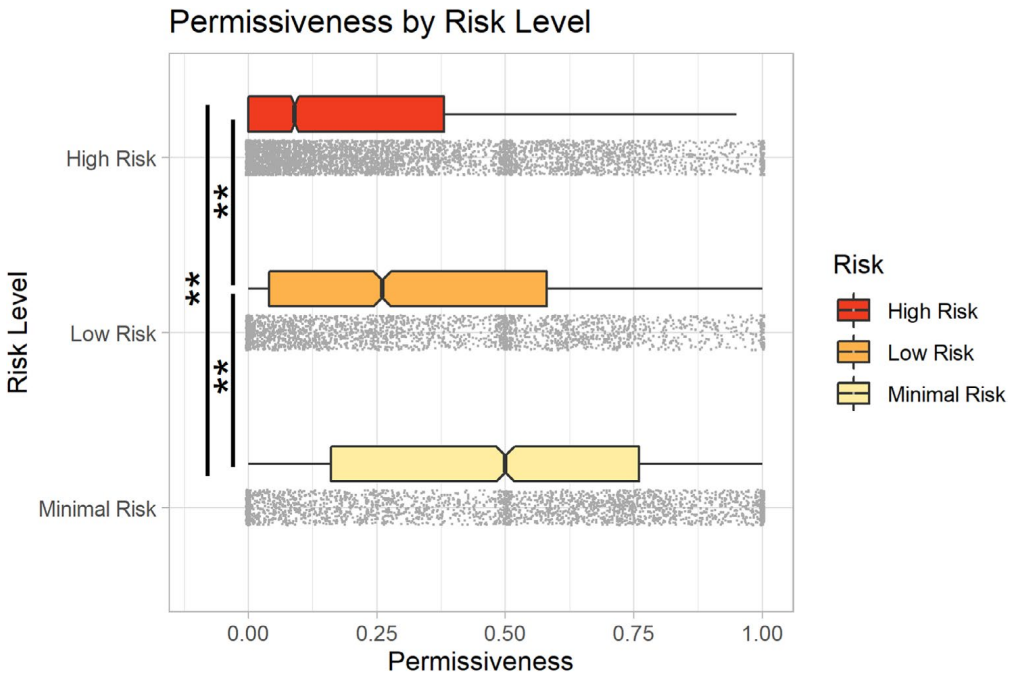

FIGURE 4 Main effects of risk; the dots represent participants' responses indicating permissiveness to going out (as indicated on a visual analogue scale, VAS) to the three risk levels (respectively: high-risk, low-risk and minimal-risk). Higher VAS responses correspond to more permissive attitudes to going out. **: $p < .001$; *: $p < .05$; $p < .1$

permissive attitudes than males ($M = .325$, $SD = .320$), OR = 1.249, 95% CI [1.083, 1.441], $t(408) = 3.05$, $p < .01$. This replicates the effects from experiment 1 (see Figure 5).

The median self-adherence ranking was significantly higher than 50, 95% CI [.795, .825], $\tilde{\chi} = 16.13$, $p < .0001$.

To see whether collecting data at the different timepoints had an effect on our dependent variable, we compared the VAS responses from June ($M = .304$, $SD = .314$) to August ($M = .309$, $SD = .336$), OR = 1.297, 95% CI [1.258, 1.339], $t(28) = 8.171$, $p < .0001$. This means that, all else being equal, participants in the first experiment were more likely to report less permissive attitudes. We found no other meaningful differences in participants' judgements between experiments.

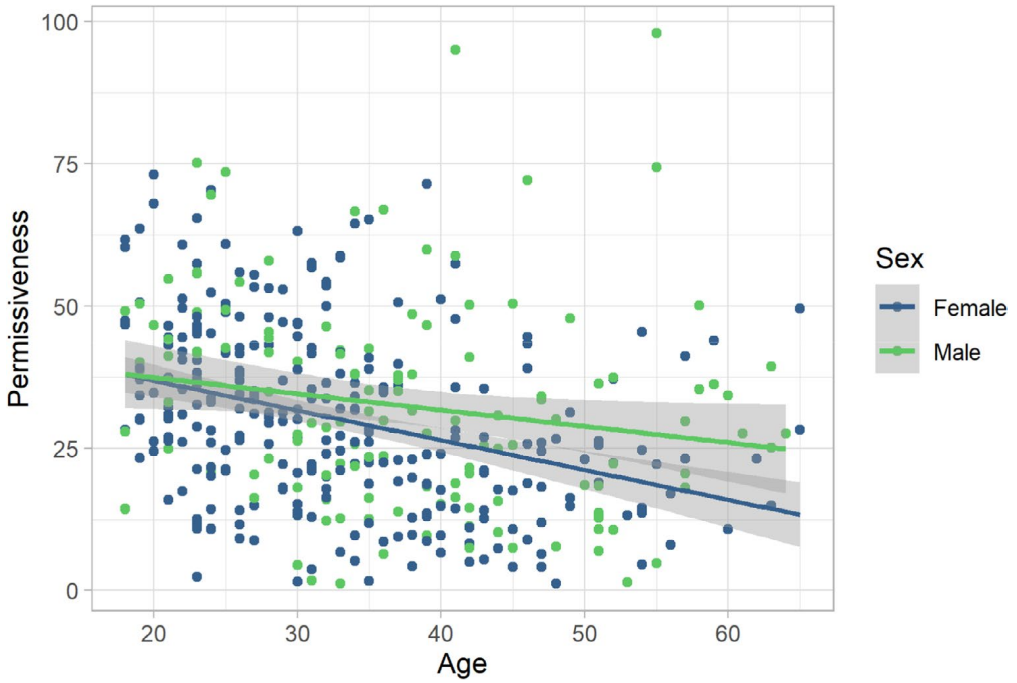

FIGURE 5 Visualization of the relationship between participants' permissiveness attitudes and demographic variables in the second experiment. The blue points are average visual analogue scale (VAS) responses for females, and the green points are average VAS responses for males. The blue and green curves are regression lines, showing the effect of age on permissiveness, partitioned by sex

Discussion of Experiment 2

The findings from the second experiment are that (1) like in the first experiment, participants' attitudes report lower permissiveness to going out in scenarios with increasing risk level. (2a) Sex is a significant factor in attitudes to staying home, with males reporting more permissive attitudes to going out than females; (2b) age is a significant factor in attitudes to staying home, with younger participants reporting more permissive attitudes to going out than older participants. (3a) There was no effect of VAS Marker, (3b) nor was there support for lesser floor and ceiling effects. However, there was a larger concentration of responses around the midpoint in the Midpoint condition. This shows that the midpoint marker encouraged fence sitting, rather than meaningfully relieving floor or ceiling effects. (4) There was no evidence to support an effect of Perspective. (5) Participants report illusory superiority, as they significantly overstate their stay-at-home order compliance in comparison to others in their community. We also found that participants in the second experiment were slightly more permissive than those in the first experiment, which may be due to the data collection happening in different stages during the pandemic. We did not find any other differences that could affect the interpretation of results.

GENERAL DISCUSSION

This study manipulated framing for a set of scenarios at three different risk levels, to investigate the association with people's attitudes to staying home or going out for specific activities. We studied the effects of imperative and personal responsibility-based framings of government advice ('people should stay home, if they can stay home' vs. 'people should consider whether they will get close to other people before they go out', and a control condition without additional framing), as well as self and other ('You'

vs. 'Someone' or 'A person in your neighbourhood') framings for the selected scenarios. Imperative framings were found to be effective at reducing permissiveness to going out, and participants show sensitivity to risk of disease transmission, but suffer from illusory superiority. We found no evidence that they judge scenarios from someone else's viewpoint more strictly than when they judge what to do themselves. Older people were less permissive of going out, as were female participants compared to males.

The imperative framing was more effective than the framing invoking personal responsibility at encouraging stringent attitudes to staying home in general, and particularly in low and minimal-risk scenarios. This interaction may be due in part to floor effects for the high-risk scenarios, where participants were already very impermissive of going out. Imperatives are simple and apply equally to all scenarios where the imperative can plausibly be followed, making them useful for cutting through uncertainty in decisions about going out, which can be effective in situations with high risk (Renn, 2008). However, even when faced with an imperative, people seem to reason about the risks of each condition, as revealed in the different levels of permissiveness per risk level. Thus, these results suggest that it may be prudent to use imperative framing when high compliance is necessary, even in minimal-risk scenarios, when presented with a virus or variant of concern with unknown transmission characteristics or an 'escape variant' that evades vaccines. A shift to communication that emphasizes personal responsibility about weighing the risk of contact with others against the need for going out may be warranted if the objective is to be more permissive overall, but in particular for outings perceived as minimal-risk.

People adjust their responses according to the risk associated with specific activities, indicating more certainty that it is not alright to go out for risky scenarios, middle levels of certainty for low-risk scenarios, and higher levels of certainty that it is alright to go out for minimal-risk scenarios (where attitudes are approximately uniformly distributed). The issue of risk perception of different scenarios is critical to understanding individual decision-making and actual risk-taking behaviour (Bran & Vaidis, 2020). Our findings are encouraging because they show that, in broad strokes, the public is aware of the riskiness of certain scenarios and takes this into account in their decision-making. Lunn et al. (2020) identify 'marginal' scenarios, defined as scenarios where people were not sure whether the activities presented were advisable, but when inspecting the responses in the current study Figures A1 and A3, it appears there is considerable disagreement on the acceptability of going out across all risk levels, but most notably in minimal-risk scenarios. This wide range of uncertainty about specific scenarios could be partially due to individual differences in abilities to accommodate loss (Osmundsen & Petersen, 2020). The disagreement in whether going out is acceptable for specific scenarios is an important issue to address in communications to the public during infectious disease outbreaks.

The self versus other framings of the scenarios did not lead to differences in responses, nor display any significant interaction with the other factors. However, the second experiment revealed that people display illusory superiority about their compliance. Understanding loophole reasoning is critical for COVID-19 because even a low number of people finding loopholes for themselves can have devastating effects (Donnarumma & Pezzulo, 2020). Illusory superiority may lead to moral licensing, and thus a self-loophole. That is, people may not *a priori* believe they are more justified in going out, but if they believe they have been morally 'good' by self-isolating rigorously, then they may be more likely to behave immorally and transgress (through moral licensing).

Several studies may pertain to our considerations about loophole reasonings. Jordan et al. (2020) investigated self-interested versus prosocial framings, a distinction that may create openings for loophole reasoning. Bilancini et al. (2020) investigated different norm-based potential communications, which may be relevant to loophole reasoning because norms can be personal or injunctive (i.e. what a person thinks other people believe is correct); they find that norm-based interventions have little impact on actual engagement. Various personality traits may also be relevant; Wolff et al. (2020) discuss impulsivity, which may relate to succumbing to temptations for loophole reasoning, and Swami and Barron (2020) discuss analytical thinking, which may relate to self-loopholes via its role in reasoning about scenarios.

Concerning risk perception, we also found that age and sex are relevant for attitudes to these scenarios. It is perhaps unsurprising that males are more permissive, less risk-averse, than females. Similarly, it

may be unsurprising that younger people are more permissive to going out than older people, given the age differences in how different groups are affected by COVID-19.

This study had some limitations that should be noted. First, the participants tested all reside in the UK. This may have had a distinct influence on the results due to the different trajectories of the COVID-19 pandemic in different countries, and might limit the extent to which these results apply to different populations. Second, the study does not provide information about actual behaviour, but rather an indication of participants' attitude or intention for (hypothetical) behaviour (cf. Sheeran & Webb, 2016). Third, even though there was reasonable spread in demographic variables, the samples were not representative of the population, which may affect the generalizability of findings, and in particular affect the effect sizes of age and sex. There is also a risk that self-reported attitudes to behaviours suffer from a social desirability bias. However, there is mixed evidence on social desirability bias around COVID-19 public health guideline compliance (for absence of an effect, see Larsen et al., 2020; for evidence of an effect, see Timmons et al., 2021). GPS data suggest that self-reported physical distancing during COVID-19 does predict actual physical distancing (Gollwitzer et al., 2020).

We chose to elicit judgements from our sample by asking them to what extent it was 'alright' to go out. This wording was chosen because it is colloquial and because it lacks significant normative meaning or jargon. And while this wording was intended to elicit comprehensive judgements that subsume moral, legal and public health considerations, participants may have interpreted this wording in various ways. For example, it is possible that our communication manipulation affected permissiveness by changing the way people interpreted the appropriate inputs to the judgement; perhaps some participants interpreted the imperative framing as emphasising that breaching the stay-at-home orders was a fineable offence, effectively increasing the salience of legal considerations. We made efforts to clarify in the instructions that the experiments concerned individual decision making (i.e. not solely legal or moral judgements), but we cannot rule out this mechanism for the effect of the imperative phrasing. Either way, we suggest that this mechanism would be interesting in itself, for example, if we could make people think about stay-at-home orders as a moral issue, then that could increase compliance (Frias-Navarro et al., 2021), and doing so indirectly—rather than by emphasizing morality directly—might be preferable in some situations. Shedding light on how imperatives from public health officials influence behaviour could be an interesting direction for future research.

The lack of accurate a priori effect size estimates for experiment 1 and the elaborate experimental design led to low power for detecting interactions, and it could explain why we did not find an effect for the personal responsibility framing (but this may also be due to ineffective wording). Future studies will be able to use effect sizes from these experiments as a baseline to power experiments more efficiently and use more focused designs to eliminate the need for larger sample sizes. This would enable a more accurate assessment of the various effects found in this study with economical use of resources.

Our findings may be relevant for public policy makers and citizens alike. The communication-style used when requesting compliance from the public may meaningfully shape people's ttitudes towards the policy. In particular, we might expect that using imperatives instils a sense of urgency that is appropriate for a future dangerous infectious disease outbreak. Further, it seems that, on aggregate, people display a relatively high degree of agreement in risk perception when a scenario is very risky, but in minimal risk scenarios attitudes differ widely—this may be useful information for policy makers and public health officials to incorporate in their models and predictions. Lastly, given we all tend to believe we self-isolate more than most others, it may be conducive to compliance, as well as potentially social cohesion, if communications seek to redress such conceptions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on OSF at https://osf.io/dbqa4/?view_only=7a024603633d4e5e8a4f7ef15179f513.

ACKNOWLEDGEMENTS

This research is supported by a Monash-Warwick Alliance Accelerator grant to SvB, JH and LW. Thank you to Lewis Barclay (Mckinnon Secondary College) who suggested inclusion of the Control condition and assisted with the scenarios. We thank Chiara Varazzano, Tim Bayne, Jonathan Robinson, Bryan Paton, John Michael and Jinhee Kim for discussion and advice.

CONFLICT OF INTEREST

All authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Simon Thomas van Baal: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Lukasz Walasek:** Conceptualization (equal); Funding acquisition (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Daniela Karanfilovska:** Methodology (equal); Validation (equal). **Allen C Cheng:** Methodology (equal); Validation (equal); Writing – review & editing (equal). **Jakob Hohwy:** Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

ETHICAL APPROVAL

This study was approved by the Monash University Human Research Ethics Committee (Project ID: 24774).

OPEN RESEARCH BADGES

This article has been awarded Open Materials, Open Data Badges. All materials and data are publicly accessible via the Open Science Framework at https://osf.io/dbqa4/?view_only=7a024603633d4e5e8a4f7ef15179f513; https://osf.io/7j4ny/?view_only=920ffe01f41a484591a850fa2301584a.

ORCID

Simon T. van Baal https://orcid.org/0000-0001-5351-4361
Lukasz Walasek https://orcid.org/0000-0002-7360-0037
Daniela Karanfilovska https://orcid.org/0000-0002-1579-1082
Allen C. Cheng https://orcid.org/0000-0003-3152-116X
Jakob Hohwy https://orcid.org/0000-0003-3906-3060

REFERENCES

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, *395*(10228), 931–934. https://doi.org/10.1016/S0140-6736(20)30567-5

Barari, S., Caria, S., Davola, A., Falco, P., Fetzer, T., Fiorin, S., Hensel, L., Ivchenko, A., Jachimowicz, J., King, G., Kraft-Todd, G., Ledda, A., MacLennan, M., Mutoi, L., Pagani, C., Reutskaja, E., & Slepoi, F. R. (2020). Evaluating COVID-19 public health messaging in italy: self-reported compliance and growing mental health concerns. *medRxiv*, 2020.2003.2027.20042820. https://doi.org/10.1101/2020.03.27.20042820

Barrios, J. M., & Hochberg, Y. (2020). Risk perception through the lens of politics in the time of the covid-19 pandemic (0898-2937).

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., … Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), 460–471. https://doi.org/10.1038/s41562-020-0884-z

Bilancini, E., Boncinelli, L., Capraro, V., Celadin, T., & Di Paolo, R. (2020). The effect of norm-based messages on reading and understanding COVID-19 pandemic response governmental rules.

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134 https://doi.org/10.1177/014616721557213410.1177/0146167215572134

Bran, A., & Vaidis, D. C. (2020). Assessing risk-taking: What to measure and how to measure it. *Journal of Risk Research*, *23*(4), 490–503. https://doi.org/10.1080/13669877.2019.1591489 https://doi.org/10.1080/13669877.2019.159148910.1080/13669877.2019.1591489

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the likert scale. *Performance Improvement*, *56*(10), 15–23. https://doi.org/10.1002/pfi.21727

Donnarumma, F., & Pezzulo, G. (2020). Moral decisions in the age of COVID-19: your choices really matter. *arXiv preprint arXiv:2004.07081*.

Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., Spiegelhalter, D., & van der Linden, S. (2020). Risk perceptions of COVID-19 around the world. *Journal of Risk Research*, *23*(7-8), 994–1006. https://doi.org/10.1080/13669877.2020.1758193

Everett, J. A. C., Colombatto, C., Chituc, V., Brady, W. J., & Crockett, M. (2020). The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic. https://doi.org/10.31234/osf.io/9yqs8

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Frias-Navarro, D., Pascual-Soler, M., Berrios-Riquelme, J., Gomez-Frias, R., & Caamaño-Rocha, L. (2021). COVID–19. Effect of moral messages to persuade the population to stay at home in Spain, Chile, and Colombia. *The Spanish Journal of Psychology*, *24*, 448–456.

Gollwitzer, A., Martel, C., Marshall, J., Höhs, J. M., & Bargh, J. A. (2020). Connecting Self-Reported Social Distancing to Real-World Behavior at the Individual and U.S. State Level. https://doi.org/10.31234/osf.io/kvnwp

Head, B. W. (2011). Australian experience: Civic engagement as symbol and substance. *Public Administration and Development*, *31*(2), 102–112.

Jordan, J., Yoeli, E., & Rand, D. G. (2020). Don't get it or don't spread it? Comparing self-interested versus prosocially framed COVID-19 prevention messaging.

Larsen, M., Nyrup, J., & Petersen, M. B. (2020). Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias? *Journal of Behavioral Public Administration*, *3*(2), 1–9. https://doi.org/10.30636/jbpa.32.164

Lunn, P. D., Belton, C. A., Lavin, C., McGowan, F. P., Timmons, S., & Robertson, D. A. (2020). Using behavioral science to help fight the Coronavirus. *Journal of Behavioral Public Administration*, *3*(1), 1–15. https://doi.org/10.30636/jbpa.31.147 https://doi.org/10.30636/jbpa.31.14710.30636/jbpa.31.147

Lunn, P. D., Timmons, S., Belton, C. A., Barjaková, M., Julienne, H., & Lavin, C. (2020). Motivating social distancing during the Covid-19 pandemic: An online experiment. *Social Science & Medicine*, *265*, 113478. https://doi.org/10.1016/j.socscimed.2020.113478 https://doi.org/10.1016/j.socscimed.2020.11347810.1016/j.socscimed.2020.113478

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541 https://doi.org/10.21105/joss.0154110.21105/joss.01541

Manuguerra, M., & Heller, G. Z. (2010). Ordinal regression models for continuous scales. *The International Journal of Biostatistics*, *6*(1), 1–18. https://doi.org/10.2202/1557-4679.1230 https://doi.org/10.2202/1557-4679.123010.2202/1557-4679.1230

Manuguerra, M., Heller, G. Z., & Ma, J. (2020). Continuous ordinal regression for analysis of visual analogue scales: the R Package ordinalCont. *Journal of Statistical Software*, *96*(1), 1–25.

Matejka, J., Glueck, M., Grossman, T., & Fitzmaurice, G. (2016). The effect of visual appearance on the performance of continuous sliders and visual analogue scales proceedings of the 2016 CHI conference on human factors in computing systems, San Jose, California, USA. https://doi-org.ezproxy.lib.monash.edu.au/ https://doi.org/10.1145/2858036.2858063

Meyer, G. (2016). In science communication, why does the idea of a public deficit always return? *Public Understanding of Science*, *25*(4), 433–446. https://doi.org/10.1177/0963662516629747 https://doi.org/10.1177/096366251662974710.1177/0963662516629747

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*(5), 711. https://doi.org/10.1037/0033-2909.130.5.711 https://doi.org/10.1037/0033-2909.130.5.71110.1037/0033-2909.130.5.711

Osmundsen, M., & Petersen, M. B. (2020). Framing political risks: Individual differences and loss aversion in personal and political situations. *Political Psychology*, *41*(1), 53–70. https://doi.org/10.1111/pops.12587

Renn, O. (2008). *Risk governance: coping with uncertainty in a complex world*. Earthscan.

Renn, O. (2020). Risk communication: Insights and requirements for designing successful communication programs on health and environmental hazards. In H. D. O'Hair, & R. L. Heath (Eds.), *Handbook of risk and crisis communication* (pp. 80–98). Routledge.

Sheeran, P., & Webb, T. L. (2016). The intention-behavior gap. *Social and Personality Psychology Compass*, *10*(9), 503–518. https://doi.org/10.1111/spc3.12265

Swami, V., & Barron, D. (2020). Analytic thinking, rejection of coronavirus (COVID-19) conspiracy theories, and compliance with mandated social-distancing: direct and indirect relationships in a nationally representative sample of adults in the united kingdom. https://doi.org/10.31219/osf.io/nmx9w

Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2021). It depends on how you ask: measuring bias in population surveys of compliance with COVID-19 public health guidance. *Journal of Epidemiology and Community Health*, *75*(4), 387–389.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105 https://doi.org/10.3758/BF0319410510.3758/BF03194105

Wise, T., Zbozinek, T. D., Michelini, G., & Hagan, C. C. (2020). Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States.

Wolff, W., Martarelli, C. S., Schüler, J., & Bieleke, M. (2020). High boredom proneness and low trait self-control impair adherence to social distancing guidelines during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, *17*(15), 5420. https://doi.org/10.3390/ijerph17155420 https://doi.org/10.3390/ijerph1715542010.3390/ijerph17155420

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin*, *146*(2), 118. https://doi.org/10.1037/bul0000218 https://doi.org/10.1037/bul000021810.1037/bul0000218

Zhang, Y., Jiang, B., Yuan, J., & Tao, Y. (2020). The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study. https://doi.org/10.1101/2020.03.04.20031187

How to cite this article: van Baal, S. T., Walasek, L., Karanfilovska, D., Cheng, A. C., & Hohwy, J. (2021). Risk perception, illusory superiority and personal responsibility during COVID-19: An experimental study of self-isolation attitudes. *British Journal of Psychology*, 00, 1–22. https://doi.org/10.1111/bjop.12554

APPENDIX A

INSTRUCTIONS EXPERIMENT 1

This study is about self-isolation decision making during the Covid-19 crisis.

Imagine that the country you live in is on lockdown during the first wave of the disease, while the number of daily new cases is starting to come down.

Everybody except essential workers is required to self-isolate in order to stop the spread of Covid-19 through the community.

We will prompt you with 30 different scenarios concerning self-isolation. Please take your time as you go through these.

For each scenario, please indicate to what extent you deem it alright for people to go out of the house. We will start with an example, so you can familiarize with the task.

You are asked to indicate how certain you are that it is acceptable for you to go out of the house.

You indicate your answer by clicking on the grey line where you decide. A slider will then appear, which you can drag if you change your mind.

For example:

You want to get the mail from the mailbox outside. How certain are you that it is alright to go out of the house?

INSTRUCTIONS EXPERIMENT 2

This study is about self-isolation decision making during the Covid-19 crisis.

Imagine that the country you live in is on lockdown during the first wave of the disease, while the number of daily new cases is starting to come down.

Everybody except essential workers is required to self-isolate in order to stop the spread of Covid-19 through the community.

We will prompt you with 30 different scenarios concerning self-isolation. Please take your time as you go through these.

For each scenario, please indicate to what extent you deem it alright for you[1] to go out of the house. We will start with an example, so you can familiarize with the task.

[1]In the Other condition, the word 'you' was replaced with 'a person in your neighbourhood'.

APPENDIX B

TABLE A1 List of the 30 scenarios used in this study, with indication of their a priori assigned Risk Level. Shown here with the Self formulation (the Other formulation is in terms‘ of ‘Someone’ rather than ‘You’). Presentation of the scenario was followed by a Communication framing (either Imperative ‘people should stay home, if they can stay home’, or Personal Responsibility ‘people should consider whether they will get close to other people before they go out’, or a Control with no framing). Participants were then asked to indicate on a visual analogue scale how certain they were that it is alright to go out.

Risk level	Personal scenario prompt
High	You want to play basketball because a few of your friends asked you to join them today. You really enjoy playing basketball.
High	You want to go see your partner because you do not live together. Your only option is to go by public transport.
High	You want to give care to a family member who is having a difficult time. You could leave it to one of your siblings, who does not need to use public transport to get there.
High	You want to see your friend who has recently returned from abroad, as their quarantine just finished. You haven't seen them in a very long time and they have invited you over for a drink.
High	You want to return to the supermarket to pick up one item you forgot to buy. This means that you will have to go back into the busy supermarket.
High	You want to go to a birthday party for one of your friends. Their apartment is not big, and there are quite a few people coming.
High	You want to see your friend who lives close by and has invited you over for dinner. Your friend has been diligently self-isolating for 4 weeks.
High	You want to have dinner with your friends. The six of you have not seen each other for a long time. Some are more serious about social distancing than others.
High	You want to go to the shops because you change your mind about what to make for dinner, but need some other ingredients to do it. You have already gone to the shops twice today.
High	You want to go deliver a meal to an elderly person in your close family. You worry that they are lonely.
High	You want to go for a barbecue in the park with your friends. There will be four of you and you have not discussed social distancing with them.
High	You want to go see your partner, but you do not live together. You will go by car, and you can park in front of their house.
High	You want to do some minor repairs around your home. You plan to go to the hardware store on a weekday morning to ask for advice and to buy some materials.
High	You want to help your friend move house because it's hard to find help when everybody is self-isolating. It will be the two of you and their partner.
High	You want to go see your close friend who lives close by. They tell you that they don't take social distancing very seriously.
High	You want to go for a swim at the beach with three of your friends. To get to the beach you will need to go together in one car.
Low	You want to get a new piece of furniture, because you have been spending a lot of time inside. It has been busy at the big furniture stores lately.
Low	You want to get a coffee. You can pay contactless at the popular café where you will have to wait outside on the busy footpath.
Low	You want to play tennis with your tennis partner. You will go early in the morning. You will arrive separately and stay at opposite sides of the court.
Low	You want to spend some time on an outdoor site for your work to finish an overdue job. There will be other people on the site, but there is lots of space.
Low	You want to meet and go for a walk in a forest with three old friends. You have discussed social distancing with them.

TABLE A1 (Continued)

Risk level	Personal scenario prompt
Low	You want to spend some time in the office at your workplace to concentrate on an important task. Most other colleagues on your floor are working from home. You will need to use the lift to get to your floor.
Low	You want to go running on a track by the beach during working hours. It is a popular, quite narrow and winding track.
Minimal	You want to go for a walk in the park at 5pm. It is a popular neighbourhood park with narrow footpaths near your house.
Minimal	You want to go for a run on a popular running track by a park. People will be breathing heavily and running both ways.
Minimal	You want to go for a vigorous walk at sunrise. The walking path is off the beaten track.
Minimal	You want to go play basketball on your own on a Sunday morning. The basketball court is located outside, on the grounds of your quiet neighbourhood school.
Minimal	You want to play a round of golf alone. You have packed drinks and food, so you will not have to buy anything unnecessarily.
Minimal	You want to go sit in the park and read your book. It is a weekday, and the park is quite large.
Minimal	You want to go for a walk in the park close to your house. You plan to go at 7am.

APPENDIX C

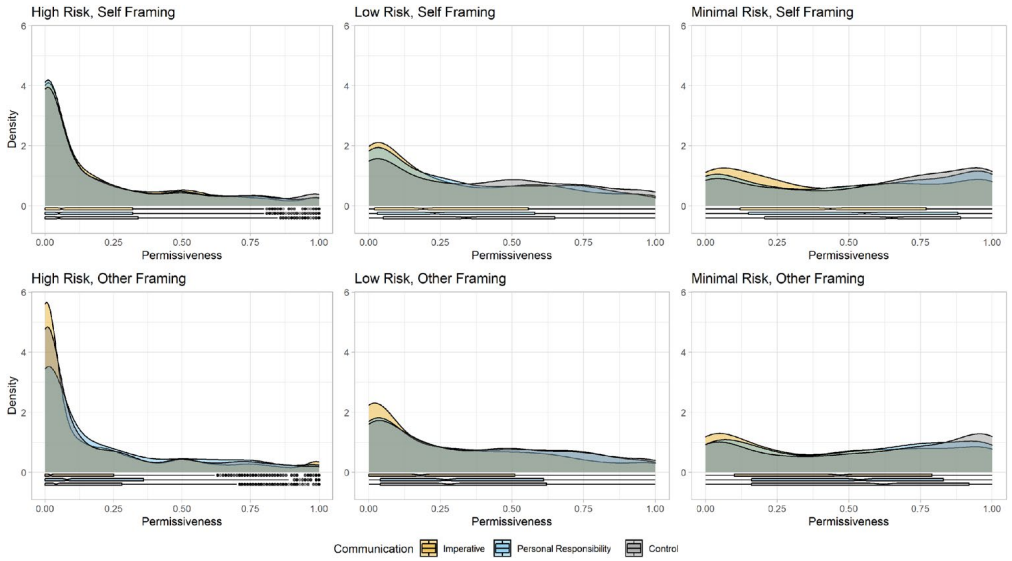

FIGURE A1 Density plots indicating each participants' 30 visual analogue scale responses in experiment 1 parsed by Communication and Risk levels. The yellow area under the curve indicates response in the Imperative condition, the blue area indicates responses in the Personal Responsibility condition, and the grey area indicates responses in the Control condition. Lower visual analogue scale responses indicate a less permissive attitude to going out. The top plots pertain to the Self framing condition, in high-risk, low-risk, and minimal-risk scenarios respectively. The bottom plots pertain to the Other framing condition with the same Risk partitioning.

APPENDIX D

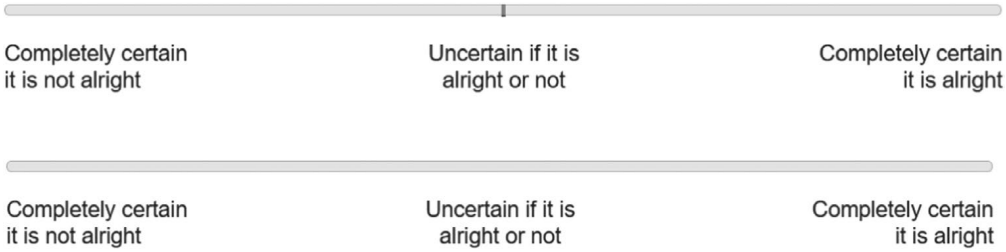

FIGURE A2 Illustrating the VAS Marker manipulation in experiment 2. The midpoint marker is visible on the bar in the top panel, and the control bar without a marker is visible on the bottom panel.

APPENDIX E

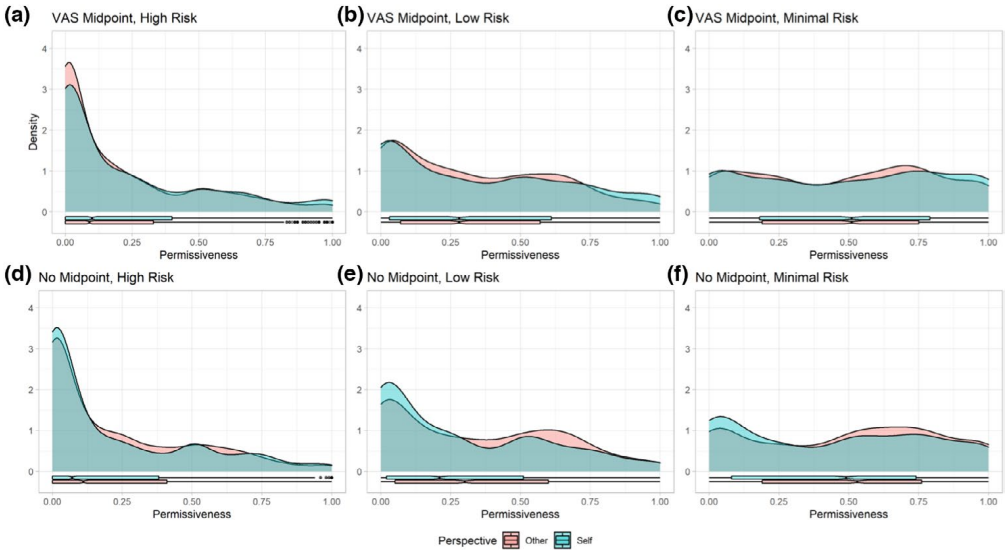

FIGURE A3 Density plots of participants' responses in experiment 2, partitioned by the Vas Marker and Risk factors. The top panels represent trials with a midpoint marker, while the bottom panels represent control trials without one (see Figure A2). The panels on the left represent high-risk scenario questions, the middle panels represent low-risk scenarios, and the panels on the right represent minimal-risk scenarios. The curve with the blue area underneath pertains to scenarios about one's own decisions, while the curve with the red area underneath pertains to scenarios about someone in the participant's neighbourhood's decisions.

Chapter 6

Learning to Control Selfish Interests for the Collective Good

6.1 Linking Text for Chapter 6

In the previous Chapter, I investigated how people judge the acceptability of courses of action that tend to the needs of themselves and those dear to them, when faced with the threat of endangering others and the public good. I showed how these judgments can be skewed towards safer behaviours by using clear language about the necessity of protecting one another in challenging times.

In analysing people's judgments of the acceptability of going out in different scenarios, I found that male participants generally were more permissive than females. This result is unsurprising when viewed through the lens of impulsivity and self-control, as males are known to be more impulsive and less self-controlled than females (e.g., see Chapple & Johnson, 2007; Reimers et al., 2009). The same line of reasoning can be applied to the finding that age negatively correlates with permissiveness of going out across scenarios. But it does imply that these differences, at least in the COVID-19 context, already enter the lead up to behaviour in our conceptual model at the beliefs, judgment, and preferences stage—rather than only through some difference in the reward drive, impulsivity, or self-control stage. As such, it appears that communication strategies are indeed a

good tool to approach these public health behaviours. Arguably, public health communication to increase people's self-control is less likely to succeed that communication targeted at changing people's attitudes.

Naturally, in this applied setting, the type of public good and private gain are relevant for the first and higher-order preferences people might hold. For example, older people, on average, should have a different perception of risk regarding COVID-19. Similarly, males suffer more from mortality-related impacts of poor mental health than females (Kiely et al., 2019). In this way, the expected value of the private gain (here: leaving the house) may differ along the age and gender axes.

Attitudes and beliefs such as these, of course, are not perfect predictors of action. This is why in the current Chapter, I develop a novel paradigm to investigate how people learn to make decisions on whether to pursue individual gain in the face of the potential for collective costs. Specifically, I will investigate how people learn to do this in an interactive, multiplayer economic game where many features of the underlying mechanics of the incentive structure remain hidden. To do this, I incentivise the same reward as was the subject of Chapter 5, namely leaving home, but provide a disincentive for a societal lockdown. I then see whether people can overcome incurring the convenient, immediate reward by exploring the action space to get better results for themselves and others.

In situations where the right course of action is unclear, social norms are strong drivers of behaviour. Relying on the behaviour of peers to inform one's own decision reduces cognitive effort and time exploring the possible options, in some cases it can also increase accuracy (i.e., wisdom of the crowd; El Zein et al., 2019). However, if we have reason to believe that others know just as little about the right course of action as we do, relying on others may have adverse long-run consequences. To elucidate this trade-off, I will investigate how people weigh descriptive information about the behaviour of their peers versus signals from the environment.

This study, then, aims to bring us closer to understanding manifestations of impulsivity and self-control by subjecting people to an environment with an unknown payoff structure. This will reveal how decision-makers deal with a disease outbreak, showing whether they respond adaptively to changes in the environment by spending effort to figure out when contributions to the public good are likely to be effective, or if they

impulsively rely on others to show them the way. This setup also allows us to explore the conditions in which participants control any self-interested tendencies they have at times when this is advantageous to both the agent but also the group in general.

The controlled environment provided by the economic game used in this study also allows for manipulation of the incentive structure, unlike the previous empirical studies in this thesis. I will explore whether changing the incentive structure can influence behaviour, and discuss ingredients for changes to the environment that are likely to achieve the desired effects.

scientific reports

Check for updates

OPEN

Modelling pandemic behaviour using an economic multiplayer game

Simon T. van Baal[1,2], Lukasz Walasek[1] & Jakob Hohwy[2,3]✉

During a pandemic, isolating oneself from the community limits viral transmission and helps avoid repeated societal lockdowns. This entails a social dilemma—either distance oneself from others for the benefit of the public good or free-ride and enjoy the benefits of freedom. It is not yet understood how the unfamiliar incentive structure and interpersonal context presented by a pandemic together modulate individuals' approach to this social dilemma. In this preregistered study, we take a game-theoretical approach and investigate people's decisions to self-isolate, using a novel iterated multiplayer game designed to capture the decision-making environment in the pandemic. To elucidate players' thinking, we use a variation of the strategy method and elicit beliefs about how much others will self-isolate. Players tend to respond to social norms with abidance, rather than transgression; they resist the temptation to freeride when others are self-isolating. However, they deal with exponential growth poorly, as they only self-isolate sufficiently when lockdowns are imminent. Further, increased collective risk can motivate more self-isolation, even though the link between self-isolation and lockdowns is stochastic. Players underreport the influence of others' choices on their own, and underestimate others' self-isolation. We discuss implications for public health, and communication to the public.

Achieving high levels of protective behaviours during a major health crisis such as the COVID-19 pandemic has proven difficult in many countries. Adopting protective behaviours, like staying home when it is feasible to do so, carries substantial individual costs, whereas benefits apply to both the individual and the larger society[1]. Decisions on isolating oneself by staying home thus constitute a social dilemma[2–4]. Due to this conflict between interests of the individual and the collective, many countries had to enforce 'lockdowns' with enforceable stay-at-home orders as high disease prevalence put strain on healthcare systems.

It is not yet known if patterns of behaviour expected in non-pandemic settings also occur in the unusual incentive structure and social context of a pandemic, nor whether that unfamiliar setting modulates attempts to resolve the social dilemma in unexpected ways. Here we study how people make decisions about self-isolation during a pandemic scenario under the threat of lockdown, and look for areas of improvement for increasing the willingness to self-isolate and avoid future lockdowns. We are interested in how people react to two types of informational inputs: situational cues that signal the current viral transmission status and incentive structure changes, and social cues that reveal dominant behaviour patterns.

Studying determinants of protective behaviour during a pandemic. We may try answering questions about the willingness to self-isolate through the use of cross-sectional surveys or longitudinal studies. There has been a proliferation of research in this area, which has brought many insights. However, drawing inferences on what drives behaviour from survey data is challenging considering the likely prevalence of recall bias[5], social desirability bias[6–9], potential confounds, and difficulty in dissociating causality from correlation. Likewise, using observational data on behaviour is often marred with privacy concerns, especially when it comes to personal information like someone's health status.

Instead, we may turn to previous social psychology literature on how humans cooperate, and how they create and respond to social norms[10–12]; social norms are powerful drivers of behaviour, especially when people are unsure of the right course of action, thus predicting high levels of conformity in a pandemic. Although it is also clear that while some act in accordance with social norms, many act in their self-interest[13]. We may also conjecture that because people are suggested to suffer from exponential growth bias[14] (people are, for example,

[1]Department of Psychology, University of Warwick, Coventry, UK. [2]Cognition and Philosophy Lab, Monash University, Melbourne, Australia. [3]Monash Centre of Consciousness and Contemplative Studies, Melbourne, Australia. ✉email: jakob.hohwy@monash.edu

not good at predicting the next number in an exponentially growing sequence of numbers, but see[15]), they may be unwilling to make costly sacrifices for public health when the caseload is still low. However, the COVID-19 pandemic has created a novel, unfamiliar situation for individuals and groups, and it is not clear how those tendencies *ceteris paribus* apply to the unique and unprecedented realities of the pandemic.

Economic games are well-suited to address questions about the determinants of willingness to self-isolate, and the effects of environmental changes because of the controlled setting in which behaviour is studied. At the same time, simply drawing inferences from existing context-free economic games carries the risk of failing to account for the unusual but ubiquitous circumstances of the pandemic. For instance, the stochastic link between behaviour (e.g., self-isolation) and the consequences to that behaviour (e.g., lockdowns) might introduce different choice patterns because it leaves more room for the influence of social norms and beliefs about others' actions. In addition, the initial exponential stage of disease growth, paired with the influence of superspreader events, makes individuals' decisions more influential than in traditional multiplayer games such as the public goods game. To deal with generalisability issues, the literature on economic games has broadened and is now rich with studies tailoring games to specific contexts, such as public goods provision in competing teams[16], contributions to combat climate change[17], building a dam to prevent flooding when there are private solutions available[18]. We contribute to this line of research with a game specifically designed to study the self-isolating behaviours in a pandemic.

In this study, we aim to investigate people's willingness to self-isolate, through the use of a novel context-specific economic game. The game mimics the incentive structure of the pandemic, where self-isolation is costly but benefits the collective through decreased frequency or avoidance of lockdowns. The virus spreads exponentially, depending on whether infected players and non-infected players self-isolate. Naturally, our game provides a simplified view of the realities of an infectious disease outbreak; two notable simplifications are that individuals do not incur the costs of being infected and that participants do not know when they are infected. This allows us to focus on players' responses to collective costs, rather than the individual cost of the disease.

We constructed this game to find out whether the incentive structure is sufficient to produce behaviour patterns that fit with the observations in the COVID-19 pandemic, and to ascertain how dynamically changing situational and social cues influence this behaviour. Important behaviour patterns that we tried to reproduce include the fact that people have been unable to avoid high disease prevalence and lockdowns without intervention, with nevertheless sustained willingness to self-isolate over time (as witnessed with the lack of "behavioural fatigue"[19]). To gain deeper insight into people's responses to situational and social cues, the game uses a variation to the strategy method[20], where participants indicate how much they would self-isolate in hypothetical scenarios and reveal their beliefs about what others will do in the next round.

The Self-Isolation Game presented here is thus related to, but contrasts with both existing context-free, classical games and games tailored to other specific scenarios. This game uses the advantages of economic games and applies them to the context of the pandemic to shed light on how individuals cooperate and behave to protect their interests as well as those of the broader social (public) good. Having participants operate under unfamiliar transmission dynamics, with a stochastic link between choices and outcomes, allows for a better understanding of what drives behaviour during infectious disease outbreaks and improves our ability to gauge the effect of manipulations in the environment.

Phenomena of interest. The current gaps in knowledge on the motivating factors for people's willingness to self-isolate that we investigate here, consist of the following four points. First, our game allows us to determine how people perceive and respond to descriptive social norms in their willingness to self-isolate, and whether they realise the influence these norms have. Descriptive social norms of high self-isolation levels could stimulate conformity[10], which usually takes the form of conditional cooperation in economic games[21–23], although descriptive social norms are less effective in regulatory contexts of prevention[24], such as in a pandemic. Thus, descriptive social norms may promote norm transgression instead; people are tempted because when others are self-isolating, they think they can go out without getting or spreading the disease.

Second, with the use of our game, we assess whether people can deal with the exponential growth of viral transmission appropriately. People tend to underestimate how fast COVID-19 cases grow, imagining they grow linearly—known as exponential growth bias[25,26]—which may inhibit people from self-isolating effectively because it limits their capacity to see the danger in low case numbers. If there is a link between exponential growth bias and behaviour, then we should expect that people respond only to relatively high case numbers.

Third, insights from our paradigm can contribute to research on the reliability of behavioural self-report measures during a pandemic, including the potential interference of social desirability bias[6,8,9]. While the flaws in self-report data have been widely documented, studies using behavioural self-report measures are still over-represented in the pandemic-behaviour literature. Considering that the unfamiliarity of the pandemic context may interact with response biases in new ways, it is important to gain more insight into how self-reports deviate from actual behaviour.

Lastly, we investigate whether increased costliness of the collective risk motivates more self-isolation behaviour when there is a stochastic link between behaviour and outcome (i.e., one individual's action does not automatically translate to viral transmission or lockdown), and compare this to income maximising behaviour, as derived from simulations.

We conduct simulations to identify what a profit maximising player would do for various levels of others' compliance, and different levels of disease prevalence. A profit maximising player would always defect (not self-isolate at all), regardless of others' self-isolation levels, but the simulations reveal that defection becomes relatively more attractive when self-isolation levels in the group are high. Further, a profit maximiser would see no reason to increase their self-isolation when faced with increasing disease prevalence. Considering the substantial compliance with public health restrictions observed during the COVID-19 pandemic, we expect to find players

adopting more cooperative strategies. Therefore, we aim to see if and how players' behaviour deviates from this profit maximising strategy while addressing the four gaps in knowledge mentioned above.

Predictions. We preregistered several predictions on people's behaviour in the Self-Isolation Game, each of which speaks to the above phenomena of interest. For ease of presentation here, the ordering of the predictions differs from the ordering in the preregistration, as follows: H2, H3a, H3b in the present text are H3, H2i and H2ii in the preregistration, respectively.

Considering the game's modelling of the exponential spread of the virus, the least costly way to avoid lockdowns would be to self-isolate to a high degree early on, when there are few infected players (for details, see Methods). We therefore predicted that (H1) players will be unable to deal with *exponential growth*, and self-isolate sufficiently only when it is too late—when there are several infected players in the group. That is, they do not account for exponential growth properly: they forfeit income even when it would be fairly inconsequential to defect because a lockdown is increasingly inevitable with more players being infected.

For the *accuracy of beliefs about others' behaviour*, we predicted that (H2) players will believe that others will self-isolate more when there are more infected players in the group. In an exploratory analysis, we also assess whether participants suffer from *illusory superiority* (also called the better-than-average-effect, referring to the tendency to regard one's own qualities and attributes as superior to others[27]), where they believe others will self-isolate less than they will themselves. We found this tendency in self-reports on staying home during a lockdown[28]. Illusory superiority is relevant in a disease outbreak context because it is known to hamper interpersonal adjustment, and it has been speculated to increase risk taking[27].

Our game bears resemblances to the volunteer dilemma[29] and the step-level public goods game[30], where a sufficient number of players needs to cooperate to produce some binary public good (here: the lack of lockdowns). In contrast to games such as the prisoners dilemma, trust games, and ultimatum bargaining, cooperation and defection tend to coexist in the equilibrium. Therefore, we predicted that (H3a) players will self-isolate less when they believe that others will self-isolate more because defection is least consequential when others are self-isolating rigorously. We also predicted that (H3b) players will report that they would self-isolate less in hypothetical scenarios where others self-isolate more (i.e., we predict that norm transgression tendencies from H3a are reflected in their hypothetical responses).

We reasoned that people would exhibit social desirability bias, and thus we evaluate whether players' *self-reports* in hypothetical scenarios indicate higher levels of self-isolation than what players *actually* choose during their incentivised trials; we predicted that (H4) players will indicate higher levels of self-isolation in hypothetical scenarios than in the subsequent actual (incentivised) trials.

Investigating whether a *collective risk* can motivate larger contributions even when there is a stochastic link between behaviour and outcomes, we predicted that (H5) costlier lockdowns lead to higher self-isolation levels. If lockdowns are costlier, then the benefit of self-isolation increases, and thus we should find that people respond to this by self-isolating more. But the probabilistic influence of self-isolation on disease prevalence may cause players to feel that whether a lockdown occurs or not is out of their control, which makes this worth investigating.

Materials and methods

Participants. The final sample consisted of 134 participants: 57 (42.5%) males, 76 (56.7%) females, and 1 (0.7%) indeterminate/intersex/unspecified with a mean age of 35.5 ($SD = 12.0$). Participants were required to be fluent in English, residents of the UK. Participants were excluded before completion of the experiment if they were diagnosed with dyslexia, dyspraxia, attention deficit hyperactivity disorder, or if they had trouble reading for any reason (including uncorrected abnormal vision).

The data collection was completed between 20 and 23 January 2021 via the Prolific online research participant database. At this stage in the COVID-19 pandemic, the UK was on the cusp of 100.000 COVID-related deaths, and two weeks earlier the National Health Service (NHS) England's national medical director urged people to physically distance because the NHS was under extreme pressure. The country was in lockdown without a clear end date. In addition, 3.07% of participants reported they had received a positive COVID-19 diagnosis.

The experiment was divided into 14 sessions with 11 participants each, for an initial total of 154 participants. The minimum group size after exclusions per session was eight. Any sessions that did not meet the cut-off were excluded from data analysis. 20 participants were excluded, most of whom dropped out while waiting for the experiment to start (19), 1 participant provided low effort data by missing too many responses. We departed from the preregistration here. We stated that we would discard data on a particular DV for participants who provided the same responses in 95% of the rounds for that DV. Seeing as there are legitimate reasons to respond the same each round, we did not go through with this exclusion procedure. Instead, if participants' pages timed out in 30% of the rounds or more, their data was excluded.

Participants were paid £2.00, with a bonus of £0.20 for every 100 points (the endowment for each round). There were 40 rounds, so participants could accrue a maximum of £10. All participants provided informed consent as approved by the Monash University Human Research Ethics Committee (Project ID: 26499).

Apparatus. This experiment was an online study, and participants were asked not to use mobile devices. No default options were used for any of the questions, and the options were always presented left to right in increasing order to avoid any confusion. The pages had timeout timers that would only appear when time was running out. This was done to streamline the experiment and make long waiting times unlikely while limiting the effect on the participants.

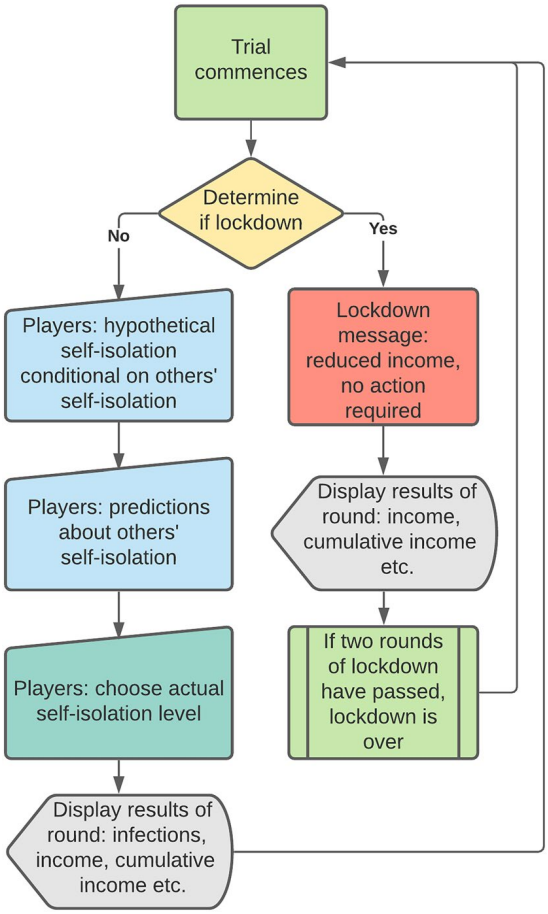

Figure 1. A trial-level schematic of the game from the players' point of view.

Procedure. Participants began the study by completing a short questionnaire, stating their age, their sex, whether they had tested positive for COVID-19 in the past, and in which region within the UK they resided (i.e., England, Scotland, Wales, or Northern Ireland). Further, as an attention check, they were asked which city was not a city in the US, with the list including Tokyo, but with a subheading that read: "Regardless of the right answer, please select Chicago". The attention check was not preregistered, but was added to flag submissions for more careful review. We could not find any reason to exclude the participants who failed the attention check, and found that excluding them would make little difference to the results (see Supplementary Materials for additional information, and a secondary analysis without these participants in the sample). Afterwards, they were presented with an introduction and instructions on how to play the game, which was created using oTree[31].

Participants were made aware that a randomly selected player in the group was infected with COVID-19 and that they could spread it to other players in the group. They did not know who this 'patient zero' was, and were told that they could be patient zero themselves. Additionally, they would not know whether they were infected at any point, nor would they experience a decrease in income if they got infected. When six players were infected with the virus, the group would go into lockdown for two rounds, costing either 60 or 90 points per round, in the Low Cost and High Cost conditions, respectively. After a lockdown, a new patient zero was randomly selected. Players were notified that they were playing to get points and that to do this they could avoid lockdowns by reducing viral transmission through self-isolation (see instructions here https://osf.io/tr6y3/?view_only=aa314332c12c45548484ac084feed2f0).

In each round, the game required several responses from the participants; if one of the pages timed out before they responded, they were not paid for the round. First, we implemented a variation to the strategy method: to see how players would respond to different levels of self-isolation in the group, players were asked "If you knew what the average self-isolation level of the others in the group was in this round, how much would you self-isolate?", with five sub-questions reading: "If the group self-isolation average were—[insert level]". Thus, each of these sub-questions presents a different hypothetical scenario (hereafter: others' self-isolation in scenario). For each sub-question, the response could be one of 5 self-isolation levels: not at all, slightly, moderately, stringently, or completely (hereafter: hypothetical self-isolation). Second, participants would provide their response to "On average, how much do you think others will self-isolate in this round?", choosing one of the aforementioned options (hereafter: beliefs about others' behaviour). Last, they answered: "How much do you actually want to self-isolate in this round?", emphasising that this response would be incentivised (hereafter: incentivised self-isolation, see below). See Fig. 1 for the flow of the experiment from the participants' point of view.

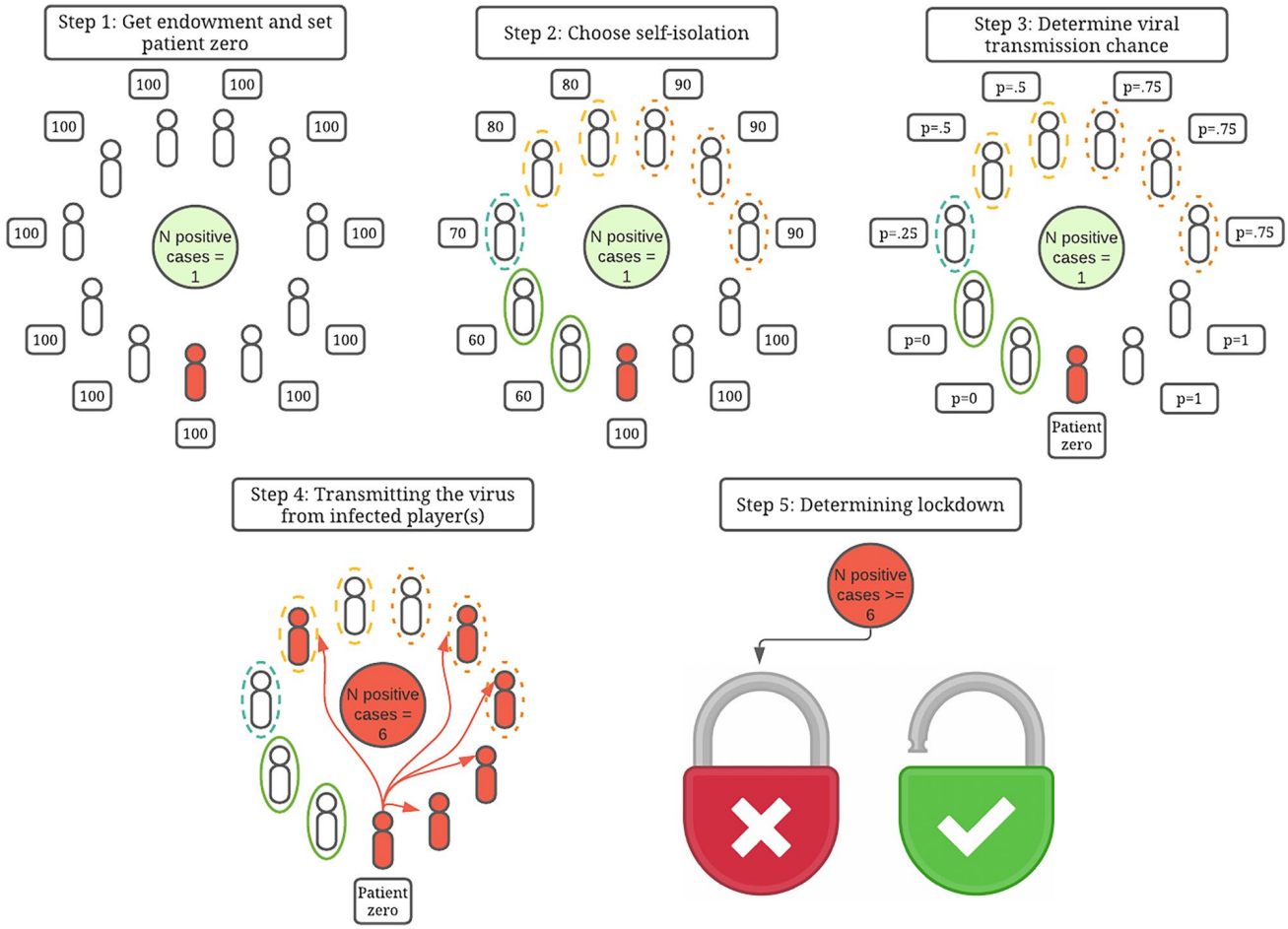

Figure 2. The costs and a demonstration of the viral transmission in the game. Participants all get 100 points of endowment in each (non-lockdown) round. They are then given a chance to choose their level of self-isolation to a maximum of 40% of their income, leaving minimally 60 points. This case depicts a scenario with only one infected player who does not self-isolate at all (at the bottom of the group, in red). Players who aren't yet infected and do not self-isolate at all have 100% chance of getting infected, whereas players who self-isolate moderately (20% of endowment sacrificed) have 50% chance to get infected. The different self-isolation levels are indicated by differences in circles around the players (e.g., the green solid line is complete self-isolation (40% of endowment sacrificed), the dark orange dashed circle is slight self-isolation (10% of endowment)).

Participants' endowment was 100 points in each round and they could sacrifice a part of their income by selecting one of the following options: sacrifice no points by choosing the self-isolation level "Not at all", sacrifice 10% (i.e., 10 points) to slightly self-isolate, 20% to moderately self-isolate, 30% to stringently self-isolate, or 40% to completely self-isolate. Participants were not made aware of how much the self-isolation would decrease the chance of transmission (because this is also not the case in the real world).

The transmission chance was determined through a stochastic process: if an infected player and a healthy player would not self-isolate at all in the same round, then the chance of transmission would be 100%, while if one of them would self-isolate moderately, this would be reduced to 50%. For example, if an infected player would self-isolate moderately, but a healthy player would not self-isolate at all, then the chance of the infected player transmitting the virus to the other player would be 50%. See Fig. 2 for the dynamics of transmission and the by-trial process.

Participants were all subjected to both lockdown cost conditions, and the order of presentation was counterbalanced. After 20 rounds in the initial lockdown cost condition, halfway through the experiment, either High Cost or Low Cost, participants were shown a message stating the new lockdown cost regime. The Low Cost condition was set to 60 points because it would equate incomes of groups where every member is completely self-isolating (i.e., for three rounds: 3*60 points), and not self-isolating at all (i.e., for three rounds: 100 points + 2*40 points in lockdown). The High Lockdown Cost condition was chosen to provide a strong contrast to this situation where not self-isolating resulted in efficiency loss; providing only 2/3 of the income of a fully self-isolating group (100 + 2*10). The range of outcomes for participants was, then, £4.98 (complete self-isolation and maximum number of lockdowns) and £10 (no self-isolation and no lockdowns).

After players entered their choices, they were presented with essential information regarding the results of that round. Specifically, they were shown: how much they had earned (i.e., endowment—self-isolation cost, or lockdown endowment); average self-isolation of others, as a percentage of complete self-isolation; how many infected players there were, and how many players there were in the group; their cumulative earnings in points.

The presented information during the various participant inputs was held constant, and so participants were always shown how many infected players there were in the group, and what the cost of a lockdown would be if it were to occur after that round. For further details of the procedure, see Supplementary Materials.

Analysis. *Simulations.* To compare participants' behaviour to profit maximising strategies we simulated the income generated by adopting different strategies along the empirical boundaries of the game. We vary other players' cooperativeness from complete defection (all other players select 'no self-isolation' in each round), to complete cooperation (all other players select 'complete self-isolation' in each round). We first simulated the game's outcomes for the five possible *static* strategies, where the fictional player of interest maintains the same level of self-isolation throughout the game, among the changing environment of the other 10 players' cooperativeness. For all strategies, income drops strongly as others become less cooperative. However, under both levels of lockdown cost, and the different levels of cooperativeness of the group, less self-isolation is more profitable.

In addition, we modelled the performance of *dynamic* strategies whereby players increase or decrease their level of self-isolation when infections rise. The fictional player of interest starts by choosing moderate self-isolation in every round but will either increase their self-isolation to complete self-isolation ('Moderate to Complete' strategy), or decrease it to no self-isolation ('Moderate to None' strategy) when there are more than two infections in the group. We compare this to the static 'Moderate Cooperator' strategy, where the player always chooses moderate self-isolation in each round.

Preregistered analyses. Based on the within-subjects design with only one manipulated variable (lockdown cost, with two levels), and various observed variables, we opted for a linear mixed modelling approach, which allowed for the controlling of the dependence between observations. There were three dependent variables (DV): incentivised self-isolation, hypothetical self-isolation, and beliefs about others' behaviour. There were three other independent variables (IV) of interest: the cost of lockdown, the number of infections in the group, and the self-isolation of others in hypothetical scenarios. Other observed variables were used as control variables, these were: the round in which the observations were made and the average self-isolation of others in the previous round.

A linear mixed model (LMM) was used to analyse the effects of the number of infected players (H1) and beliefs about others' self-isolation (H3a) on incentivised self-isolation. In this model, the DV was incentivised self-isolation, and the IVs were the number of infections, the self-isolation of others in the previous round, people's beliefs about others' self-isolation, and the round number. The model included participants as a random intercept and the round number as a random slope.

Another LMM was used to analyse the effects of the number of infections on beliefs about others' self-isolation (H2). In this model, the DV was beliefs about others' self-isolation, the IVs were the round number, the number of infections, the self-isolation of others in the previous round. The model also included the participants' IDs as the random intercept and the round number as the random slope. A third LMM was used to analyse the effects of others' self-isolation in a scenario on self-reported hypothetical self-isolation (H3b). In this model, the DV was hypothetical self-isolation, and the IVs were the round number, the number of infections, the self-isolation of others in the previous round, and others' self-isolation in the scenario. The model also included participants' IDs as the random intercept and the round number as the random slope.

A Wilcoxon Signed Rank (WSR) test was used to measure the difference between hypothetical self-isolation and incentivised self-isolation (H4). Here we averaged self-isolation for each category (i.e., hypothetical or incentivised trials) per participant, such that each participant had one observation in each category. Only the hypothetical scenario was used that the participant believed would happen (e.g., when the participant indicated they believed others would choose moderate self-isolation levels, then that hypothetical self-isolation response would be used for this analysis). To measure the effect of the cost of lockdown on incentivised self-isolation (H5), we also used a WSR test. The WSR test was used in these instances because the tests are designed to only detect a main effect, disregarding any interactions.

Analyses were conducted using R[32], mainly relying on the "afex" package[33] and the "emmeans" package[34]. Results were considered significant based on a false discovery rate adjusted α of 0.05.

Exploratory analysis. One exploratory analysis was conducted to investigate whether participants consistently self-isolated more than they believed others would. This analysis consisted of a WSR, testing the difference between incentivised self-isolation (aggregated per player) and beliefs about others' behaviour (aggregated per player).

The preregistration can be found here https://osf.io/xhf7w/?view_only=c3f6f12d380d414f8584e19b7a934294; the data, analysis code, and code for the experimental software are all publicly available on the Open Science Foundation website here https://osf.io/jcmwt/?view_only=d7b6fe0e5e53447ca871312b2b10f779.

Ethical approval. This research was approved by the Monash University Human Research Ethics Committee (Project ID: 26499). The experiment was performed in accordance with relevant named guidelines and regulations. Informed consent was obtained from all participants and/or their legal guardians.

Results

Simulations. The simulations show that the most profitable static strategy is to defect (by never self-isolating), regardless of what the rest of the group chooses to do. It is also clear that the less the others self-isolate, the less profitable each strategy becomes, but the ordering of the profitability of each strategy is largely maintained as the most cooperative strategies are the least profitable. The socially optimal outcomes are that everybody chooses 'complete self-isolation' in the High Lockdown Cost condition, and 'stringent self-isolation' in the Low Lock-

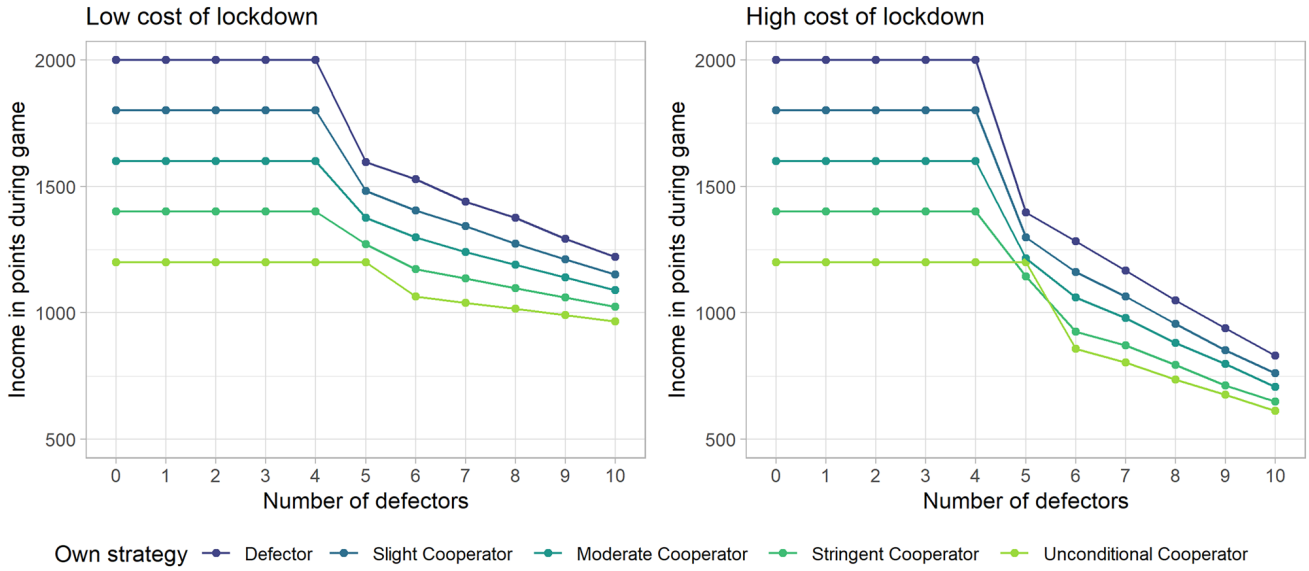

Figure 3. A comparison of average cumulative incomes achieved in simulations of the game through all five possible static strategies. The panel on the left represents the Low Lockdown Cost condition, wherein players earned 40 points per round in lockdown, whereas the panel on the right shows the condition wherein players earned 10 points per round in lockdown. The income earned through adopting each strategy (y-axis) is shown for a different number of defectors and unconditional cooperators in the group (x-axis); the left-most points describe groups where all other players choose 'no self-isolation' in every round (defectors), while the right-most points depict groups where all other players choose 'complete self-isolation' in every round (unconditional cooperators). The upper, purple lines represent the income one would receive when adopting a strategy of choosing 'no self-isolation' every round, and each line below shows the income received through a one increment increase in self-isolation.

down Cost condition. There is also a steep drop in profitability when there are 6 or more players in the group who are not self-isolating at all. This is because lockdowns will happen when six players are infected, affecting players' endowments. See Fig. 3 for a depiction of the cumulative income attained by players per strategy for varying levels of cooperativeness of the others in the group.

When comparing dynamic strategies, where players start with one of the static strategies but adjust their self-isolation level when there are 2 or more infections, we found that decreasing self-isolation when cases increase is strictly dominant over a strategy of increasing self-isolation as well as holding the same strategy. This is especially true when there are relatively few defectors in the group, while producing the same amount of income for < 2 defectors and for > 5 defectors. However, a complete defection strategy is still strictly dominant. See Fig. 4 for a comparison between these dynamic strategies with some of the static strategies mentioned before.

Experimental results. Participants maintained high levels of self-isolation throughout the experiment (see Figure S1 in supplementary materials for average levels of self-isolation per round). Female participants on average sacrificed more of their income to self-isolate ($M = 23.43$) than males ($M = 21.63$), and there was no apparent association between age and self-isolation.

Our results show that H1 and H2 were supported. First, there was a significant effect of the number of infections on both incentivised self-isolation, $b = 1.345$, $SE = 0.115$, $t(3317) = 11.684$, $p < 0.0001$, and on players' beliefs about others' self-isolation, $b = 2.833$, $SE = 0.105$, $t(3300) = 27.106$, $p < 0.0001$. This means that the more infections there were in the group, the more they would self-isolate, and that players (correctly) believed that others would do the same. However, the exploratory analysis on illusory superiority showed that beliefs about others' self-isolation ($Mdn = 20.84$) were lower than players' own incentivised self-isolation ($Mdn = 21.94$), $z = 2.198$, $p = 0.02$. See Fig. 5 for the influence of the number of infections on self-isolation and beliefs about others, and Table 1 for a summary of the output of the three regressions. This means that players believed others' self-isolation would be lower than their own self-isolation.

We found an effect of descriptive social norms in the opposite direction of H3a: there was a significant effect of beliefs about others' self-isolation on incentivised self-isolation, $b = 0.475$, $SE = 0.017$, $t(3420) = 27.745$, $p < 0.0001$. When participants believed that others would choose high levels of self-isolation, they chose higher levels of self-isolation themselves. Further, the results concerning H3a were mirrored in hypothetical scenarios (H3b). Participants reported they would self-isolate more in hypothetical scenarios where others were also self-isolating more, $b = 0.108$, $SE = 0.005$, $t(16{,}810) = 20.470$, $p < 0.0001$, but this effect was markedly smaller. See Fig. 6 for the influence of beliefs about others' behaviour on incentivised self-isolation, compared to the effect of others' self-isolation in a hypothetical scenario. This figure is relevant to evaluating what information about others motivates players' behaviour.

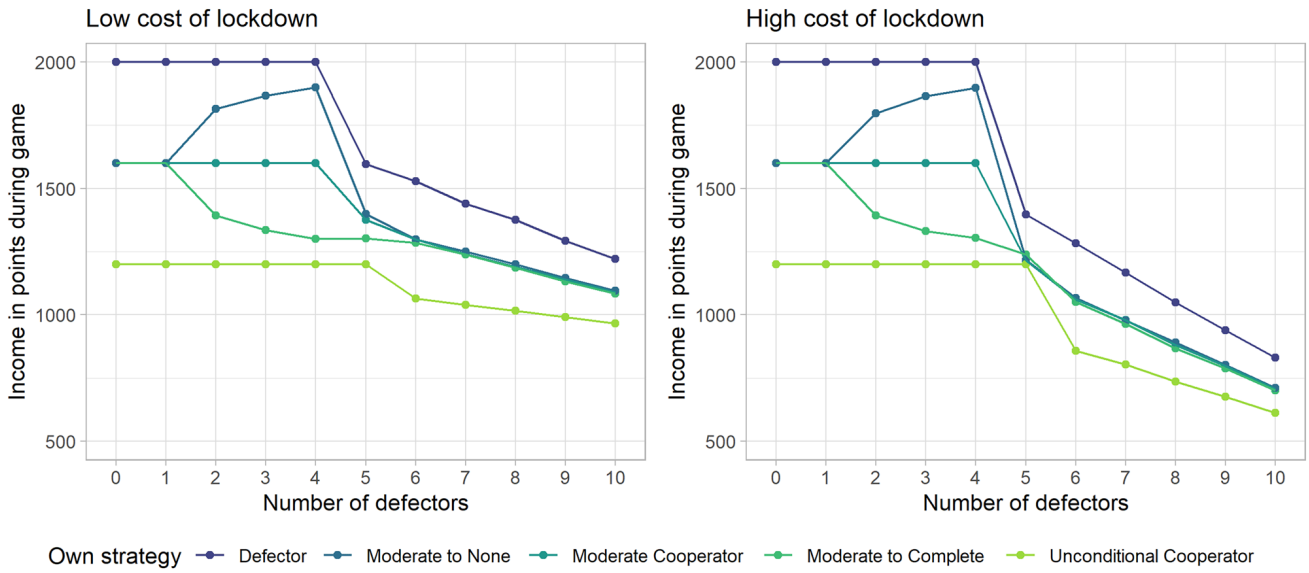

Figure 4. A comparison between dynamic strategies and static strategies. Two dynamic strategies are depicted, namely 'Moderate to None' and 'Moderate to Complete', in these strategies, participants choose 'moderate self-isolation' if there are only one or two infected players in the group. When there are more infections in the group, players who adopt the Moderate to None strategy always choose 'no self-isolation' whereas those adopting the Moderate to Complete strategy then choose 'complete self-isolation'. The other three strategies are static strategies, where defectors always choose 'no self-isolation', moderate cooperators choose 'moderate self-isolation', and unconditional cooperators choose 'complete self-isolation'.

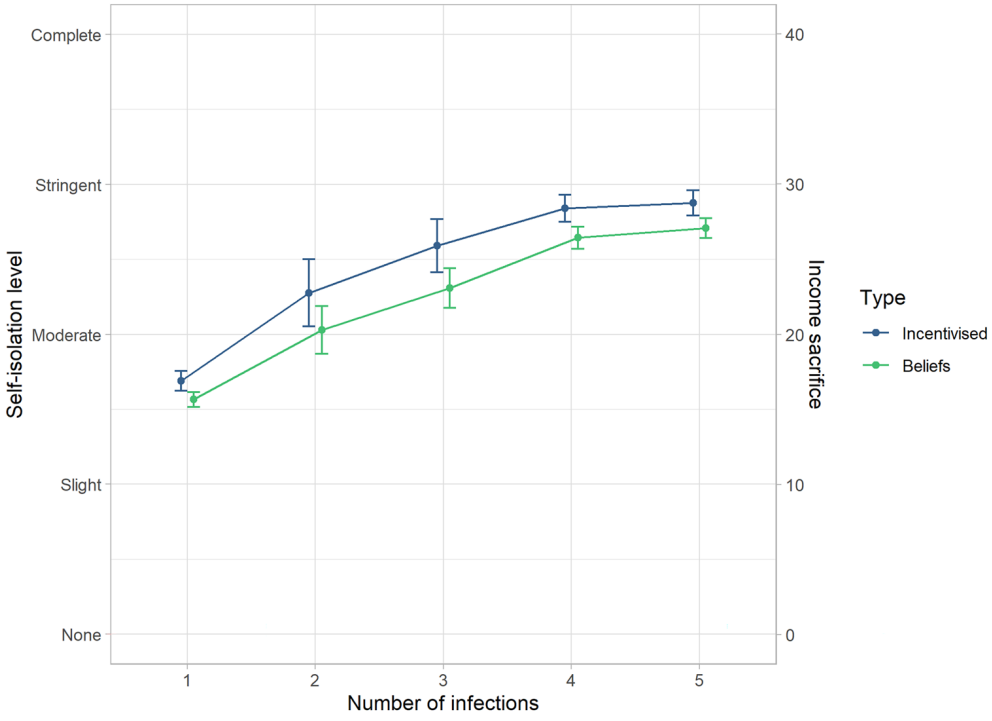

Figure 5. A visualisation of the effects of the number of infections in the group (x-axis) on incentivised self-isolation (in blue) and beliefs about others' self-isolation (in green). The error bars represent 95% confidence intervals.

Turning to the differences between self-reports and incentivised choices, we found no evidence for H4; self-isolation levels in hypothetical scenarios ($Mdn = 22.48$) were not significantly higher than incentivised self-isolation ($Mdn = 22.13$), $z = 0.572$, $p = 0.28$.

H5 was supported, however. There was a significant effect of the cost of lockdown on incentivised self-isolation. Incentivised self-isolation was lower in the Low Lockdown Cost condition ($Mdn = 20.00$) than in the High

	Dependent variable		
	Incentivised self-isolation (1)	Beliefs about others (2)	Hypothetical self-isolation (3)
Regression results H1–H3			
Round number	0.022 (0.025)	− 0.014 (0.023)	0.007 (0.022)
Number of infections	1.345*** (0.115)	2.833*** (0.105)	1.803*** (0.063)
Others in previous round	0.009 (0.007)	0.009 (0.007)	0.007* (0.004)
Beliefs about others	0.475*** (0.017)		
Self-isolation of others in scenario			0.108*** (0.005)
Constant	22.328*** (0.698)	20.776*** (0.556)	22.803*** (0.770)
Log likelihood	− 12,091.020	− 12,064.220	−63,638.690

Table 1. A regression table showing the effects of various predictors on incentivised self-isolation, beliefs about self-isolation of others, and hypothetical self-isolation. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

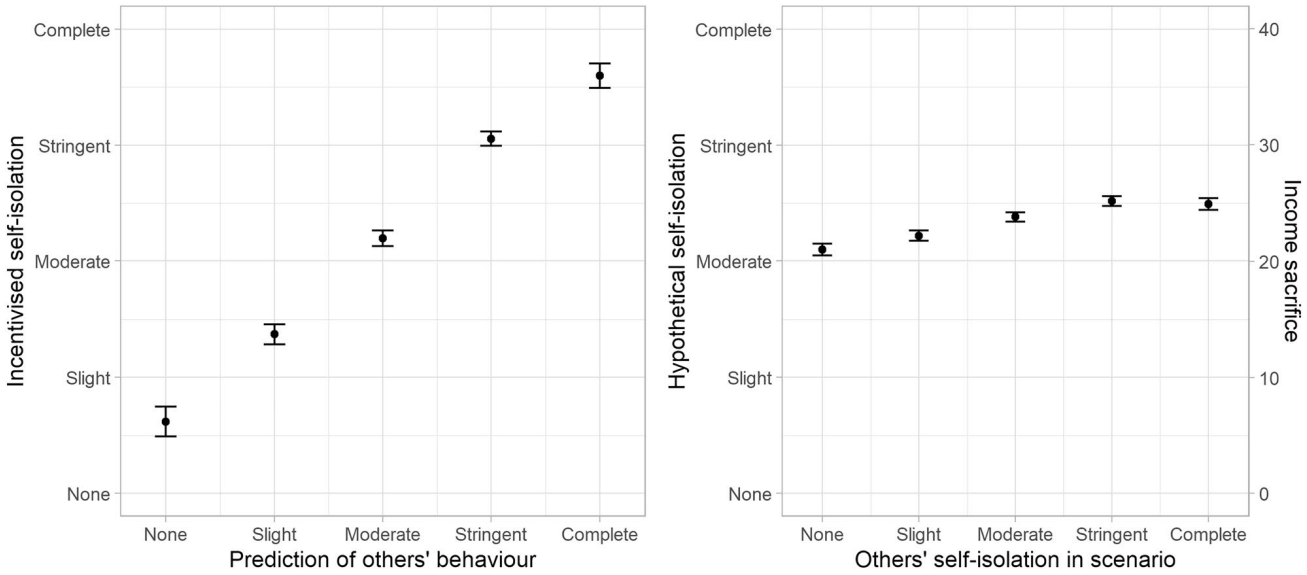

Figure 6. The effects of a player's beliefs about other players' self-isolation (x-axis in the left panel) behaviour on their incentivised self-isolation (y-axis in the left panel), and the effects of the self-isolation of others in each hypothetical scenario (x-axis in the right panel) on self-isolation levels (y-axis in the right panel). The error bars represent 95% confidence intervals.

Lockdown Cost condition ($Mdn = 23.66$), $z = 3.667$, $p < 0.001$. This means that people self-isolated more when the cost of lock-down was high. See Fig. 7 for the interaction between the cost of lockdown and the number of infections on incentivised self-isolation.

Discussion

The Self-Isolation Game we present in this study was designed to capture key properties of the self-isolation decisions that individuals face during an infectious disease crisis, such as the COVID-19 pandemic. The results were as follows: players only tended to self-isolate more as there were more infections in the group, when lock-downs were practically unavoidable. Players systematically underestimate other people's willingness to self-isolate compared to their own—displaying illusory superiority. They also tend to respond to social norms with norm abidance, rather than norm transgression, even though decisions were private. Players' willingness to self-isolate was affected by the cost of lockdown, but only in the second block. We discuss these results in detail below.

Players maintain relatively high levels of cooperation (i.e., self-isolation) throughout the game. In many other repeated games, contributions normally dwindle[35–37], whereas in the Self-Isolation Game, players do not tend to decrease their self-isolation, even though they garner no direct benefits from their input. This implies that the incentive structure and context produce an environment where people remain motivated to cooperate, more so than in a classic repeated public goods game, but mimicking the absence of 'behavioural fatigue'[19] in the general public during the pandemic (see Figure S1 for the development of players' average self-isolation levels

Figure 7. The effect of the cost of lockdown (60 points per trial in green, 90 points per trial in purple) on incentivised self-isolation levels (y-axis, left panel), parsed by the number of infections in the group (x-axis). In the right panel, a line plot of the interaction effect between the self-isolation of others in each hypothetical scenario (x-axis) and the cost of lockdown, on self-isolation (y-axis). The error bars represent 95% confidence intervals.

throughout the game). One other explanation for the continued high levels of self-isolation may be that players receive information on the average contributions of the others in the group, rather than about each individual player's contributions. The way information is presented is particularly relevant during the pandemic because prominent defectors (such as vocal anti-vaccination advocates) can instil feelings of helplessness and encourage non-adherence to guidelines[38]. Collective risk social dilemma games, which also bear significant similarities to the current game, can produce high levels of cooperation under the right conditions[17,39]. These conditions appear to be met with the current parameters of the game, even though the relationship between self-isolation and lockdowns is stochastic, and there is minimal efficiency gain of self-isolation when the cost of lockdown is low.

We predicted that (H1) participants' self-isolation would increase as a function of disease prevalence. We reasoned that people would underestimate the importance of early intervention to stop the spread of the virus; that they would increase their self-isolation only when lockdowns were nearly inevitable. This prediction was supported, as participants persistently displayed relatively low self-isolation when there was only one case in the group, but ramped up their self-isolation gradually as disease prevalence rose, reaching maximum levels when lockdowns were already imminent (see Fig. 5). This result relates to a recent study reporting that COVID-conscious people were focused on the current disease prevalence numbers (rather than their possible contribution to a rise in those numbers)[40]. It is significant because if their goal was to avoid lockdowns, then maintaining a high level of self-isolation throughout would be a superior strategy. In addition, our simulations show that such a strategy is weakly dominated as a profit maximising strategy because players could earn more income by defecting after cases increased.

It thus appears that players are unable to deal with exponential growth appropriately, in a pattern that is consistent with exponential growth bias[25]. This pattern could also have arisen due to players initially exploring the consequences of low self-isolation levels, and subsequently responding to a rise in infections. Either way, rather than players ultimately overcoming this underestimation of the importance of self-isolating early, this pattern then gets ingrained, possibly because it turns into a descriptive social norm. This implies that waiting with the implementation of public health measures until popular support for those measures is attained may cause an ineffective, slow response. Arguably, this is one of the successes in many countries' policy responses to the pandemic; the first containment measures were usually introduced within days or weeks after the first case was detected[41].

Players believed (correctly) that other players would self-isolate more when the number of infected players increased, and thus H2 was supported. But, their beliefs about others' self-isolation levels were systematically lower than their own self-isolation (if beliefs were accurate, these would be equal on average), see Fig. 5. This suggests that players suffered from illusory superiority; they tended to believe others would self-isolate less than they did themselves, a pattern we also found in another study[28]. However, it is more surprising in this context because players have complete knowledge about average self-isolation in previous rounds by other players. The type of illusory superiority we found in this study is also surprising because it goes against previous findings on the *holier-than-thou* effect where people overestimate themselves, but do not underestimate others[42].

Illusory superiority is positively correlated with wellbeing and self-esteem and negatively correlated with depression[43]. Thus, it might be helpful to people to think they are more willing to adopt protective behaviour,

such as self-isolation, than others; it makes them feel like they are good citizens who are contributing to the public good. Illusory superiority also potentially has negative implications for the management of a pandemic because their underestimation of other people's compliance means that people may be more pessimistic about the probability of successful outbreak management. That is, when people underestimate the likelihood that others adhere to public health guidelines, they will infer that the measures are likely to fail. This may increase other forms of social distrust and non-compliance. In addition, although the link has seemingly not been explicitly tested, it has been proposed that illusory superiority may encourage people to engage in moral licensing[44,45]. That is to say that people may believe they have been behaving relatively morally (because they believe others are uncompliant), making it appear more permissible to transgress[46].

The influence of social norms was the opposite of our predictions (H3a, H3b): we hypothesised that participants would be tempted to take advantage of others' higher compliance and self-isolate less because free-riding is less likely to be consequential when others self-isolate more. The simulations show that the extra income that can be earned by defecting is much higher when others are diligent self-isolators, which is similar to the incentive structure in classic games such as the trust game[47]. However, players' self-isolation was highly correlated with their beliefs about others' willingness to self-isolate, favouring an explanation in accordance with the literature on social norms[48] (See Fig. 4), which predicts that people behave per the descriptive social norm especially when they are uncertain about the correct policy[49], but also that cooperative behaviour tends to cascade through social networks[50]. Another possibility is that the context of the pandemic changed how the norms are perceived, turning the behaviour of others into a prescriptive, rather than a descriptive, norm[11]. This is not to say that freerider behaviour does not exist in the pandemic, but it provides grounds to believe that the urge to follow social norms may be strong in this type of context.

The cooperation pattern we observed also relates to the literature on conditional cooperation in public goods games, a well-studied phenomenon[21–23], but the difference here is that the stochastic nature of infections leaves room for sustained conditional cooperation throughout the repeated game. It has also been suggested that many people may not perceive behaviour in the pandemic as a social dilemma, and that general levels of trust and cooperation (i.e., beliefs and behaviours not specifically related to pandemic behaviour) are not a good predictor of intentions to adopt protective behaviours[51]. Thus, managing beliefs about descriptive social norms on willingness to socially distance and stay home could be a powerful communication tool for governments, but more research needs to be done on this topic and how it relates to different domains of behaviour during a pandemic.

We predicted that players would indicate more willingness to self-isolate in hypothetical scenarios than in incentivised trials (H4), as one would expect social desirability bias to work in this direction; people report that they would self-isolate a lot in a certain scenario (often a socially desirable answer), but might not do so. This hypothesis was not supported, closely aligning with the results of Larsen et al.[9] and Gollwitzer et al.[8], who also found no evidence of overreporting willingness to comply with government guidelines. The social desirability bias measurement in this experiment was stringent, however, because unlike in many social desirability bias studies, the self-report measurement was not retrospective (leaving no room for recall bias)—self-reports were temporally proximate to the actual behaviour.

Although, there was a downward bias in the extent to which they indicated their decisions were influenced by the descriptive social norm (see the difference between the left and right panels of Fig. 5). Tuncgenc et al.[52] found that people's decisions in pandemic contexts are highly influenced by close peers. Underestimating or being unaware of that influence would be particularly harmful in a pandemic context because people ('self-isolation role models') could use that information to influence their peers by setting a good example, which could improve their own outcomes (e.g., by avoiding a societal lockdown). It may thus be risky to trust people's insights on what influences their decisions in self-report studies, and this is brought out in an experiment like our game that contrasts a type of self-report with actual incentivised behaviour.

We predicted that (H5) the cost of a lockdown would affect the level of self-isolation players chose in the game because people may be sensitive to the magnitude of the collective risk. This prediction was supported, but it was driven mostly by the second half of the game (after the cost of lockdown was changed). Players increased their self-isolation when they transitioned from the Low Cost to the High Cost condition, and vice versa, while their self-isolation levels in the first half of the experiment were comparable over the two levels of lockdown cost (see Figure S1). Therefore, it is likely that this is a framing effect; people think they need to self-isolate more when the collective cost of not doing so increases. This reinforces the importance of thoughtful management of the public's perception of the collective risks to the community.

Limitations and future directions. The experimental design left out any adverse effects of contracting the virus. The current paper is meant to not only stand on its own, but also serve as a template for how people's behaviour under various payoff structures in future pandemics could be studied, and thus parsimony of the model was an important consideration. Omitting personal costs served to mimic the payoff structure for a key demographic that is difficult to motivate to adhere to public health guidelines: those who believe that catching the virus will not have an impact on them (i.e., the 'infection indifferent')[40]. It also served to ensure that we could capture key features of decision making during the pandemic, namely, how people respond to the risks and costs to the collective, and not whether they respond to their individual cost function. Nonetheless, gauging the impact of different individual costs associated with the virus on behaviour would be a meaningful addition to the current study because it could interact with the collective risk in various ways.

Another issue is that some of the findings in this study will be dependent on parameterization, the country of residence of the sample, or the framing, and teasing these influences apart is challenging. Changing the parameters of the game, the framing, and the sample will likely change the relationships that we found in this paper: for example, if we introduce inequality in endowments, some group members might be unresponsive to changes

in the number of infections; if we change the framing of the game to remove all mentions of the pandemic, the baseline willingness to self-isolate may be lower; different cultures might handle information about other group members differently. Thus, for future research it is important that the rest of the realistic parameter space, the influence of different cultures and nations, and neutral (non-pandemic) framing, be explored to compare to the original findings.

For some findings, though, it seems unlikely that they would be affected by these manipulations. Illusory superiority would likely remain a feature of behaviour during the pandemic we also recorded it in a self-report study[28], and because the game already provides participants with average compliance information. Similarly, we deem it likely to be a robust feature of human behaviour during the pandemic that people's compliance with, and support for containment measures may accumulate slower than would normally be optimal from a public health perspective. That is, people will usually respond to new outbreaks (or perhaps 'threats' in neutral framing) when the pathogen has already taken hold in the community because only then will it become clear that it is a 'problem' and not an overreaction to try to contain it.

The group size in the experiment was kept at eleven participants, even in case of dropouts. Participants who dropped out were marked as complete self-isolators during the game, and therefore the self-isolation rigour in the group was artificially inflated (although these observations were dropped from analysis). This means that information presented about others' self-isolation rigour was often exaggerated slightly (19 dropouts over 14 sessions led to an average ~ 6% overestimation of others' compliance). With this in mind, the illusory superiority found here is even more striking because, given that others' self-isolation was slightly inflated in this experiment, one would expect that players believe others self-isolate more than they do themselves. And even though we see no reason that the other findings of this experiment would change through allowing the number of players in the group to decrease after dropouts, further study should investigate any differences that would occur.

Further, the experimental software would reset and switch to a new patient zero if there was no transmission for three consecutive rounds. This modelled effective management of outbreaks and was inspired by the situation in countries where there was no community transmission, such as New Zealand (at the time of the study being conducted). Namely, if there is one positive case, and they do not spread the virus, then they will no longer be infectious after a while and the risk they pose to the community dissipates. In this scenario, the next possibility for a new chain of infection would be a different patient zero (e.g., from international travel). However, in all other scenarios, players would not similarly lose their infectiousness after three rounds, which would be more realistic, but we decided not to include this to preserve the parsimony of the model. Further study could look at how people's behaviour changes if they know that they are infected and if they know when they will no longer be infected.

Conclusion

The Self-Isolation Game shows several important behavioural tendencies in a disease transmission suppression setting that would have been difficult to investigate with classical games or survey-based research. We found that people exhibit illusory superiority, they can be motivated to sacrifice through a collective risk, and they tend to follow social norms even when doing so is disadvantageous for themselves and the group.

Policymakers should be aware of the effect that perceived descriptive social norms have on people's willingness to self-isolate. Creating the perception of social norms around willingness to maintain physical distance and to stay home may alleviate the need to impose restrictions. Careful framing of the high cost of a lockdown may also induce more willingness to adopt protective behaviours.

Waiting for public support to impose measures against viral transmission may also be ill-advised; people are late to respond to growing outbreaks, which is why policy interventions may need to be introduced before the public perceive the threat as serious. Finally, people should be aware overestimating the rigour of one's own protective behaviours compared to others is common, and use this to promote understanding and compassion for each other during difficult times.

Data availability

All data are available here, together with the analysis code, and the code for the experimental software.

Received: 23 February 2022; Accepted: 28 July 2022
Published online: 05 August 2022

References

1. Fazio, R. H. *et al.* Social distancing decreases an individual's likelihood of contracting COVID-19. *Proc. Natl. Acad. Sci.* https://doi.org/10.1073/pnas.2023131118 (2021).
2. Dawes, R. M. Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193 (1980).
3. Kerr, N. L. Motivation losses in small groups: A social dilemma analysis. *J. Pers. Soc. Psychol.* **45**, 819. https://doi.org/10.1037/0022-3514.45.4.819 (1983).
4. Van Lange, P. A., Balliet, D. P., Parks, C. D. & Van Vugt, M. *Social dilemmas: Understanding human cooperation* (Oxford University Press, 2014).
5. Coughlin, S. S. Recall bias in epidemiologic studies. *J. Clin. Epidemiol.* **43**, 87–91 (1990).
6. Timmons, S., McGinnity, F., Belton, C., Barjaková, M. & Lunn, P. It depends on how you ask: Measuring bias in population surveys of compliance with COVID-19 public health guidance. *J. Epidemiol. Commun. Health* **75**, 387–389 (2021).
7. Van de Mortel, T. F. Faking it: Social desirability response bias in self-report research. *Aust. J. Adv. Nurs.* **25**, 40 (2008).
8. Gollwitzer, A. *et al.* Linking self-reported social distancing to real-world behavior during the COVID-19 pandemic. *Soc. Psychol. Personal. Sci.* https://doi.org/10.1177/19485506211018132 (2021).
9. Larsen, M., Nyrup, J. & Petersen, M. B. Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias?. *J. Behav. Public Adm.* https://doi.org/10.30636/jbpa.32.164 (2020).

10. McDonald, R. I. & Crandall, C. S. Social norms and social influence. *Curr. Opin. Behav. Sci.* **3**, 147–151 (2015).
11. Bicchieri, C. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (Oxford University Press, 2016).
12. Fehr, E. & Fischbacher, U. Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 185–190 (2004).
13. Fehr, E. & Gintis, H. Human motivation and social cooperation: Experimental and analytical foundations. *Ann. Rev. Sociol.* **33**, 43–64 (2007).
14. Levy, M. R. & Tasoff, J. Exponential-growth bias and overconfidence. *J. Econ. Psychol.* **58**, 1–14 (2017).
15. Hamann, H. On getting it right by being wrong: A case study of how flawed research may become self-fulfilling at last. *Proc. Natl. Acad. Sci.* **119**, e2122274119. https://doi.org/10.1073/pnas.2122274119 (2022).
16. Kugler, T., Rapoport, A. & Pazy, A. Public good provision in inter-team conflicts: Effects of asymmetry and profit-sharing rule. *J. Behav. Decis. Mak.* **23**, 421–438 (2010).
17. Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A. & Marotzke, J. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc. Natl. Acad. Sci.* **105**, 2291–2294 (2008).
18. Gross, J. & De Dreu, C. K. Individual solutions to shared problems create a modern tragedy of the commons. *Sci. Adv.* **5**, eaau7296 (2019).
19. Sibony, A.-L. The UK COVID-19 response: A behavioural irony?. *Eur. J. Risk Regul.* **11**, 350–357 (2020).
20. Selten, R. (Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie).
21. Keser, C. & Van Winden, F. Conditional cooperation and voluntary contributions to public goods. *Scand. J. Econ.* **102**, 23–39 (2000).
22. Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J. & Sutter, M. Conditional cooperation on three continents. *Econ. Lett.* **101**, 175–178 (2008).
23. Fischbacher, U., Gächter, S. & Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).
24. Melnyk, V., van Herpen, E., Fischer, A. R. & van Trijp, H. C. Regulatory fit effects for injunctive versus descriptive social norms: Evidence from the promotion of sustainable products. *Mark. Lett.* **24**, 191–203 (2013).
25. Banerjee, R., Bhattacharya, J. & Majumdar, P. Exponential-growth prediction bias and compliance with safety measures related to COVID-19. *Soc. Sci. Med.* **268**, 113473. https://doi.org/10.1016/j.socscimed.2020.113473 (2020).
26. Lammers, J., Crusius, J. & Gast, A. Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proc. Natl. Acad. Sci.* **117**, 16264–16266. https://doi.org/10.1073/pnas.2006048117 (2020).
27. Zell, E., Strickhouser, J. E., Sedikides, C. & Alicke, M. D. The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychol. Bull.* **146**, 118 (2020).
28. van Baal, S. T., Walasek, L., Karanfilovska, D., Cheng, A. C. & Hohwy, J. Risk perception, illusory superiority and personal responsibility during COVID-19: An experimental study of attitudes to staying home. *Br. J. Psychol.* **113**, 608–629 (2022).
29. Archetti, M. & Scheuring, I. Coexistance of cooperation and defection in public goods games. *Evolution* **65**, 1140–1148. https://doi.org/10.1111/j.1558-5646.2010.01185.x (2011).
30. Suleiman, R. & Rapoport, A. Provision of step-level public goods with continuous contribution. *J. Behav. Decis. Mak.* **5**, 133–153. https://doi.org/10.1002/bdm.3960050205 (1992).
31. Chen, D. L., Schonger, M. & Wickens, C. oTree—An open-source platform for laboratory, online, and field experiments. *J. Behav. Exp. Finance* **9**, 88–97 (2016).
32. R Core Team, R. (R Foundation for Statistical Computing Vienna, Austria, 2013).
33. Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. S. *afex: Analysis of factorial experiments*. R package version 0.13-145 (2015).
34. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. *Emmeans: Estimated marginal means, aka least-squares means*. R package version, vol. 1, 3 (2018).
35. Engle-Warnick, J. & Slonim, R. L. The evolution of strategies in a repeated trust game. *J. Econ. Behav. Organ.* **55**, 553–573 (2004).
36. Andreoni, J. Why free ride?: Strategies and learning in public goods experiments. *J. Public Econ.* **37**, 291–304 (1988).
37. Camerer, C. & Weigelt, K. Experimental tests of a sequential equilibrium reputation model. *Econom. J. Econom. Soc.* **56**, 1–36 (1988).
38. Prieto Curiel, R. & González Ramírez, H. Vaccination strategies against COVID-19 and the diffusion of anti-vaccination views. *Sci. Rep.* **11**, 6626. https://doi.org/10.1038/s41598-021-85555-1 (2021).
39. Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci.* **108**, 11825–11829 (2011).
40. Shelat, S., Cats, O. & van Cranenburgh, S. Traveller behaviour in public transport in the early stages of the COVID-19 pandemic in the Netherlands. *Transp. Res. Part A Policy Pract.* **159**, 357–371. https://doi.org/10.1016/j.tra.2022.03.027 (2022).
41. Waitzberg, R. *et al.* Early health system responses to the COVID-19 pandemic in Mediterranean countries: A tale of successes and challenges. *Health Policy* **126**, 465–475. https://doi.org/10.1016/j.healthpol.2021.10.007 (2022).
42. Epley, N. & Dunning, D. Feeling" holier than thou": Are self-serving assessments produced by errors in self-or social prediction?. *J. Pers. Soc. Psychol.* **79**, 861 (2000).
43. Taylor, S. E. & Brown, J. D. Illusion and well-being: A social psychological perspective on mental health. *Psychol. Bull.* **103**, 193 (1988).
44. Leviston, Z. & Uren, H. V. Overestimating one's "green" behavior: Better-than-average bias may function to reduce perceived personal threat from climate change. *J. Soc. Issues* **76**, 70–85. https://doi.org/10.1111/josi.12365 (2020).
45. Bergquist, M. Most people think they are more pro-environmental than others: A demonstration of the better-than-average effect in perceived pro-environmental behavioral engagement. *Basic Appl. Soc. Psychol.* **42**, 50–61. https://doi.org/10.1080/01973533.2019.1689364 (2020).
46. Blanken, I., van de Ven, N. & Zeelenberg, M. A meta-analytic review of moral licensing. *Pers. Soc. Psychol. Bull.* **41**, 540–558 (2015).
47. Berg, J., Dickhaut, J. & McCabe, K. Trust, reciprocity, and social history. *Games Econ. Behav.* **10**, 122–142 (1995).
48. Frey, B. S. & Meier, S. Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *Am. Econ. Rev.* **94**, 1717–1722. https://doi.org/10.1257/0002828043052187 (2004).
49. Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W. & Laland, K. N. The evolutionary basis of human social learning. *Proc. R. Soc. B Biol. Sci.* **279**, 653–662. https://doi.org/10.1098/rspb.2011.1172 (2012).
50. Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proc. Natl. Acad. Sci.* **107**, 5334–5338. https://doi.org/10.1073/pnas.0913149107 (2010).
51. Romano, A. *et al.* Cooperation and trust across societies during the COVID-19 pandemic. *J. Cross Cult. Psychol.* **52**, 622–642. https://doi.org/10.1177/00220221209889l3 (2021).
52. Tuncgenc, B. *et al.* We distance most when we believe our social circle does. https://doi.org/10.31234/osf.io/u74wc (2020).

Acknowledgements

This research is supported by a Monash-Warwick Alliance Accelerator grant to S.T.vB., J.H. and L.W. The Three Springs Foundation supports J.H. through the Monash Centre for Consciousness and Contemplative Studies. We would like to thank Jinhee Kim for helping to test the experimental software.

Author contributions

S.T.v.B.: Conceptualisation, Methodology, Software, Validation, Data Curation, Formal Analysis, Investigation, Writing—Original Draft, Writing—Review and Editing, Visualisation, Project Administration; L.W.: Conceptualisation, Methodology, Resources, Writing—Original Draft, Writing—Review and Editing, Supervision, Funding Acquisition; J.H.: Conceptualisation, Methodology, Resources, Writing—Original Draft, Writing—Review and Editing, Supervision, Funding Acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-17642-w.

Correspondence and requests for materials should be addressed to J.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6.8 Supplementary Materials

Here we report details of the experimental procedure that will be beneficial for the replicability of the study, as well as the results of the secondary analysis that checks the robustness of the results against excluding all participants who did not pass the attention check.

6.8.1 Methods

Participants were not made aware of how many rounds they had completed, and were told the experiment would be anywhere from 30-50 rounds (the sessions were always 40 rounds), in order to prevent any 'last round effects', meaning to prevent participants from returning to the Nash Equilibrium of the stage game (a one-round equivalent), which is to not self-isolate at all. This is the Nash Equilibrium in this particular setup because lockdowns (i.e., endowment deductions) follow in the next round, which is not possible in a stage game.

Participants were required to provide their choices before pages timed out, otherwise they would not be paid for that round. If participants did not submit their answers on time, the experimental software would mark their answer as complete self-isolation, so as not to negatively impact other players in the group (this happened infrequently, on average .32 times per participant after exclusions). This was done to ensure data quality. Participants were then able to progress to the results section, and subsequently to the next round.

We also chose to maintain the 11-player group size, even if there were dropouts, so as not to confuse the participants and change the dynamics of the experiment too much. In order to do this, dropouts had to be treated as complete self-isolators (this way the others would not be adversely affected by their absence). As a result, 19 dropouts over 14 sessions led to a minor overestimation (average 6%) of others' compliance.

The experiment would also automatically 'reset' and randomly select a new patient zero if none of the players transmitted the virus to anyone else for 3 rounds. This resetting modelled people losing infectiousness after a while, and it meant that the players had successfully managed a chain of transmission by self-isolating enough. However, this is

not part of the results as it is not particularly informative in itself, given that this is highly dependent on patient zero's self-isolation rigour. This happened 1.93 times per session on average (maximum 3 times).

Participants were shown a 'lockdown message' during each lockdown round, specifying how costly the lockdown would be (either 60 or 90 points), that it would last two rounds, and that they would not need to make any of the decisions they normally would. The ambiguous effect of the cost of lockdown can be seen in Figure S1. Participants start at a similar level of self-isolation until after the cost of lockdown is changed, at which points the two groups diverge—presumably because they infer increasing (decreasing) their self-isolation in response to an increase (decrease) in the cost of lockdown is the correct response.

6.8.2 Results

For this experiment, the sample was split into two groups: those who were subjected to the high lockdown cost first and the low lockdown cost after, and those who were subjected to the reverse order. To show the effect of the cost of lockdown on participants' behaviour, accounting for order effects, we plot the average contribution of participants per round in each group (see Figure S1).

6.8.3 Secondary Analysis

In the original analysis we did not exclude those who failed the attention check. There were no signs that data quality was compromised for these participants, nor was there any difference in the statistical test results. For transparency, we reanalysed the data without those participants to show that the results are robust to their exclusion.

Just as in the primary analysis, H1 and H2 are supported, and we observed an effect in the opposite of the predicted direction for H3a and H3b. See Table S1 for the regression results pertaining to H1, H2, and H3a,b. The results for H4 and H5 were also mirrored in the secondary analysis: H4 was not supported, $z = .695$, $p = .24$, but H5 was supported, $z = 3.862$, $p < .0001$. Furthermore, we found that participants also exhibited illusory superiority in this reduced sample, $z = 2.245$, $p = .01$.

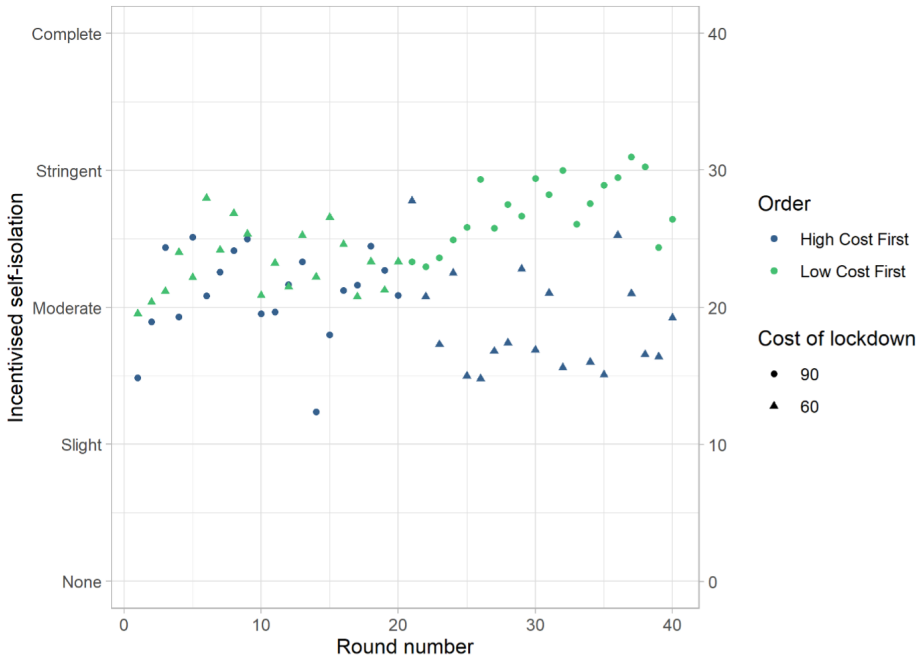

FIGURE S1: The average self-isolation levels throughout the game. Each point represents the average self-isolation (y-axis) of all groups during that round (x-axis) in one of the two orders in which the lockdown cost was changed, provided they were not in lockdown. The colour of the points is determined by the order in which the cost of lockdown was administered (blue is when the high lockdown cost condition was first, and green is when the low lockdown cost condition was first). The shape is determined by the cost of lockdown in that round (circles for high cost of lockdown, triangles for low cost of lockdown).

Chapter 7

Conclusion

This thesis has explored the manifestations of impulsivity and self-control in the context of time preferences, controlling urges and cravings, and selfish versus prosocial motivation in the public health context. Through a combination of theoretical arguments, decision-making and behavioural experiments, survey studies, and computational modelling, we have gained new insights into the nature of these constructs and their role in public health.

The current thesis set out to conceptually dissociate impulsivity, self-control, and time preferences, to study the manifestations of these concepts empirically, and to see if these manifestations may be stimulated or avoided by changing the way individuals think or by adjusting the environment.

7.1 Dissociating Impulsivity and Self-Control and Time Preferences

To dissociate impulsivity, self-control, and time preferences, I used varied conceptualisations from the literature to identify the core components of each construct. I used examples and surveyed empirical evidence from the literature to build theoretical arguments on the overlap of these constructs and identify in which scenarios they come apart.

I have argued that future-oriented time preferences are theoretically (though not always practically) orthogonal to impulsivity and self-control. People can be willing to wait large amounts of time for incrementally larger rewards because they value them more, but they might well be doing this impulsively. That is, they might choose to wait for future rewards without thinking about their decision, possibly forfeiting more meaningful rewards in the present.

Similarly, the relationship between a preference for future rewards and self-control is not a straightforward one. While it is commonly thought that a willingness to wait for future rewards is a mark of self-control, both colloquially and in the literature on intertemporal choice (Ainslie, 1975), it is also possible future-orientedness and self-control are opposing forces within an agent. One may value spending more time with one's family but be tempted to instead work hard and long hours in the office for a better future. Some may then understandably regret their inability to resist this temptation to work long hours. This is an example of a situation where their self-control is ill-equipped to tip the balance of their preference for future rewards over those in the present.

My theorising about impulsivity, self-control, and time preferences led to a concept model for how these cognitive processes interact and influence behaviour. This model is one where the valuation of rewards in the present and the future is separate from, but influences, impulsive and self-controlled behaviour.

The conceptualisation and model I offered for the relationship between impulsivity, self-control, and time preferences (see Figure 1.3) also resulted in testable predictions: the model suggests that time preferences have a causal relationship with impulsivity and self-control. Namely, when an agent is highly biased toward immediate rewards, then they will be more likely to act impulsively when in proximity to an immediate reward, whereas if an agent is highly future-oriented, they might act impulsively when presented with saving and investment decisions, committing themselves to a course of action that limits their options in the meantime. Moreover, I posited that higher-order time preferences, where different from first-order preferences, can motivate an agent to use their self-control, but that higher-order preferences need not be more future-oriented than first-order preferences.

Our findings suggest, therefore, that impulsivity and self-control are complex and multifaceted constructs that can have both positive and negative consequences. They both

relate to time preferences, but the nature of this relationship depends on the dispositions and context of the agent.

7.2 Higher-Order Preferences in Intertemporal Choice

Using this understanding of how impulsivity and self-control influence behaviour, I set out to explore whether directing deliberation toward higher-order preferences and evaluative judgment could systematically change people's decision-making processes and outcomes.

I found that when deliberation is directed toward such higher-order preferences, people tend to become more *future-oriented* for decisions with small stakes, and more *dynamically consistent* with larger stakes (i.e., people revert their preferences less as the smaller-sooner reward becomes more immediate). But we found that there is much room for further exploration of how the underlying decision-making process is altered by different ways of deliberating a decision.

Further, considering that people change their decisions when they think about options in relation to their higher-order preferences, it appears that a promising future area of study is to see whether focusing on higher-order preferences may help people use their self-control. This focus might increase the perceived discrepancy between first-order and higher-order preferences, or decrease the 'pull' of the first-order preferences, thus motivating the agent to use self-control.

7.3 The Influence of State Changes on Urges and Self-Control

In an experimental investigation of how people deal with urges in their daily lives, working toward the goal of understanding manifestations of impulsivity and self-control, I showed that hunger and being in a heightened impulsive state amplifies the experience of urges, making them more difficult to control (the latter is a finding that agrees with the model put forward in Chapter 1). Hunger also impaired self-control in our sample, making it more challenging to control urges, even when controlling for the intensity of

those urges. I view these findings through the lens of a dual-process model where the potency of people's first-order preferences (which I take to be loosely equivalent to urges), depends highly on their current state, but that their capacity for self-control remains more constant—only the motivation to use it might diminish (for a similar argument on ego depletion, see Friese et al., 2019) as the concern with one's higher-order preferences fluctuates.

I then showed how urges in the infectious diseases domain can also be influenced by subtly changing people's mental states. Administering daily prompts that ask participants to imagine a better future or to be compassionate with people who have been adversely affected by the infectious disease in question can decrease the intensity of urges. This again resonates with the idea introduced in Chapter 2 that temptations and urges (i.e., the pull of first-order preferences) can be sensitised or desensitised by mental states, perhaps through lowering momentary impulsivity, (see also Milyavskaya et al., 2015), which would agree with the findings in Chapter 3. The model introduced in Chapter 1 suggests that the effects observed in Chapter 4 are caused by diminishing state impulsivity, which moderates the relationship between reward valuation and behaviour. Though this is yet to be explicitly tested because we did not administer a state impulsivity scale in Chapter 4, it appears that the evidence from Chapters 2, 3, and 4 speak to the strengths of the model from Chapter 1.

7.4 Epistemology and Learning in Balancing Own Needs and the Collective Good

After exploring how urges are sensitised and dealt with in different states, and seeing whether changing people's states can change the experience and the way people deal with urges, I turned to investigate the beliefs and motivations behind people's behaviour in the public health context. For the next two studies, I conceptualised that people would feel urges to leave their homes, but that they would arrive at differing views on whether this would be acceptable in different scenarios (a kind of higher-order preference, or evaluative judgment, depending on the agent). These two Chapters, then, speak to the drivers of the potential intra-personal conflict of desires and urges to leave home on the

one hand, and one's evaluative judgment and higher-order preferences on the other, as discussed in Chapter 1.

I first focused on what influences people's beliefs about the acceptability of leaving their homes during an infectious disease-related lockdown. I showed that people exhibit a degree of sophistication as to which scenarios are most risky from a public health perspective, by comparing their judgments to those of two public health experts. I also found that even though people do not judge the acceptability of others' behaviour more harshly (at least *a priori*), most suffer from illusory superiority in the public health context; people are inclined to believe they are diligent rule followers when compared to others. I speculate that this might result in a moral licensing effect, where participants judge it to be acceptable to give in to their impulses once in a while because they deem themselves to be morally good on the whole. Putting this into the perspective of my conceptualisation of impulsivity and self-control, this means they likely think their past behaviour fits with their higher-order preferences, and thus there is no reason to use their self-control to reign in their first-order preferences.

In an effort to gauge whether it is possible to influence people's attitudes toward the permissiveness of going out during a lockdown and their tendency to view low-risk scenarios differently, I paired scenarios with communication strategies that either emphasised personal responsibility or were imperative. Imperative communication was effective in reducing permissiveness, especially so in lower-risk scenarios, effectively mirroring that communication can be a useful tool for increasing contributions to a public good (Isaac & Walker, 1988). One possible mechanism here is that using an imperative to communicate a social norm (e.g., stay home if you can stay home) effectively moves the boundary for what being a good member of society is, and thus, people whose higher-order preferences contain the wish to be a good member of society may be motivated to control themselves more.

In Chapter 6, a study based on an economic game, we observed how participants learn about the decision-making environment in an infectious disease outbreak. In this game, participants were incentivised to want to leave their isolation; I operationalised urges to leave home as a monetary incentive. There were many parallels in the findings between this study and Chapter 5. I found, for example, that participants tended to

underestimate others' contributions to the public good, even though they had perfect information about previous trials.

Chapter 6 also showed that, instead of exploring the action space, people tended to rely overly on what others do in an unfamiliar environment, most likely to reduce cognitive effort. In the case of this study, that policy almost universally led to a sub-optimal outcome for individuals and the collective alike. Participants tried to 'cash in' when there were few infected players in the group, while it would have been prudent, both from an individual and a collective perspective, to postpone doing this until later when avoiding a lockdown was no longer possible. When we cast this into the light of the model devised in Chapter 1, it appears that most players were unable to break the link between their reward valuation (here, the incentives for private gains by leaving isolation) and their behaviour. They could have used their self-control to explore the action space, effectively seeking information, to either maximise overall individual income or the income of the group, depending on the agent's evaluative judgment.

It is worth noting that the research in Chapters 4, 5, and 6 proved fruitful in that it led to further research on whether behavioural interventions could be incorporated in a government-funded chatbot aimed at decreasing infectious disease transmission during the pandemic (van Baal, Le, Fatehi, Hohwy, et al., 2022; van Baal, Le, Fatehi, Verdejo-Garcia, et al., 2022).

7.5 Implications

The implications of my proposed conceptual framework of impulsivity, self-control, and time preferences is that we should bring nuance to our discussion of how each cognitive process relates to and influences the other. There also seems to be a tendency in the literature and in society more generally to regard patterns relating to the three constructs through a normative lens. Future orientation and self-control are seen as positive traits that we ought to cultivate. I argue that it is important we are aware that too much of both can have detrimental outcomes too. Impulsivity, in contrast, is often regarded as negative in the public eye and in the literature; it gets much attention for its association with psychopathology. This too is often misguided because people tend to become

impulsive when their environment rewards them for it, or conversely, does not reward them for deep contemplation and information seeking.

In any case, diversity in impulsivity and self-control is beneficial because a society likely needs people who take risks (a key component of impulsivity), such that some people become entrepreneurs, driving innovation. Likewise, high levels of self-control can lead to a host of problems such as rigidity, and a failure to enjoy oneself—it is often good to have people be more flexible.

Thus, one conclusion one may draw from this discussion is that how much impulsivity or self-control is good or bad depends on the situation. Aristotle's doctrine of the mean appears to be a good starting point for an average lifespan trait score, but arguably it is our adaptability (i.e., state changes, see Chapter 3 and Chapter 4) that matters. The adaptability of our impulsivity, self-control, and time preferences can not only ensure that we act rationally in changing environments and phenotypical states but also makes people receptive to good policy.

Theoretically, the conceptualisation of impulsivity and self-control I developed in this thesis, together with the experimental evidence suggests that measures of impulsivity and self-control might need to be re-evaluated, but they also cast new light on theories of impulsive behaviour and addiction. In starting the methodological re-evaluation effort, I have focused on time preferences, but the implications of my theory of impulsivity and self-control need not be limited to paradigms aimed at examining this topic.

Regarding the theories of impulsive behaviour and addiction, the findings in this thesis echo earlier calls that attempts to increase self-control have lower chances of success than attempts to lower the intensity of urges, especially in moments when participants are likely to be in an impulsive state. For many scenarios with smaller stakes, the best approach is likely to remove the tempting reinforcer (Duckworth et al., 2018), while for decisions with larger stakes where removing the tempting option is not possible, focusing on one's higher-order preferences might desensitise the urge to take the tempting option.

In summary, this thesis has provided new insights into the nature of impulsivity and self-control, and how we can intervene in these processes. The findings of this research have important implications for public health. Further research is needed to continue to

explore the complex relationship between these constructs and to determine how best to promote healthy behaviours in the public health context.

7.6 Future Directions

I believe the topic of adaptability in impulsivity and self-control to be one that will become increasingly important, as one of the major tasks for citizens and key decision-makers in the next few decades is responding well to changing environments. For example, people will need to adapt to a confusing and misleading information market (Pennycook & Rand, 2021), changing social norms on acceptable environmental behaviour (for a discussion, see Cialdini & Jacobson, 2021), and potential further pandemics (Harrington et al., 2021). Another frontier of particular contemporary interest is what happens to our impulsivity and self-control as the attention economy increasingly competes for our eyeballs and clicks.

One encouraging new development for psychological research on impulsivity and self-control is the development of gamified measures (e.g., Verdejo-Garcia et al., 2020). These give researchers more freedom in constructing and changing the environment and allow them to observe impulsivity and self-control in a more ecologically valid setting than, for example, preferential choice.

A historical focus on trait scales of impulsivity, self-control, and behavioural measures of time preferences has left us largely in the dark about absolute levels of, and individual differences in, the adaptability of these tendencies to new environments with different requirements (Fenneman & Frankenhuis, 2020). More research on this topic is needed.

In this thesis, I have presented a novel theoretical framework that spells out the differences and the relationships among impulsivity, self-control, and time preferences. The conceptual framework guided my approach and fundamental thinking in each of the empirical Chapters, where I carried out research investigating manifestations of impulsivity and self-control. I have also discussed two studies which extend my conceptualisation of impulsivity and self-control into the selfish versus prosocial domain, identifying that to contribute to the public good, people need to realise there may be a large discrepancy between their first-order preferences and what they ought to do. Finally, I have identified

implications as well as new questions for future research, pertaining to re-evaluations of existing theories and methodologies for studying impulsivity and self-control.

Bibliography

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

Ainslie, G. (2021). Willpower with and without effort. *Behavioral and Brain Sciences*, *44*.

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, *9*, 211–217. https://doi.org/10.2147/JMDH.S104807

Amasino, D. R., Sullivan, N. J., Kranton, R. E., & Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour*, *3*(4), 383–392. https://doi.org/10.1038/s41562-019-0537-2

Ameriks, K., & Clarke, D. M. (2000). *Aristotle: Nicomachean Ethics*. Cambridge University Press. Retrieved December 13, 2022, from https://books.google.co.uk/books?hl=en&lr=&id=p1NmFDSlMfgC&oi=fnd&pg=PR6&dq=nicomachean+ethics&ots=mXq23hDZbs&sig=t8AxhSqAzqHJI7QEuqSJ1HWg8fA&redir_esc=y#v=onepage&q=nicomachean%20ethics&f=false

Amlung, M., Vedelago, L., Acker, J., Balodis, I., & MacKillop, J. (2017). Steep delay discounting and addictive behavior: A meta-analysis of continuous associations. *Addiction*, *112*(1), 51–62. https://doi.org/10.1111/add.13535

Anderberg, R. H., Hansson, C., Fenander, M., Richard, J. E., Dickson, S. L., Nissbrandt, H., Bergquist, F., & Skibicka, K. P. (2016). The stomach-derived hormone ghrelin increases impulsive behavior. *Neuropsychopharmacology*, *41*(5), 1199–1209. https://doi.org/10.1038/npp.2015.297

Andreoni, J. (1988). Why free ride?: Strategies and learning in public goods experiments. *Journal of public Economics*, *37*(3), 291–304.

Archetti, M., & Scheuring, I. (2011). Coexistance of cooperation and defection in public goods games. *Evolution*, *65*(4), 1140–1148. https://doi.org/10.1111/j.1558-5646.2010.01185.x

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*(2), 87–98. https://doi.org/10.1002/bdm.501

Association, A. P. (2010). *Diagnostic and statistical manual of mental disorders, text revision (DSM-IV-TR®)*. American Psychiatric Publishing.

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders*.

Bakhshani, N.-M. (2014). Impulsivity: A Predisposition Toward Risky Behaviors. *International Journal of High Risk Behaviors & Addiction*, *3*(2), e20428. https://doi.org/10.5812/ijhrba.20428

Ballard, T., Luckman, A., & Konstantinidis, E. (2023). A systematic investigation into the reliability of inter-temporal choice model parameters. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-022-02241-7

Banerjee, R., Bhattacharya, J., & Majumdar, P. (2020). Exponential-growth prediction bias and compliance with safety measures related to COVID-19. *Social Science & Medicine*, *268*, 113473. https://doi.org/10.1016/j.socscimed.2020.113473

Barari, S., Caria, S., Davola, A., Falco, P., Fetzer, T., Fiorin, S., Hensel, L., Ivchenko, A., Jachimowicz, J., King, G., Kraft-Todd, G., Ledda, A., MacLennan, M., Mutoi, L., Pagani, C., Reutskaja, E., & Slepoi, F. R. (2020). Evaluating COVID-19 Public Health Messaging in Italy: Self-Reported Compliance and Growing Mental Health Concerns. *medRxiv*, 2020.03.27.20042820. https://doi.org/10.1101/2020.03.27.20042820

Barker, V., Romaniuk, L., Cardinal, R., Pope, M., Nicol, K., & Hall, J. (2015). Impulsivity in borderline personality disorder. *Psychological medicine*, *45*(9), 1955–1964. https://doi.org/10.1017/S0033291714003079

Baumann, A. A., & Odum, A. L. (2012). Impulsivity, risk taking, and timing. *Behavioural Processes*, *90*(3), 408–414. https://doi.org/10.1016/j.beproc.2012.04.005

Baumeister, R. F., & Alghamdi, N. G. (2015). Role of self-control failure in immoral and unethical actions. *Current Opinion in Psychology*, *6*, 66–69. https://doi.org/10.1016/j.copsyc.2015.04.001

Baumeister, R. F., & Vohs, K. D. (2003). Willpower, choice, and self-control.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current directions in psychological science*, *16*(6), 351–355. https://doi.org/10.1111%2Fj.1467-8721.2007.00534.x

Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., ... Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), 460–471. https://doi.org/10.1038/s41562-020-0884-z

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature neuroscience*, *8*(11), 1458–1463. https://doi.org/10.1038/nn1584

Becker, G. M., Degroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, *8*(1), 41–55. https://doi.org/10.1002/bs.3830080106

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232. https://doi.org/10.1002/bs.3830090304

Bembenutty, H. (2011). Academic delay of gratification and academic achievement. *Ed. Bembunutty, H. Self Regulated Learning: New Direction for Teaching and Learning. San Francisco. Wiley Periodicals, İnc.*

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *Journal of Neuroscience*, *31*(18), 6771–6779.

Berg, J. M., Latzman, R. D., Bliwise, N. G., & Lilienfeld, S. O. (2015). Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the UPPS for psychopathology. *Psychological assessment*, *27*(4), 1129. https://doi.org/10.1037/pas0000111

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Bergquist, M. (2020). Most People Think They Are More Pro-Environmental than Others: A Demonstration of the Better-than-Average Effect in Perceived Pro-Environmental Behavioral Engagement. *Basic and Applied Social Psychology*, *42*(1), 50–61. https://doi.org/10.1080/01973533.2019.1689364

Berry, G., Parsons, A., Morgan, M., Rickert, J., & Cho, H. (2022). A review of methods to reduce the probability of the airborne spread of COVID-19 in ventilation systems and enclosed spaces. *Environmental Research*, *203*, 111765. https://doi.org/10.1016/j.envres.2021.111765

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bickel, W. K., & Johnson, M. W. (2003). Delay discounting: A fundamental behavioral process of drug dependence. In *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 419–440). Russell Sage Foundation.

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134

Bran, A., & Vaidis, D. C. (2020). Assessing risk-taking: What to measure and how to measure it. *Journal of Risk Research*, *23*(4), 490–503. https://doi.org/10.1080/13669877.2019.1591489

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, *9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *Journal of consulting and clinical psychology*, *80*(1), 128. https://doi.org/10.1037/a0026582

Bulley, A., Lempert, K., Conwell, C., Irish, M., & Schacter, D. L. (2021). Intertemporal choice reflects value comparison rather than self-control: Insights from confidence judgments. https://doi.org/10.31234/osf.io/w5zuk

Bulley, A., & Schacter, D. L. (2020). Deliberating trade-offs with the future. *Nature Human Behaviour*, *4*(3), 238–247. https://doi.org/10.1038/s41562-020-0834-9

Camerer, C., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica: Journal of the Econometric Society*, 1–36. https://doi.org/10.2307/1911840

Cameron, J. D., Goldfield, G. S., Finlayson, G., Blundell, J. E., & Doucet, É. (2014). Fasting for 24 hours heightens reward from food and food-related cues. *PloS one*, *9*(1), e85970. https://doi.org/10.1371/journal.pone.0085970

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of personality and social psychology*, *67*(2), 319. https://doi.org/10.1037/0022-3514.67.2.319

Caswell, A. J., Bond, R., Duka, T., & Morgan, M. J. (2015). Further evidence of the heterogeneous nature of impulsivity. *Personality and Individual Differences*, *76*, 68–74. https://doi.org/10.1016/j.paid.2014.11.059

Chabris, C. F., Laibson, D. I., & Schuldt, J. P. (2010). Intertemporal choice. In *Behavioural and Experimental Economics* (pp. 168–177). Springer.

Chapple, C. L., & Johnson, K. A. (2007). Gender Differences in Impulsivity. https://doi.org/10.1177/1541204007301286

Chater, N., & Loewenstein, G. (2022). The i-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray. https://doi.org/10.2139/ssrn.4046264

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Cherek, D. R., Moeller, F. G., Dougherty, D. M., & Rhoades, H. (1997). Studies of violent and nonviolent male parolees: II. Laboratory and psychometric measurements of impulsivity. *Biological Psychiatry*, *41*(5), 523–529. https://doi.org/10.1016/S0006-3223(96)00426-X

Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the R package ordinal. *Submitted in J. Stat. Software*.

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement*, *56*(10), 15–23. https://doi.org/10.1002/pfi.21727

Cialdini, R. B., & Jacobson, R. P. (2021). Influences of social norms on climate change-related behaviors. *Current Opinion in Behavioral Sciences*, *42*, 1–8. https://doi.org/10.1016/j.cobeha.2021.01.005

Coffey, S. F., Gudleski, G. D., Saladin, M. E., & Brady, K. T. (2003). Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals. *Experimental and clinical psychopharmacology*, *11*(1), 18. https://doi.org/10.1037/1064-1297.11.1.18

Cohen, J., Ericson, K. M., Laibson, D., & White, J. M. (2020). Measuring Time Preferences. *Journal of Economic Literature*, *58*(2), 299–347. https://doi.org/10.1257/jel.20191074

Constandt, B., Thibaut, E., De Bosscher, V., Scheerder, J., Ricour, M., & Willem, A. (2020). Exercising in times of lockdown: An analysis of the impact of COVID-19 on levels and patterns of exercise among adults in Belgium. *International journal of environmental research and public health*, *17*(11), 4144. https://doi.org/10.3390/ijerph17114144

Coskunpinar, A., & Cyders, M. A. (2013). Impulsivity and substance-related attentional bias: A meta-analytic review. *Drug and alcohol dependence*, *133*(1), 1–14. https://doi.org/10.1016/j.drugalcdep.2013.05.008

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. *Journal of clinical epidemiology*, *43*(1), 87–91. https://doi.org/10.1016/0895-4356(90)90060-3

Cubitt, R., McDonald, R., & Read, D. (2018). Time Matters Less When Outcomes Differ: Unimodal vs. Cross-Modal Comparisons in Intertemporal Choice. *Management Science*, *64*(2), 873–887. https://doi.org/10.1287/mnsc.2016.2613

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical psychology review*, *31*(6), 965–982. https://doi.org/10.1016/j.cpr.2011.06.001

Dahl, C. J., Lutz, A., & Davidson, R. J. (2016). Cognitive processes are central in compassion meditation. *Trends in Cognitive Sciences*, *20*(3), 161–162. https://doi.org/10.1016/j.tics.2015.12.005

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, *143*(4), 1489. https://doi.org/10.1037/a0035976

Dawes, R. M. (1980). Social dilemmas. *Annual review of psychology*, *31*(1), 169–193. https://doi.org/10.1146/annurev.ps.31.020180.001125

De Wit, H. (2009). Impulsivity as a determinant and consequence of drug use: A review of underlying processes. *Addiction biology*, *14*(1), 22–31. https://doi.org/10.1111/j.1369-1600.2008.00129.x

De Wit, H., Flory, J. D., Acheson, A., McCloskey, M., & Manuck, S. B. (2007). IQ and nonplanning impulsivity are independently associated with delay discounting in middle-aged adults. *Personality and Individual Differences*, *42*(1), 111–121. https://doi.org/10.1016/j.paid.2006.06.026

DeShong, H. L., Mullins-Sweatt, S. N., Miller, J. D., Widiger, T. A., & Lynam, D. R. (2016). Development of a Short Form of the Five-Factor Borderline Inventory. *Assessment*, *23*(3), 342–352. https://doi.org/10.1177/1073191115581475

DeSteno, D., Li, Y., Dickens, L., & Lerner, J. S. (2014). Gratitude: A Tool for Reducing Economic Impatience. *Psychological Science*, *25*(6), 1262–1267. https://doi.org/10.1177/0956797614529979

Dick, D. M., Smith, G., Olausson, P., Mitchell, S. H., Leeman, R. F., O'Malley, S. S., & Sher, K. (2010). REVIEW: Understanding the construct of impulsivity and its relationship to alcohol use disorders. *Addiction Biology*, *15*(2), 217–226. https://doi.org/10.1111/j.1369-1600.2009.00190.x

Donnarumma, F., & Pezzulo, G. (2021). Moral decisions in the age of COVID-19: Your choices really matter. *Social Sciences & Humanities Open*, *4*(1), 100149. https://doi.org/10.1016/j.ssaho.2021.100149

Doran, N., Spring, B., & McChargue, D. (2007). Effect of impulsivity on craving and behavioral reactivity to smoking cues. *Psychopharmacology*, *194*(2), 279–288. https://doi.org/10.1007/s00213-007-0832-x

Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J. E., & Treibich, T. (2020). Rational Heuristics? Expectations and Behaviors in Evolving Economies with Heterogeneous Interacting Agents. *Economic Inquiry*, *58*(3), 1487–1516. https://doi.org/10.1111/ecin.12897

Doyle, J. R. (2012). Survey of time preference, delay discounting models. *Delay Discounting Models (April 20, 2012)*.

Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., Spiegelhalter, D., & van der Linden, S. (2020). Risk perceptions of COVID-19 around the world. *Journal of Risk Research*, 1–13. https://doi.org/10.1080/13669877.2020.1758193

Duckworth, A. L., Milkman, K. L., & Laibson, D. (2018). Beyond Willpower: Strategies for Reducing Failures of Self-Control. *Psychological Science in the Public Interest*, *19*(3), 102–129. https://doi.org/10.1177/1529100618821893

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *Journal of abnormal psychology*, *118*(1), 195. https://doi.org/10.1037/a0014868

Eiselt, A.-K., Chen, S., Chen, J., Arnold, J., Kim, T., Pachitariu, M., & Sternson, S. M. (2021). Hunger or thirst state uncertainty is resolved by outcome evaluation in medial prefrontal cortex to guide decision-making. *Nature Neuroscience*, 1–6. https://doi.org/10.1038/s41593-021-00850-4

El Zein, M., Bahrami, B., & Hertwig, R. (2019). Shared responsibility in collective decisions. *Nature Human Behaviour*, *3*(6), 554–559. https://doi.org/10.1038/s41562-019-0596-4

Engle-Warnick, J., & Slonim, R. L. (2004). The evolution of strategies in a repeated trust game. *Journal of Economic Behavior & Organization*, *55*(4), 553–573. https://doi.org/10.1016/j.jebo.2003.11.008

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472–5477. https://doi.org/10.1073/pnas.1818430116

Epley, N., & Dunning, D. (2000). Feeling" holier than thou": Are self-serving assessments produced by errors in self-or social prediction? *Journal of personality and social psychology*, *79*(6), 861. https://doi.org/10.1037/0022-3514.79.6.861

Ericson, K., White, J. M., Laibson, D., & Cohen, J. (2015). *Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better than Delay Discounting*

(tech. rep. w20948). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w20948

Escrivá-Martínez, T., Herrero, R., Molinari, G., Rodríguez-Arias, M., Verdejo-García, A., & Baños, R. M. (2020). Binge Eating and Binge Drinking: A Two-Way Road? An Integrative Review. *Current pharmaceutical design*, *26*(20), 2402–2415. https://doi.org/10.2174/1381612826666200316153317

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*(4), 348–361. https://doi.org/10.1007/pl00005481

Everett, J. A. C., Colombatto, C., Chituc, V., Brady, W. J., & Crockett, M. (2020). The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic. https://doi.org/10.31234/osf.io/9yqs8

Farah, M. J., & Hook, C. J. (2017). Trust and the poverty trap. *Proceedings of the National Academy of Sciences*, *114*(21), 5327–5329. https://doi.org/10.1073/pnas.1704798114

Farmer, J. D., & Geanakoplos, J. (2009). Hyperbolic discounting is rational: Valuing the far future with uncertain discount rates.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fazio, R. H., Ruisch, B. C., Moore, C. A., Samayoa, J. A. G., Boggs, S. T., & Ladanyi, J. T. (2021). Social distancing decreases an individual's likelihood of contracting COVID-19. *Proceedings of the National Academy of Sciences*, *118*(8). https://doi.org/10.1073/pnas.2023131118

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in cognitive sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Gintis, H. (2007). Human Motivation and Social Cooperation: Experimental and Analytical Foundations. *Annual Review of Sociology*, *33*(1), 43–64. https://doi.org/10.1146/annurev.soc.33.040406.131812

Feltenstein, M., & See, R. (2008). The neurocircuitry of addiction: An overview. *British journal of pharmacology*, *154*(2), 261–274. https://doi.org/10.1038%2Fbjp.2008.51

Fenneman, J., & Frankenhuis, W. E. (2020). Is impulsive behavior adaptive in harsh and unpredictable environments? A formal model. *Evolution and Human Behavior*, *41*(4), 261–273. https://doi.org/10.1016/j.evolhumbehav.2020.02.005

Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: A review of its development, causes, and consequences. *Drug and Alcohol Dependence*, *97*(1), 1–20. https://doi.org/10.1016/j.drugalcdep.2008.03.030

Finne, E., Englert, C., & Jekauc, D. (2019). On the importance of self-control strength for regular physical activity. *Psychology of Sport and Exercise*, *43*, 165–171. https://doi.org/10.1016/j.psychsport.2019.02.007

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fiske, D. W. (1971). Measuring the concepts of personality.

Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, *107*(12), 5334–5338. https://doi.org/10.1073/pnas.0913149107

Frankfurt, H. G. (1988). Freedom of the Will and the Concept of a Person. In M. F. Goodman (Ed.), *What Is a Person?* (pp. 127–144). Humana Press. https://doi.org/10.1007/978-1-4612-3950-5_6

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, *40*(2), 351–401.

Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *American Economic Review*, *94*(5), 1717–1722. https://doi.org/10.1257/0002828043052187

Frias-Navarro, D., Pascual-Soler, M., Berrios-Riquelme, J., Gomez-Frias, R., & Caamaño-Rocha, L. (2021). COVID–19. Effect of Moral Messages to Persuade the Population to Stay at Home in Spain, Chile, and Colombia. *The Spanish journal of psychology*, *24*.

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*, *23*(2), 107–131. https://doi.org/10.1177%2F1088868318762183

Gabaix, X., & Laibson, D. (2017). *Myopia and discounting* (Report). National bureau of economic research.

Gigerenzer, G. (2016). Towards a Rational Theory of Heuristics. In R. Frantz & L. Marsh (Eds.), *Minds, Models and Milieux: Commemorating the Centennial of the Birth of Herbert Simon* (pp. 34–59). Palgrave Macmillan UK. https://doi.org/10.1057/9781137442505_3

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational behavior and human decision processes*, *88*(1), 430–444. https://doi.org/10.1006/obhd.2001.2982

Gloster, A. T., Walder, N., Levin, M. E., Twohig, M. P., & Karekla, M. (2020). The empirical status of acceptance and commitment therapy: A review of meta-analyses. *Journal of Contextual Behavioral Science*, *18*, 181–192. https://doi.org/10.1016/j.jcbs.2020.09.009

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological bulletin*, *136*(3), 351. https://doi.org/10.1037/a0018807

Goldstein, R. Z., & Volkow, N. D. (2002). Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. *American Journal of Psychiatry*, *159*(10), 1642–1652. https://doi.org/10.1176/appi.ajp.159.10.1642

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature reviews neuroscience*, *12*(11), 652–669. https://doi.org/10.1038/nrn3119

Gollwitzer, A., Martel, C., Marshall, J., Höhs, J. M., & Bargh, J. A. (2020). Connecting self-reported social distancing to real-world behavior at the individual and us state level. https://doi.org/10.31234/osf.io/kvnwp

Gollwitzer, A., McLoughlin, K., Martel, C., Marshall, J., Höhs, J. M., & Bargh, J. A. (2021). Linking Self-Reported Social Distancing to Real-World Behavior During the COVID-19 Pandemic. *Social Psychological and Personality Science*, *0*(0), 19485506211018132. https://doi.org/10.1177/19485506211018132

Gough, H. G. (2000). The California psychological inventory. *Testing and assessment in counseling practice*, *45*, 71.

Green, L., Fristoe, N., & Myerson, J. (1994). Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin & Review*, *1*(3), 383–389. https://doi.org/10.3758/BF03213979

Green, L., & Myerson, J. (2013). How many impulsivities? A discounting perspective. *Journal of the experimental analysis of behavior*, *99*(1), 3–13. https://doi.org/10.1002/jeab.1

Gross, J., & De Dreu, C. K. (2019). Individual solutions to shared problems create a modern tragedy of the commons. *Science advances*, *5*(4), eaau7296. https://doi.org/10.1126/sciadv.aau7296

Gul, F., & Pesendorfer, W. (2001). Temptation and Self-Control. *Econometrica*, *69*(6), 1403–1435. Retrieved June 30, 2022, from https://www.jstor.org/stable/2692262

Gullo, M. J., & Dawe, S. (2008). Impulsivity and adolescent substance use: Rashly dismissed as "all-bad"? *Neuroscience and Biobehavioral Reviews*, *32*(8), 1507–1518. https://doi.org/10.1016/j.neubiorev.2008.06.003

Hamann, H. (2022). On getting it right by being wrong: A case study of how flawed research may become self-fulfilling at last. *Proceedings of the National Academy of Sciences*, *119*(15), e2122274119. https://doi.org/10.1073/pnas.2122274119

Hamilton, K. R., Mitchell, M. R., Wing, V. C., Balodis, I. M., Bickel, W. K., Fillmore, M., Lane, S. D., Lejuez, C. W., Littlefield, A. K., Luijten, M., Mathias, C. W., Mitchell, S. H., Napier, T. C., Reynolds, B., Schütz, C. G., Setlow, B., Sher, K. J., Swann, A. C., Tedford, S. E., … Moeller, F. G. (2015). Choice impulsivity: Definitions, measurement issues, and clinical implications. *Personality Disorders: Theory, Research, and Treatment*, *6*(2), 182–198. https://doi.org/10.1037/per0000099

Hanegraaf, L., van Baal, S. T., Hohwy, J., & Verdejo-Garcia, A. (2021). A systematic review and meta-analysis of 'Systems for Social Processes' in borderline personality and substance use disorders. *Neuroscience & Biobehavioral Reviews*, *127*, 572–592. https://doi.org/10.1016/j.neubiorev.2021.04.013

Harrington, W. N., Kackos, C. M., & Webby, R. J. (2021). The evolution and future of influenza pandemic preparedness. *Experimental & Molecular Medicine*, *53*(5), 737–749. https://doi.org/10.1038/s12276-021-00603-0

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, *27*(1), 1–17. https://doi.org/10.1080/1047840X.2016.1082418

Hawker, C. O., Merkouris, S. S., Youssef, G. J., & Dowling, N. A. (2021). A smartphone-delivered ecological momentary intervention for problem gambling (GamblingLess: Curb Your Urge): Single-arm acceptability and feasibility trial. *Journal of Medical Internet Research*, *23*(3), e25786. https://doi.org/10.2196/25786

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006

He, L., Golman, R., & Bhatia, S. (2019). Variable time preference. *Cognitive psychology*, *111*, 53–79. https://doi.org/10.1016/j.cogpsych.2019.03.003

He, L., Zhao, W. J., & Bhatia, S. (2022). An ontology of decision models. *Psychological Review*, *129*(1), 49–72. https://doi.org/10.1037/rev0000231

Head, B. W. (2011). Australian experience: Civic engagement as symbol and substance. *Public Administration and Development*, *31*(2), 102–112. https://doi.org/10.1002/pad.599

Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What people desire, feel conflicted about, and try to resist in everyday life. *Psychological science*, *23*(6), 582–588. https://doi.org/10.1177%2F0956797612437426

Holton, R. (2009). *Willing, wanting, waiting*. Oxford University Press on Demand.

Hursh, S. R., & Schwartz, L. P. (2022). A general model of demand and discounting. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*. https://doi.org/10.1037/adb0000848

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in cognitive sciences*, *18*(3), 127–133. https://doi.org/10.1016/j.tics.2013.12.009

Ioannidis, K., Hook, R., Wickham, K., Grant, J. E., & Chamberlain, S. R. (2019). Impulsivity in gambling disorder and problem gambling: A meta-analysis. *Neuropsychopharmacology*, *44*(8), 1354–1361. https://doi.org/10.1038/s41386-019-0393-9

Isaac, R. M., & Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry*, *26*(4), 585–608. https://doi.org/10.1111/j.1465-7295.1988.tb01519.x

Jachimowicz, J. M., Chafik, S., Munrat, S., Prabhu, J. C., & Weber, E. U. (2017). Community trust reduces myopic decisions of low-income individuals. *Proceedings of the National Academy of Sciences*, *114*(21), 5401–5406. https://doi.org/10.1073/pnas.1617395114

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*(2), 159–186. https://doi.org/10.1017/bpp.2018.43

Jordan, J. J., Yoeli, E., & Rand, D. G. (2021). Don't get it or don't spread it: Comparing self-interested versus prosocial motivations for COVID-19 prevention behaviors. *Scientific Reports*, *11*(1), 20222. https://doi.org/10.1038/s41598-021-97617-5

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kakoschke, N., Hawker, C., Castine, B., de Courten, B., & Verdejo-Garcia, A. (2018). Smartphone-based cognitive bias modification training improves healthy food choice in obesity: A pilot study. *European Eating Disorders Review*, *26*(5), 526–532. https://doi.org/10.1002/erv.2622

Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler Publishing Company.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American psychologist*, *43*(1), 23. https://doi.org/10.1037/0003-066X.43.1.23

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of personality and social psychology*, *45*(4), 819. https://doi.org/10.1037/0022-3514.45.4.819

Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, *102*(1), 23–39. https://doi.org/10.1111/1467-9442.00182

Kiely, K. M., Brady, B., & Byles, J. (2019). Gender, mental health and ageing. *Maturitas*, *129*, 76–84. https://doi.org/10.1016/j.maturitas.2019.09.004

Kirby, K. N., & Herrnstein, R. (1995). Preference Reversals Due to Myopic Discounting of Delayed Reward. *Psychological Science*, *6*(2), 83–89. https://doi.org/10.1111/j.1467-9280.1995.tb00311.x

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental psychology: general*, *128*(1), 78. https://doi.org/10.1037/0096-3445.128.1.78

Kivetz, R., & Keinan, A. (2006). Repenting Hyperopia: An Analysis of Self-Control Regrets. *Journal of Consumer Research*, *33*(2), 273–282. https://doi.org/10.1086/506308

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economics letters*, *101*(3), 175–178.

Konstantinidis, E., van Ravenzwaaij, D., Güney, Ş., & Newell, B. R. (2020). Now for sure or later with a risk? Modeling risky intertemporal choice as accumulated preference. *Decision*, *7*(2), 91. https://doi.org/10.1037/dec0000103

Kooistra, E. B., & van Rooij, B. (2020). Pandemic Compliance: A systematic review of influences on social distancing behaviour during the first wave of the COVID-19 outbreak. *Available at SSRN 3738047*.

Koomen, R., Grueneisen, S., & Herrmann, E. (2020). Children delay gratification for cooperative ends. *Psychological science*, *31*(2), 139–148. https://doi.org/10.1177%2F0956797619894205

Koval, P., Hinton, J., Dozo, N., Gleeson, J., Alvarez, M., Harrison, A., Vu, D., Susanto, R., Jayaputera, G., & Sinnott, R. (2019). SEMA3: Smartphone Ecological Momentary Assessment, Version 3. http://www.sema3.com

Kraut, R. (2001). Aristotle's ethics.

Kugler, T., Rapoport, A., & Pazy, A. (2010). Public good provision in inter-team conflicts: Effects of asymmetry and profit-sharing rule. *Journal of Behavioral Decision Making*, *23*(4), 421–438. https://doi.org/10.1002/bdm.667

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478. https://doi.org/10.1162/003355397555253

Lammers, J., Crusius, J., & Gast, A. (2020). Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proceedings of the National Academy of Sciences*, *117*(28), 16264–16266. https://doi.org/10.1073/pnas.2006048117

Lane, S. D., Cherek, D. R., Rhodes, H. M., Pietras, C. J., & Tcheremissine, O. V. (2003). Relationships among laboratory and psychometric measures of impulsivity: Implications in substance abuse and dependence. *Addictive Disorders & Their Treatment*, *2*(2), 33–40. https://doi.org/10.1097/00132576-200302020-00001

Larsen, M., Nyrup, J., & Petersen, M. B. (2020). Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias? *Journal of Behavioral Public Administration*, *3*(2). https://doi.org/10.30636/jbpa.32.164

Latkin, C. A., Dayton, L., Yi, G., Colon, B., & Kong, X. (2021). Mask usage, social distancing, racial, and gender correlates of COVID-19 vaccine intentions among adults in the US. *PloS one*, *16*(2), e0246970. https://doi.org/10.1371/journal.pone.0246970

Laude, J. R., Pattison, K. F., & Zentall, T. R. (2012). Hungry pigeons make suboptimal choices, less hungry pigeons do not. *Psychonomic bulletin & review*, *19*(5), 884–891. https://doi.org/10.3758/s13423-012-0282-2

Lawrence, K., Allen, J., & Chanen, A. (2010). Impulsivity in Borderline Personality Disorder: Reward-Based Decision-Making and its Relationship to Emotional Distress. *Journal of personality disorders*, *24*, 786–99. https://doi.org/10.1521/pedi.2010.24.6.785

Lempert, K. M., & Phelps, E. A. (2016). The Malleability of Intertemporal Choice. *Trends in Cognitive Sciences*, *20*(1), 64–74. https://doi.org/10.1016/j.tics.2015.09.005

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Leviston, Z., & Uren, H. V. (2020). Overestimating One's "Green" Behavior: Better-Than-Average Bias May Function to Reduce Perceived Personal Threat from Climate Change. *Journal of Social Issues*, *76*(1), 70–85. https://doi.org/10.1111/josi.12365

Levy, M. R., & Tasoff, J. (2017). Exponential-growth bias and overconfidence. *Journal of Economic Psychology*, *58*, 1–14. https://doi.org/10.1016/j.joep.2016.11.001

Li, E., Hing, N., Russell, A. M., & Vitartas, P. (2020). Impulsive Sports Betting: The Effects of Food or Substance Consumption. *Journal of gambling studies*, *36*(2), 539–554. https://doi.org/10.1007/s10899-020-09938-1

Loewenstein, G. (1988). Frames of Mind in Intertemporal Choice. *Management Science*, *34*(2), 200–214. https://doi.org/10.1287/mnsc.34.2.200

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational behavior and human decision processes*, *65*(3), 272–292. https://doi.org/10.1006/obhd.1996.0028

Loewenstein, G. (2018). Self-control and its discontents: A commentary on Duckworth, Milkman, and Laibson. *Psychological Science in the Public Interest*, *19*(3), 95–101.

Loewenstein, G., O'Donoghue, T., & Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, *2*(2), 55–81. https://doi.org/10.1037/dec0000029

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, *107*(2), 573–597. https://doi.org/10.2307/2118482

Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal choice. *Journal of Economic perspectives*, *3*(4), 181–193. https://doi.org/10.1257/jep.3.4.181

Logue, A. W. (1988). Research on self-control: An integrating framework. *Behavioral and Brain Sciences*, *11*(4), 665–679. https://doi.org/10.1017/S0140525X00053978

Lubman, D. I., Allen, N. B., Peters, L. A., & Deakin, J. W. (2008). Electrophysiological evidence that drug cues have greater salience than other affective stimuli in opiate addiction. *Journal of psychopharmacology*, *22*(8), 836–842. https://doi.org/10.1177%2F0269881107083846

Lunn, P. D., Belton, C. A., Lavin, C., McGowan, F. P., Timmons, S., & Robertson, D. A. (2020). Using Behavioral Science to help fight the Coronavirus. *Journal of Behavioral Public Administration*, *3*(1). https://doi.org/10.30636/jbpa.31.147

Lunn, P. D., Timmons, S., Belton, C. A., Barjaková, M., Julienne, H., & Lavin, C. (2020). Motivating social distancing during the Covid-19 pandemic: An online experiment. *Social Science & Medicine*, *265*, 113478. https://doi.org/10.1016/j.socscimed.2020.113478

Madden, G. J., & Bickel, W. K. (2010). *Impulsivity: The behavioral and neurological science of discounting*. American Psychological Association.

Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control patients: Drug and monetary rewards. *Experimental and clinical psychopharmacology*, *5*(3), 256. https://doi.org/10.1037/1064-1297.5.3.256

Madden, G. J., Raiff, B. R., Lagorio, C. H., Begotka, A. M., Mueller, A. M., Hehli, D. J., & Wegener, A. A. (2004). Delay discounting of potentially real and hypothetical rewards: II. Between-and within-subject comparisons. *Experimental and clinical psychopharmacology*, *12*(4), 251. https://doi.org/10.1037/1064-1297.12.4.251

Magen, E., Dweck, C. S., & Gross, J. J. (2008). The hidden-zero effect: Representing a single choice as an extended sequence reduces impulsive choice. *Psychol Sci*, *19*(7), 648–9. https://doi.org/10.1111/j.1467-9280.2008.02137.x

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.

Manuguerra, M., & Heller, G. Z. (2010). Ordinal regression models for continuous scales. *The International Journal of Biostatistics*, *6*(1). https://doi.org/10.2202/1557-4679.1230

Manuguerra, M., Heller, G. Z., & Ma, J. (2020). Continuous ordinal regression for analysis of visual analogue scales: The R Package ordinalCont. *Journal of Statistical Software*, *96*(1), 1–25. https://doi.org/10.18637/jss.v096.i08

Massey, A., & Hill, A. J. (2012). Dieting and food craving. A descriptive, quasi-prospective study. *Appetite*, *58*(3), 781–785. https://doi.org/10.1016/j.appet.2012.01.020

Matejka, J., Glueck, M., Grossman, T., & Fitzmaurice, G. (2016). *The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales*. https://doi.org/10.1145/2858036.2858063

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Quantitative analyses of behavior*, *5*, 55–73.

McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, *3*, 147–151. https://doi.org/10.1016/j.cobeha.2015.04.006

McLean, D. C., Nakamura, J., & Csikszentmihalyi, M. (2017). Explaining System Missing: Missing Data and Experience Sampling Method. *Social Psychological and Personality Science*, *8*(4), 434–441. https://doi.org/10.1177/1948550617708015

Mele, A. R. (2010). Weakness of will and akrasia. *Philosophical Studies*, *150*(3), 391–404. https://doi.org/10.1007/s11098-009-9418-2

Mele, A. R. (2012). *Backsliding: Understanding weakness of will*. Oxford University Press.

Melnyk, V., van Herpen, E., Fischer, A. R., & van Trijp, H. C. (2013). Regulatory fit effects for injunctive versus descriptive social norms: Evidence from the promotion of sustainable products. *Marketing Letters*, *24*(2), 191–203. https://doi.org/10.1007/s11002-013-9234-5

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological review*, *106*(1), 3. https://doi.org/10.1037/0033-295x.106.1.3

Meyer, G. (2016). In science communication, why does the idea of a public deficit always return? *Public Understanding of Science*, *25*(4), 433–446. https://doi.org/10.1177/0963662516629750

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological bulletin*, *130*(5), 711. https://doi.org/10.1037/0033-2909.130.5.711

Michael, J., & van Baal, S. T. (2021). Key themes in recent research on interpersonal functioning in borderline personality disorder. *Current Opinion in Psychiatry*, *34*(1), 44–47. https://doi.org/10.1097/YCO.0000000000000649

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A., & Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences*, *105*(7), 2291–2294. https://doi.org/10.1073/pnas.0709546105

Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making - Katherine L. Milkman, Todd Rogers, Max H. Bazerman, 2008. *Perspectives on Psychological Science*, *3*(4). https://doi.org/10.1111/j.1745-6924.2008.0008

Milyavskaya, M., Inzlicht, M., Hope, N., & Koestner, R. (2015). Saying "no" to temptation: Want-to motivation improves self-regulation by reducing temptation rather than by increasing self-control. *Journal of Personality and Social Psychology*, *109*(4), 677. https://doi.org/10.1037/pspp0000045

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, *16*(2), 329. https://doi.org/10.1037/h0029815

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, *244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Mitchell, S. H. (1999). Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology*, *146*(4), 455–464. https://doi.org/10.1007/PL00005491

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American journal of psychiatry*, *158*(11), 1783–1793. https://doi.org/10.1176/appi.ajp.158.11.1783

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 653–662. https://doi.org/10.1098/rspb.2011.1172

Morgan, T. A., & Aljabari, R. (2019). Using Acceptance and Commitment Therapy in the Treatment of Borderline Personality Disorder: Rationale, Preliminary Evidence, and Future Directions. *Current Treatment Options in Psychiatry*, *6*(4), 271–283. https://doi.org/10.1007/s40501-019-00185-9

Morina, N., Kip, A., Hoppen, T. H., Priebe, S., & Meyer, T. (2021). Potential impact of physical distancing on physical and mental health: A rapid narrative umbrella

review of meta-analyses on the link between social connection and health - PMC. *BMJ Open*, *11*(3). https://doi.org/10.1136/bmjopen-2020-042335

Moss, M. (2013). *Salt, sugar, fat: How the food giants hooked us*. Random House.

Muela, I., Navas, J. F., Ventura-Lucena, J. M., & Perales, J. C. (2022). How to pin a compulsive behavior down: A systematic review and conceptual synthesis of compulsivity-sensitive items in measures of behavioral addiction. *Addictive Behaviors*, *134*, 107410. https://doi.org/10.1016/j.addbeh.2022.107410

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, *76*(2), 235–243. Retrieved June 28, 2022, from https://onlinelibrary.wiley.com/doi/abs/10.1901/jeab.2001.76-235

Niedzwiedz, C. L., Green, M. J., Benzeval, M., Campbell, D., Craig, P., Demou, E., Leyland, A., Pearce, A., Thomson, R., & Whitley, E. (2021). Mental health and health behaviours before and during the initial phase of the COVID-19 lockdown: Longitudinal analyses of the UK Household Longitudinal Study. *J Epidemiol Community Health*, *75*(3), 224–231. https://doi.org/10.1136/jech-2020-215060

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of child psychology and psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

O'Donoghue, T., & Rabin, M. (1998). Procrastination in preparing for retirement. *University of California-Berkeley Working Paper*.

O'Donoghue, T., & Rabin, M. (2015). Present bias: Lessons learned and to be learned. *American Economic Review*, *105*(5), 273–79. https://doi.org/10.1257/aer.p20151085

Odum, A. L. (2011). Delay Discounting: I'm a k, You're a k. *Journal of the Experimental Analysis of Behavior*, *96*(3), 427–439. https://doi.org/10.1901/jeab.2011.96-423

Odum, A. L., & Baumann, A. A. (2010). Delay discounting: State and trait variable. In *Impulsivity: The behavioral and neurological science of discounting.* (pp. 39–65). American Psychological Association.

Omoto, A. M., Malsch, A. M., & Barraza, J. A. (2009). Compassionate acts: Motivations for and correlates of volunteerism among older adults. *The science of compassionate love: Theory, research, and applications*, 257–282.

Osmundsen, M., & Petersen, M. B. (2020). Framing Political Risks: Individual Differences and Loss Aversion in Personal and Political Situations. *Political Psychology*, *41*(1), 53–70. https://doi.org/10.1111/pops.12587

Papachristou, H., Nederkoorn, C., Havermans, R., van der Horst, M., & Jansen, A. (2012). Can't stop the craving: The effect of impulsivity on cue-elicited craving for alcohol in heavy and light social drinkers. *Psychopharmacology*, *219*(2), 511–518. https://doi.org/10.1007/s00213-011-2240-5

Parker, J. D., & Bagby, R. (1997). Impulsivity in adults: A critical review. *Impulsivity: Theory, assessment, and treatment*, 142–157.

Patock-Peckham, J. A., Cheong, J., Balhorn, M. E., & Nagoshi, C. T. (2001). A social learning perspective: A model of parenting styles, self-regulation, perceived drinking control, and alcohol use and problems. *Alcoholism: clinical and experimental research*, *25*(9), 1284–1292. https://doi.org/10.1111/j.1530-0277.2001.tb02349.x

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6⟨768::AID-JCLP2270510607⟩3.0.CO;2-1

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, *66*(1), 138–148. https://doi.org/10.1016/j.neuron.2010.03.026

Pieh, C., Budimir, S., Delgadillo, J., Barkham, M., Fontaine, J. R., & Probst, T. (2021). Mental health during COVID-19 lockdown in the United Kingdom. *Psychosomatic medicine*, *83*(4), 328–337. https://doi.org/10.1097/PSY.0000000000000871

Potenza, M. N., Steinberg, M. A., Skudlarski, P., Fulbright, R. K., Lacadie, C. M., Wilber, M. K., Rounsaville, B. J., Gore, J. C., & Wexler, B. E. (2003). Gambling urges in pathological gambling: A functional magnetic resonance imaging study. *Archives of general psychiatry*, *60*(8), 828–836. https://doi.org/10.1001/archpsyc.60.8.828

Prieto Curiel, R., & González Ramírez, H. (2021). Vaccination strategies against COVID-19 and the diffusion of anti-vaccination views. *Scientific Reports*, *11*(1), 6626. https://doi.org/10.1038/s41598-021-85555-1

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the experimental analysis of behavior*, *17*(1), 15–22. https://doi.org/10.1901/jeab.1972.17-15

Rawn, C. D., & Vohs, K. D. (2006). The Importance of Self-Regulation for Interpersonal Functioning. In *Self and relationships: Connecting intrapersonal and interpersonal processes.* (pp. 15–31). The Guilford Press.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of risk and uncertainty*, *23*(1), 5–32.

Read, D. (2004). Intertemporal choice. *Blackwell handbook of judgment and decision making*, 424–443.

Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four Score and Seven Years from Now: The Date/Delay Effect in Temporal Discounting. *Management Science*, *51*(9), 1326–1335. https://doi.org/10.1287/mnsc.1050.0412

Reddy, S. D., Negi, L. T., Dodson-Lavelle, B., Ozawa-de Silva, B., Pace, T. W., Cole, S. P., Raison, C. L., & Craighead, L. W. (2013). Cognitive-based compassion training: A promising prevention strategy for at-risk adolescents. *Journal of Child and Family Studies*, *22*(2), 219–230. https://doi.org/10.1007/s10826-012-9571-7

Reichenberger, J., Richard, A., Smyth, J. M., Fischer, D., Pollatos, O., & Blechert, J. (2018). It's craving time: Time of day effects on momentary hunger and food craving in daily life. *Nutrition*, *55*, 15–20. https://doi.org/10.1016/j.nut.2018.03.048

Reimers, S., Maylor, E. A., Stewart, N., & Chater, N. (2009). Associations between a one-shot delay discounting measure and age, income, education and real-world impulsive behavior. *Personality and Individual Differences*, *47*(8), 973–978. https://doi.org/10.1016/j.paid.2009.07.026

Renn, O. (2008). *Risk governance: Coping with uncertainty in a complex world*. Earthscan.

Renn, O. (2020). Risk communication: Insights and requirements for designing successful communication programs on health and environmental hazards. In *Handbook of risk and crisis communication* (pp. 80–98). Routledge.

Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and individual differences*, *40*(2), 305–315. https://doi.org/10.1016/j.paid.2005.03.024

Robinson, E., Boyland, E., Chisholm, A., Harrold, J., Maloney, N. G., Marty, L., Mead, B. R., Noonan, R., & Hardman, C. A. (2021). Obesity, eating behavior and physical activity during COVID-19 lockdown: A study of UK adults. *Appetite*, *156*, 104853. https://doi.org/10.1016/j.appet.2020.104853

Romano, A., Spadaro, G., Balliet, D., Joireman, J., Van Lissa, C., Jin, S., Agostini, M., Bélanger, J. J., Gützkow, B., Kreienkamp, J., & Leander, N. P. (2021). Cooperation and Trust Across Societies During the COVID-19 Pandemic. *Journal of Cross-Cultural Psychology*, *52*(7), 622–642. https://doi.org/10.1177/0022022120988913

Romer, D., & Jamieson, K. H. (2020). Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. *Social Science & Medicine*, *263*, 113356. https://doi.org/10.1016/j.socscimed.2020.113356

Rubia, K., Halari, R., Christakou, A., & Taylor, E. (2009). Impulsiveness as a timing disturbance: Neurocognitive abnormalities in attention-deficit hyperactivity disorder during temporal processes and normalization with methylphenidate. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1525), 1919–1931. https://doi.org/10.1098/rstb.2009.0014

Rubinstein, A. (2003). "Economics and psychology"? The case of hyperbolic discounting. *International Economic Review*, *44*(4), 1207–1216. https://doi.org/10.1111/1468-2354.t01-1-00106

Samuelson, P. A. (1937). A note on measurement of utility. *The review of economic studies*, *4*(2), 155–161.

Saper, C. B., Chou, T. C., & Elmquist, J. K. (2002). The need to feed: Homeostatic and hedonic control of eating. *Neuron*, *36*(2), 199–211. https://doi.org/10.1016/S0896-6273(02)00969-8

Schlam, T. R., Wilson, N. L., Shoda, Y., Mischel, W., & Ayduk, O. (2013). Preschoolers' delay of gratification predicts their body mass 30 years later. *The Journal of pediatrics*, *162*(1), 90–93. https://doi.org/10.1016/j.jpeds.2012.06.049

Scholten, H., Scheres, A., de Water, E., Graf, U., Granic, I., & Luijten, M. (2019). Behavioral trainings and manipulations to reduce delay discounting: A systematic review. *Psychonomic Bulletin & Review*, *26*(6), 1803–1849. https://doi.org/10.3758/s13423-019-01629-2

Selten, R. (1965). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes.

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological bulletin*, *140*(2), 374. https://doi.org/10.1037/a0034418

Sheeran, P., & Webb, T. L. (2016). The Intention–Behavior Gap. *Social and Personality Psychology Compass*, *10*(9), 503–518. https://doi.org/10.1111/spc3.12265

Shelat, S., Cats, O., & van Cranenburgh, S. (2022). Traveller behaviour in public transport in the early stages of the COVID-19 pandemic in the Netherlands. *Transportation Research Part A: Policy and Practice*, *159*, 357–371. https://doi.org/10.1016/j.tra.2022.03.027

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Sibony, A.-L. (2020). The UK COVID-19 response: A behavioural irony? *European Journal of Risk Regulation*, *11*(2), 350–357. https://doi.org/10.1017/err.2020.22

Singer, J. (2022). *There are no accidents: The deadly rise of injury and disaster—who profits and who pays the price*. Simon; Schuster.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). Afex: Analysis of factorial experiments. *R package version 0.13–145*.

Skrynka, J., & Vincent, B. T. (2019). Hunger increases delay discounting of food and non-food rewards. *Psychonomic bulletin & review*, *26*(5), 1729–1737. https://doi.org/10.3758/s13423-019-01655-0

Snider, S. E., LaConte, S. M., & Bickel, W. K. (2016). Episodic future thinking: Expansion of the temporal window in individuals with alcohol dependence. *Alcoholism:*

clinical and experimental research, *40*(7), 1558–1566. https://doi.org/10.1111/acer.13112

Solnick, J. V., Kannenberg, C. H., Eckerman, D. A., & Waller, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation*, *11*(1), 61–77. https://doi.org/10.1016/0023-9690(80)90021-1

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and individual differences*, *47*(5), 385–395. https://doi.org/10.1016/j.paid.2009.04.008

Stedman, T. L. (1995). *Stedman's Medical Dictionary: 26th Edition* (M. Spraycar, Ed.; 26th). Williams & Wilkins.

Stein, J. S., Sze, Y. Y., Athamneh, L., Koffarnus, M. N., Epstein, L. H., & Bickel, W. K. (2017). Think fast: Rapid assessment of the effects of episodic future thinking on delay discounting in overweight/obese participants. *J Behav Med*, *40*(5), 832–838. https://doi.org/10.1007/s10865-017-9857-8

Stein, J. S., Wilson, A. G., Koffarnus, M. N., Daniel, T. O., Epstein, L. H., & Bickel, W. K. (2016). Unstuck in time: Episodic future thinking reduces delay discounting and cigarette smoking. *Psychopharmacology*, *233*(21), 3771–3778. https://doi.org/10.1007/s00213-016-4410-y

Steinglass, J., Figner, B., Berkowitz, S., Simpson, H. B., Weber, E. U., & Walsh, B. T. (2012). Increased Capacity to Delay Reward in Anorexia Nervosa. *Journal of the International Neuropsychological Society : JINS*, *18*(4), 773–780. https://doi.org/10.1017/S1355617712000446

Stevens, J. R., & Stephens, D. W. (2010). The adaptive nature of impulsivity. https://doi.org/10.1037/12069-013

Strickland, J. C., & Johnson, M. W. (2020). Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. *Psychological review*, *128*. https://doi.org/10.1037/rev0000263

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The review of economic studies*, *23*(3), 165–180.

Stroud, S. (2021). Weakness of Will. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. Retrieved December 20, 2022, from https://plato.stanford.edu/archives/win2021/entries/weakness-will/

Stuart, R. M., Abeysuriya, R. G., Kerr, C. C., Mistry, D., Klein, D. J., Gray, R. T., Hellard, M., & Scott, N. (2021). Role of masks, testing and contact tracing in preventing COVID-19 resurgences: A case study from New South Wales, Australia. *BMJ open*, *11*(4), e045941. https://doi.org/10.1136/bmjopen-2020-045941

Suleiman, R., & Rapoport, A. (1992). Provision of step-level public goods with continuous contribution. *Journal of Behavioral Decision Making*, *5*(2), 133–153. https://doi.org/10.1002/bdm.3960050205

Sun, J., Rhemtulla, M., & Vazire, S. (2021). Eavesdropping on Missing Data: What Are University Students Doing When They Miss Experience Sampling Reports? *Personality and Social Psychology Bulletin*, *47*(11), 1535–1549. https://doi.org/10.1177/0146167220964639

Swami, V., & Barron, D. (2020). Analytic Thinking, Rejection of Coronavirus (COVID-19) Conspiracy Theories, and Compliance with Mandated Social-Distancing: Direct and Indirect Relationships in a Nationally Representative Sample of Adults in the United Kingdom. https://doi.org/10.31219/osf.io/nmx9w

Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, *108*(29), 11825–11829. https://doi.org/10.1073/pnas.1102493108

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological bulletin*, *103*(2), 193. https://doi.org/10.1037/0033-2909.103.2.193

Team, R. C. (2013). R: A language and environment for statistical computing.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Thaler, R. H. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207. https://doi.org/10.1016/0165-1765(81)90067-7

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

THE EMPLOYMENT SITUATION — MAY 2021 (tech. rep. USDL-21-0980). (2021). U.S. Department of Labor.

Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2021). It depends on how you ask: Measuring bias in population surveys of compliance with COVID-19 public health guidance. *J Epidemiol Community Health*, *75*(4), 387–389. https://doi.org/10.1136/jech-2020-215256

Tomko, R. L., Solhan, M. B., Carpenter, R. W., Brown, W. C., Jahng, S., Wood, P. K., & Trull, T. J. (2014). Measuring impulsivity in daily life: The momentary impulsivity scale. *Psychological assessment*, *26*(2), 339. https://doi.org/10.1037/a0035083

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440. https://doi.org/10.1037/a0018963

Tuncgenc, B., El Zein, M., Sulik, J., Newson, M., Zhao, Y., Dezecache, G., & Deroy, O. (2020). We distance most when we believe our social circle does. https://doi.org/10.31234/osf.io/u74wc

Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), 40.

Van den Bergh, B., Dewitte, S., & Warlop, L. (2008). Bikinis instigate generalized impatience in intertemporal choice. *Journal of Consumer Research*, *35*(1), 85–97. https://doi.org/10.1086/525505

Van Lange, P. A., Balliet, D. P., Parks, C. D., & Van Vugt, M. (2014). *Social dilemmas: Understanding human cooperation*. Oxford University Press.

van Baal, S. T., Hohwy, J., & Verdejo-Garcia, A. (2021). Resisting urges of non-compliance during lockdown in the pandemic: Data & Code. https://doi.org/10.17605/OSF.IO/PJ9FT

van Baal, S. T., Le, S. T. T., Fatehi, F., Hohwy, J., & Verdejo-Garcia, A. (2022). Cory COVID-Bot: An Evidence-Based Behavior Change Chatbot for COVID-19. *Informatics and Technology in Clinical Care and Public Health*, 422–425. https://doi.org/10.3233/SHTI210948

van Baal, S. T., Le, S. T. T., Fatehi, F., Verdejo-Garcia, A., & Hohwy, J. (2022). Effecting behaviour change using an artificial intelligence chatbot: A pilot randomised controlled study. https://doi.org/10.31234/osf.io/2xuat

van Baal, S. T., Moskovsky, N., Hohwy, J., & Verdejo-García, A. (2022). State impulsivity amplifies urges without diminishing self-control. *Addictive Behaviors*, *133*, 107381. https://doi.org/10.1016/j.addbeh.2022.107381

van Baal, S. T., Walasek, L., Karanfilovska, D., Cheng, A. C., & Hohwy, J. (2022). Risk perception, illusory superiority and personal responsibility during COVID-19: An experimental study of attitudes to staying home. *British Journal of Psychology*, *n/a*, 1–22. https://doi.org/10.1111/bjop.12554

van Baal, S. T., Walasek, L., Verdejo-Garcia, A., & Hohwy, J. (2022). Impulsivity and Self-Control as Timeless Concepts. https://doi.org/10.17605/OSF.IO/9EVFT

Vanderveldt, A., Oliveira, L., & Green, L. (2016). Delay Discounting: Pigeon, Rat, Human – Does it Matter? *Journal of experimental psychology. Animal learning and cognition*, *42*(2), 141–162. https://doi.org/10.1037/xan0000097

Vassileva, J., & Conrod, P. J. (2019). Impulsivities and addictions: A multidimensional integrative framework informing assessment and interventions for substance use disorders. *Philosophical Transactions of the Royal Society B*, *374*(1766), 20180137. https://doi.org/10.1098/rstb.2018.0137

Verdejo-Garcia, A., Tiego, J., Kakoschke, N., Moskovsky, N., Voigt, K., Anderson, A., Koutoulogenis, J., Lubman, D. I., & Bellgrove, M. A. (2021). A unified online test battery for cognitive impulsivity reveals relationships with real-world impulsive behaviours. *Nature Human Behaviour*, *5*(11), 1562–1577. https://doi.org/10.1038/s41562-021-01127-3

Verdejo-Garcia, A., Tiego, J., Kakoschke, N., Moskovsky, N., Voigt, K., Lubman, D. I., & Bellgrove, M. A. (2020). A unified test of cognitive impulsivity reveals novel relationships with trait impulsivity and real world impulsive behaviour.

Verdejo-García, A., Benbrook, A., Funderburk, F., David, P., Cadet, J.-L., & Bolla, K. I. (2007). The differential relationship between cocaine use and marijuana use on decision-making performance over repeat testing with the Iowa Gambling Task. *Drug and Alcohol Dependence*, *90*(1), 2–11. https://doi.org/10.1016/j.drugalcdep.2007.02.004

Verdejo-García, A., Lawrence, A. J., & Clark, L. (2008). Impulsivity as a vulnerability marker for substance-use disorders: Review of findings from high-risk research, problem gamblers and genetic association studies. *Neuroscience and biobehavioral reviews.*, *32*(4), 777–810. https://doi.org/10.1016/j.neubiorev.2007.11.003

Verdejo-García, A., Moreno-Padilla, M., Garcia-Rios, M. C., Lopez-Torrecillas, F., Delgado-Rico, E., Schmidt-Rio-Valle, J., & Fernandez-Serrano, M. J. (2015). Social stress increases cortisol and hampers attention in adolescents with excess weight. *PLoS One*, *10*(4). https://doi.org/10.1371/journal.pone.0123565

Vincent, B. T., & Stewart, N. (2020). The case of muddled units in temporal discounting. *Cognition*, *198*, 104203. https://doi.org/10.1016/j.cognition.2020.104203

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Waitzberg, R., Hernández-Quevedo, C., Bernal-Delgado, E., Estupiñán-Romero, F., Angulo-Pueyo, E., Theodorou, M., Kantaris, M., Charalambous, C., Gabriel, E., Economou, C., Kaitelidou, D., Konstantakopoulou, O., Vildiridi, L. V., Meshulam, A., de Belvis, A. G., Morsella, A., Bezzina, A., Vincenti, K., Figueiredo Augusto, G., … Maresso, A. (2022). Early health system responses to the COVID-19 pandemic in Mediterranean countries: A tale of successes and challenges. *Health Policy*, *126*(5), 465–475. https://doi.org/10.1016/j.healthpol.2021.10.007

Weatherly, J. N., & Terrell, H. K. (2010). Delay Discounting of Different Commodities II: Confirmatory Analyses. *The Journal of General Psychology*, *138*(1), 35–48. https://doi.org/10.1080/00221309.2010.532522

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric Discounting in Intertemporal Choice: A Query-Theory Account. *Psychological Science*, *18*(6), 516–523. https://doi.org/10.1111/j.1467-9280.2007.01932.x

Weng, H. Y., Fox, A. S., Shackman, A. J., Stodola, D. E., Caldwell, J. Z., Olson, M. C., Rogers, G. M., & Davidson, R. J. (2013). Compassion training alters altruism and neural responses to suffering. *Psychological science*, *24*(7), 1171–1180. https://doi.org/10.1177/0956797612469537

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J., & Stouthamer-Loeber, M. (1994). Measuring impulsivity and examining its relationship to

delinquency. *Journal of abnormal psychology*, *103*(2), 192. https://doi.org/10.1037/0021-843X.103.2.192

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and individual differences*, *30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Wiers, R., Ames, S. L., Hofmann, W., Krank, M., & Stacy, A. (2010). Impulsivity, impulsive and reflective processes and the development of alcohol use and misuse in adolescents and young adults. *Frontiers in psychology*, *1*, 144. https://doi.org/10.3389%2Ffpsyg.2010.00144

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60*(2), 181–192. https://doi.org/10.1037/0022-3514.60.2.181

Wise, T., Zbozinek, T. D., Michelini, G., Hagan, C. C., & Mobbs, D. (2020). Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. *Royal Society Open Science*, *7*(9), 200742. https://doi.org/10.1098/rsos.200742

Wismans, A., Letina, S., Wennberg, K., Thurik, R., Baptista, R., Burke, A., Dejardin, M., Janssen, F., Santarelli, E., & Torrès, O. (2021). The role of impulsivity and delay discounting in student compliance with COVID-19 protective measures. *Personality and Individual Differences*, 110925. https://doi.org/10.1016/j.paid.2021.110925

Wolff, W., Martarelli, C., Schüler, J., & Bieleke, M. (2020). High boredom proneness and low trait self-control impair adherence to social distancing guidelines during the COVID-19 pandemic. https://doi.org/10.31234/osf.io/jcf95

Xu, A. J., Schwarz, N., & Wyer, R. S. (2015). Hunger promotes acquisition of nonfood objects. *Proceedings of the National Academy of Sciences*, *112*(9), 2688–2692. https://doi.org/10.1073/pnas.1417712112

Yarmush, D. E., Manchery, L., Luehring-Jones, P., & Erblich, J. (2016). Gender and impulsivity: Effects on cue-induced alcohol craving. *Alcoholism: Clinical and Experimental Research*, *40*(5), 1052–1057. https://doi.org/10.1111/acer.13030

Yates, J. F., & de Oliveira, S. (2016). Culture and decision making. *Organizational Behavior and Human Decision Processes*, *136*, 106–118. https://doi.org/10.1016/j.obhdp.2016.05.003

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin*, *146*(2), 118. https://doi.org/10.1037/bul0000218

Zhang, Y., Jiang, B., Yuan, J., & Tao, Y. (2020). The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study. https://doi.org/10.1101/2020.03.04.20031187

Zheng, H., Ton, H., Yang, L., Liufu, N., Dong, Y., Zhang, Y., & Xie, Z. (2019). Acute fasting does not induce cognitive impairment in mice. *Frontiers in neuroscience*, *13*, 896. https://doi.org/10.3389/fnins.2019.00896