School of Economics
University of Nottingham

Essays on Time Allocation

Richard Mills

PhD in Behavioural Economics
Network for Integrated Behavioural Sciences
Centre for Decision Research and Experimental Economics

January, 2023

Thesis submitted to the University of Nottingham
for the Degree of Doctor of Philosophy

Richard Mills: *Essays on Time Allocation.* PhD in Behavioural Economics. Supervised by Professor Chris Starmer and Professor Robin Cubitt. ©January, 2023

Acknowledgements

An African proverb says, *"it takes a village to raise a child"*. This certainly rings true for me if that 'child' is this thesis. There are several people in my 'village' that I would like to acknowledge. First and foremost, to my supervisors, Robin Cubitt and Chris Starmer, thank you for your incredible support over the years. I am extremely grateful for the opportunity offered to undertake this PhD as well as your wise guidance, knowledge and razor-sharp logic and reasoning. You have always gone above and beyond to help me, not only in developing the work presented in this thesis but in other areas of PhD life too.

I want to express my sincerest gratitude to NIBS for funding this PhD studentship through the *ESRC*[1] *NIBS*[2] *Postgraduate Research Scholarship*. It has also been a privilege to be affiliated with CeDEx[3]. Together, NIBS and CeDEx have allowed me to explore a field of research that I feel passionate about and have fostered an environment conducive to learning and camaraderie. A special thanks to Suzanne Robey who was always there to help, often sorting out my logistical woes as an international student. I am also grateful for the opportunities to present research and attend summer schools in the UK and mainland Europe. Many thanks to those who provided helpful comments at the NIBS workshops, PhD conferences, CeDEx, CREED[4] and CBESS[5] meetings, as well as CeDEx brownbags.

I am deeply grateful to my friends, especially Malte Baader, Ollie Baines, Sarah Bowen, Jose Guinot Saporta, Anna Hochleitner, Lara Suraci, and Yuliet Verbel Bustamante. Thank you for the many adventures and memories we have made and continue to create together. This PhD experience *truly* would not have been the same without you. All of you contributed in some shape or form to this thesis, whether it involved programming, helping collect data, proofreading, statistics, or generally providing useful feedback and advice —not to mention the many exchanges of baked goods.

To Mom, Dad, and Matthew —thank you for your unconditional love and encouragement. Last but certainly not least, to Candice Ford, for providing me with an abundance of love, support, laughter, and croissants (especially when I needed them the most). But primarily, thank you for ensuring that I maintain focus on the things in life that are important, but *not* urgent.

[1] Economic and Social Research Council
[2] Network for Integrated Behavioural Sciences
[3] Centre for Decision Research and Experimental Economics, University of Nottingham
[4] Centre for Research in Experimental Economics and Political Decision Making, University of Amsterdam
[5] Centre for Behavioural and Experimental Social Science, University of East Anglia

Contents

Overview of Figures

Overview of Tables

General introduction

"Few questions are as perplexing and profound as those that relate to time."
—Buonomano (2017, p.2)

Time is an integral part of the modern world for humans. The most commonly used noun in the English language is 'time' (Buonomano, 2017).[1] However, our modern preoccupation with time is a relatively recent construct. Davies (2019) discusses that humans first began measuring time as early as 5,500 years ago, where a stick was placed in the ground, casting shadows of varying lengths as the day proceeded. These 'sundials' were pretty sophisticated for the period, though they were rendered useless on days with cloud cover (unfortunate for those living in the UK). Indoors, candles, incense sticks, water clocks and even sand clocks were used. The earliest mechanical clocks came about in the 14th century, with their power derived from either an unwinding spring or falling weights. Over time, the accuracy of such mechanisms improved to the benefit of astronomers and scientists.

However, most people in a largely agrarian economy still lived by the natural progressions of the days and the seasons. Humans had a general disregard for 'clock time' —time was local to their area (Davies, 2019). The industrial revolution completely changed how we measure and perceive time (Chong, 2020; Davies, 2019; Thompson, 2016). Ocean travel meant that more accurate time instruments were needed, the railroad necessitated that time be standardised, and factories demanded better time-management (Thompson, 2016). Time became a measurable resource that employers could exploit to maximise output and profits from their employees. In this way, it eventually grew to *"...permeate all levels of society and industry"*, and it continues to have a profound influence on the world that we live in today (Chong, 2020).[2] In his seminal work titled *Time, Work and Industrial Capitalism*, Thompson (1967, p.61), wrote: *"Time is now currency: it is not passed but spent"*.[3]

Keynes (1930, p.322-329) predicted that due to technological advancements, *"our grandchildren"* in *"a hundred years hence"* would work around *"three hours a day"* —and that this would most likely only be by choice. However, this prediction hardly seems an accurate representation of the modern world. If anything, people currently appear to be afflicted by a perennial 'time-scarcity' problem (Pearson, 2015). At-least in the West, it is common to hear *"everyone's always rushing"* and *"the pace of life sure has sped up; you hardly have time to think now"* (Eisendrath, 1991, p.1). Various terms have emerged relating to 'time-famine' (Perlow, 1999), 'time-bind' (Hochschild & Arlie, 1997), and 'harried' (Linder, 1970) —all referring to the feeling of having

[1] In fact, Buonomano (2017) states that according to the Oxford dictionary, three of the five most regularly used nouns are related to 'time' —time, person, year, way and day.

[2] There are distinct differences between how parts of the world view time —'clock time' (clocks determine when tasks begin and end) vs. 'event time' (tasks move in a sequence and individuals transition between them) (see *A Geography of Time* by Levine (2008)).

[3] In factory production, time is a resource controlled by employers. However, in most modern occupations (perhaps due to increasing task complexity), decisions about how time is spent are made by employees who have the autonomy to determine the course of their day. Worker autonomy has been an important topic in recent years (Gagné & Bhave, 2011; Budd, 2018), and this may increase due to working-from-home predicted to gain in momentum and traction (Dingel & Neiman, 2020; Barrero et al., 2021).

too much to do and too little time to do it in.[4] Reflecting on this issue, Hamermesh & Lee (2007, p.374) wrote: *"Time stress is a problem analogous to poverty: both reflect scarcity of resources, time in the former case, goods in the latter. The only difference is that in a growing economy the goods constraint will relax over time, while the time constraint cannot. The time crunch will become relatively more binding for more people."*

Given the significance of time in our lives and its arguably growing scarcity, understanding how we choose to allocate it, particularly in the face of scarcity, is of utmost importance. In this thesis, we investigate what we believe are some interesting questions surrounding time allocation, particularly the various circumstances that may lead to poor time-use decisions. Understanding such circumstances can help provide a practical understanding of how we could use our time more effectively. We provide three largely independent but also related chapters (Chapters 1, 2 and 3), examining various ways in which people deviate from 'optimal' time allocation behaviour. This work departs from the standard economic optimisation paradigm, where we do not assume a rational 'optimising' agent.[5] Specifically, we focus on literature at the intersection between how various forms of *scarcity*[6] affect judgement and decision making as well as how individuals *bracket*[7] their choices.

In Chapter 1, the main question we ask is: *'How robust is the mere urgency effect?'* (Zhu et al., 2018). The mere urgency effect is a surprising phenomenon demonstrating that people tend to pursue tasks characterised by a sense of *urgency* (an illusion of expiration), despite these tasks being monetarily inferior options. The authors argue that studying this effect allows researchers to gain an understanding of a highly relevant real-world issue, *"...people's tendency to procrastinate on what is important to finish what is urgent."* (Zhu et al., 2018, p.674). While the effect has important implications for organisations and everyday life, prior evidential support is limited to Zhu et al.'s (2018) paper. As a result, we attempt to replicate the mere urgency effect. Beyond a standard replication, we conduct an independent and in-depth replication exercise attempting to rule out hypothesised confounds in the original experimental design.

The effect is investigated experimentally in a simple one-shot dichotomous choice environment (Zhu et al., 2018).[8] Through *four* replication studies, we find that, while the mere urgency effect attenuates after removing hypothesised confounds in the original experimental design, it remains robust. We also observe that *relative attentional focus* on attributes to do with timing over payoffs plays an essential role in choice behaviour. This work demonstrates that tasks characterised by a sense of urgency capture attention and result in higher propensities of suboptimal

[4] There has been a proliferation of books titled: *Work Without End* (Hunnicutt, 1988), *Fighting for Time* (Epstein & Kalleberg, 2004), *Busier than Ever* (Darrah, 2007), *Pressed for Time* (Wajcman, 2014) ,*The Ruthless Elimination of Hurry* (Comer, 2019) and many more.

[5] Research on *time allocation* within economics has historically focused on an 'optimal consumer' who allocates their time. For example, in his seminal work, *Theory of allocation of time*, Becker (1965) laid the foundation of time as an important input in the household production function and spurred a large body of subsequent research on the topic (see Juster & Stafford (1991) for a comprehensive review). Our focus departs from this perspective in that we do not assume a 'rational' agent (Simon, 1979; Camerer, 1999).

[6] This refers to *scarcity theory*. In Chapters 1 and 3, we provide the relevant literature surrounding this research area. Some important papers are Mullainathan & Shafir (2013); Zhu et al. (2018); Zhao & Tomm (2018); Hamilton et al. (2019a); Cannon et al. (2019); De Bruijn & Antonides (2021); and Bernheim et al. (2019).

[7] This refers to *choice bracketing*. In Chapter 2, we provide a literature review surrounding this work. The most relevant papers are Herrnstein et al. (1993); Read et al. (1999b); Koch & Nafziger (2019), and Ellis & Freeman (2020).

[8] Subjects are presented with two options that have the same underlying typing task, but which vary with respect to (i) time available and (ii) payoffs. They make a single choice about which task to work on and cannot switch after this decision.

behaviour. It also provides strong support for the mere urgency effect, adding to its evidential base, and contributes to the current discourse surrounding the importance of replication studies within the social sciences.

In Chapter 2, we move to a continuous time allocation environment and examine the role that choice bracketing plays.[9] The key questions we ask are: *'Does narrow bracketing in time allocation exist? And if so, how robust is it?'* Choice bracketing can be separated into two distinct behaviours: *narrow* and *broad* bracketing. In their seminal work, Read et al. (1999b, p.171,193) describe these distinct types as the following: *"When making many choices, a person can broadly bracket them by assessing the consequences of all of them taken together, or narrowly bracket them by making each choice in isolation...Narrow bracketing generally shifts people's attention from the macro level to the micro level —a level at which many of the most pernicious patterns of decision making seem to occur."*[10] While there is extensive literature showing the importance of choice bracketing in various decision contexts, surprisingly little research, to our knowledge, explores choice bracketing concerning time allocation.

In this work, we set out to experimentally investigate whether narrow bracketing in time allocation occurs, and if so, how sensitive it is in the studied environment. As an initial step, we construct a simple theoretical framework, which allows us to disentangle narrow from broad bracketing behaviour. Through *three* studies, we find that there is a significant prevalence of narrow bracketing in time-use decisions and that it remains remarkably robust in the studied environment. We attempt to unpick the underlying motivations for such persistent occurrence of narrow bracketing and explore the correlates surrounding it. While financial considerations are important in determining narrow bracketing in our experimental environment (individuals are lured by the initially higher rewards of some tasks, neglecting the implications of this for broader payoff maximisation), the time-restriction (a form of time-scarcity) placed on some tasks also plays a role in choice. Moreover, we find some important correlates of suboptimal behaviour. To our knowledge, this research is a first of its kind, providing strong evidence for the high prevalence of narrow bracketing in time-use decisions in a controlled experimental environment. We also add to the debate surrounding how cognitive ability and specific demographics relate to narrow bracketing behaviour.

In Chapter 3, utilising the novel theoretical and experimental design constructed in Chapter 2, we explore the effect that two forms of time-scarcity have on time allocation behaviour. Specifically, the core question we ask is *'To what extent do urgency and busyness induce narrow bracketing in time allocation?'* These dimensions are interesting because while they are distinct in and of themselves, both fall under the same category of 'scarcity' —that Mullainathan & Shafir (2013) argue forms its own distinct psychology.[11] Specifically, the authors conjecture that scarcity

[9]Subjects are free to allocate themselves in *real* time across options that vary in (i) time available and (ii) payoffs. They can switch across options as they please, subject to availability.

[10]As an example, consider the simple decision to purchase a daily coffee. Someone making a series of local choices each day might reason to themselves that the costs of a single coffee are relatively trivial —consequently buying their daily coffee. However, if that same individual thought about the global consequences of purchasing a daily coffee for a more extended period, say a year, they might realise that the costs add up to a significant amount (a £3 coffee everyday adds up to over £1,000 in a year) and might not choose to buy the daily coffee. This is especially true if that person prefers to spend that same money on something else, such as a holiday. While a fairly trivial example, such accounting has profound implications for almost every decision environment that involves the cumulative effects of multiple choices (arguably almost all of the choices we face in life).

[11]The psychology of scarcity focuses on the logic and consequences of 'scarcity' —what happens to our cognition

of *any* kind captures our attention, focusing cognitive resources on what is perceived as scarce, but at the neglect of broader things outside of where attention is placed. In our context, such elements of scarcity are predicted to increase focus on narrow choices at the disregard of the broader picture —consistent with higher propensities of narrow bracketing.

This work directly extends our research from Chapters 1 and 2, by exploring the mere urgency effect in continuous time. Moreover, to our knowledge, we are the first to investigate busyness and its effects on choice bracketing in an experimental framework. To investigate urgency and busyness, we experimentally manipulate (i) the time at which certain tasks are available, and (ii) the extent to which subjects are short of time. Contrasting the predictions that such elements of scarcity result in higher levels of narrow bracketing, an experiment shows that these manipulations do not adversely impact time-use behaviour of subjects. That said, in the experiment we also measure subject *perceptions* and those who perceive tasks as more urgent or themselves busier are significantly more likely to engage in suboptimal behaviour. Since the key conjecture under the psychology of scarcity is that perceptions matter for suboptimal time-use, these results align to some extent with the theory posited by Mullainathan & Shafir (2013). Ultimately, we find that individual perceptions lie at the heart of choice behaviour, not necessarily the physical constraints we impose. This research provides an important springboard for future research that explores the role of scarcity on suboptimal time-use.

This thesis concludes with a summary of the key takeaways from Chapters 1, 2 and 3 and offers a more interpretive discussion surrounding the results, particularly across the three chapters as a collective.

when we feel we have too little of something and how this affects judgement and decision-making. We all have a physical constraint of 24 hours in the day, yet the scarcity of time is much more apparent on days when our schedules are rammed with things to do. Similarly, how cold we feel depends on the absolute temperature outside, but also on our own physiology (Mullainathan & Shafir, 2013). Specifically, Mullainathan & Shafir (2013) conjecture that when people *feel* resources are low relative to their needs, a scarcity 'mindset' emerges, which can have profound implications for how they make decisions —largely to their detriment (Mullainathan & Shafir, 2013; Mani et al., 2013; Shah et al., 2012, 2015).

CHAPTER 1

The Mere Urgency Effect —an artefact of experimental design?

Abstract

In the modern world, individuals are frequently confronted with decisions about how to allocate their time between tasks that vary in their degree of *urgency* and *importance*. This research investigates the *mere urgency effect* (Zhu et al., 2018), a novel finding showing that individuals are more likely to pursue unimportant tasks when they are characterised by *spurious* time-pressure. We conduct four experiments to replicate this effect and examine if it remains robust to the removal of hypothesised confounds in the experimental design. With a combined sample size of $N = 2,003$, our results show that although the mere urgency effect significantly attenuates after controlling for confounds, it remains robust, providing further compelling evidence for its existence. Moreover, we find strong evidence for the *attention-based psychological account* as an explanation behind the effect. Subjects in our experiments have an overwhelming tendency to pursue tasks that are made urgent, irrespective of whether it makes financial sense for them to do so, and this is mediated by relative attentional focus on temporal versus pecuniary elements in the decision-environment.

1.1 Introduction and motivation

"I have two kinds of problems: the urgent and the important. The urgent are not important, and the important are never urgent."

—Dwight D. Eisenhower (1954)

Few would deny that time is central to the way we live. The modern world with the continuous barrage of social media, 24-hour newsfeeds, endless piles of to-do-lists, and family demands has left many feeling an acute shortage of time (Mullainathan & Shafir, 2013; Perlow, 1999; Hochschild & Arlie, 1997; Linder, 1970). One perennial and difficult aspect of our daily lives is having to choose how to allocate our limited time and attention between tasks that vary in their urgency (completion-windows) and importance (net benefit). As an example, you might come into work and find two emails awaiting a response, one regarding something that is unimportant but should be done very quickly (e.g., briefing a colleague for an imminent meeting on a minor issue) and the other regarding something very important but not so time-sensitive (e.g., discussing next steps in a long-term research project). Which do you decide to respond to?

Through five experiments, Zhu et al. (2018, p.673) find that *"...people are more likely to perform unimportant tasks (i.e., tasks with objectively lower payoffs) over important tasks (i.e., tasks with objectively better payoffs), when the unimportant tasks are characterised merely by spurious urgency*

This chapter is based on joint work with Chris Starmer and Robin Cubitt. We would like to thank the CeDEx members, subjects at the 2019 NIBS workshop, the 2020 CCC meeting, and other friends and colleagues for their invaluable input. We also thank CeDEx and NIBS for funding this research. NIBS grant number: RA1175. The experiments were approved by the Nottingham School of Economics' Research Ethics Committee.

(e.g., an illusion of expiration)."[1] The authors coin this phenomenon the *mere urgency effect*, and argue that it violates the basic normative principle of dominance. Specifically, there are multiple *rational* reasons for why one might prioritise urgent tasks over important ones in some contexts.

For example, urgent tasks might be contingent on each-other (Zhu et al., 2018), have high demand or low supply (Brehm, 1966; Brock, 1968; Greenwald, 1968; Worchel et al., 1975; Brock & Brannon, 1992; Cialdini, 2009), or have payoffs that are realised sooner (Frederick et al., 2002; McClure et al., 2004). Important tasks, on the other hand, might be harder (O'Donoghue & Rabin, 2001), and have more distant or uncertain goal completion (Hull, 1932; Kivetz et al., 2006). Moreover, in certain situations, individuals want to complete urgent tasks first before moving to important ones later (Shin & Ariely, 2004). These examples are a non-exhaustive list of normative reasons for urgency pursuit. However, what is interesting and surprising about the mere urgency effect is that Zhu et al. (2018) employ an experimental design where the unimportant task is unequivocally dominated by the important one; yet, under spurious time pressure, they find a preference for the unimportant task. The time pressure is 'spurious' because it is incidental and uninformative for rational choice, in the sense that decision-makers always have sufficient time to make a decision and finish a task. Zhu et al. (2018, p.674) use the mere urgency effect *"...to gain an insight into an issue of high real-world relevance: people's tendency to procrastinate on what is important to finish what is urgent."*

The phenomenon is captured by the following *"...highly simplified yet well-controlled and flexible..."* experimental design provided in Table 1.1 (Zhu et al., 2018, p.676). Subjects are ran-

	Bonus	**Expiration time**
Control condition:		
Task A	\$0.12	50 minutes
Task B	\$0.16	50 minutes
Urgency treatment:		
Task A	\$0.12	5 minutes
Task B	\$0.16	50 minutes

Table 1.1: Experimental design

Source. Design from Experiment IIB (Zhu et al., 2018).
Notes. This experiment involves a between-subject design, with subjects allocated to either control or urgency treatment (with 0.50 probability). A one-shot choice between 'Task A' or 'Task B'. The nature of the tasks are identical: type as many randomly generated six-letter strings in reverse order (e.g., 'oiggcj' to 'jcggio') in 3 minutes. The only difference between the control and urgency conditions is that the task availability differs in the urgency treatment. Subjects are informed that they are only allowed to participate in the experiment once. Zhu et al. (2018, p.676) motivate that the choice set-up is designed such that the restricted option is always *"clearly"* dominated by the nonrestricted one. Specifically, the authors state that the restricted option does not have tasks that are dependent on each other, higher demand or lower supply, payoffs that are realised sooner, differences in goal progress, difficulty or length, and finally, any order effect present as individuals are only allowed to work on one task.

[1] The five experiments from Zhu et al. (2018) are as follows. In Experiment I, the authors showed preliminary evidence for the mere urgency effects with students. In Experiments IIA and IIB they tested the attention-based psychological account through *ex-post* questions. In Experiment III, they examined whether reminding subjects about payoffs at moment of choice attenuates the mere urgency effect. In Experiment IV, they examined if perceived busyness exacerbates the effect and whether it persists over time. For reasons explained below (see Section 1.3.2), we focus on Experiment IIB. Link to the paper.

domly allocated to a control or urgency condition, where they need to make a one-shot choice between one of two tasks to work on: 'Task A' or 'Task B'. The tasks both involve an identical typing assignment that last for a fixed duration of 3-minutes. Task *importance* and *urgency* are manipulated through varying the bonus and completion windows offered to finish the task.[2]

In the control condition, subjects are presented with a choice between a low-bonus ($0.12) and higher-bonus ($0.16) task, with the same non-urgent expiration time (50 minutes). In the urgency treatment, while subjects are presented with the same choice between a low-bonus ($0.12) and a higher-bonus ($0.16) task, the low-paying task has a 5-minute completion window (creating urgency), and the higher-paying task a non-urgent 50-minute expiration (as in the control condition). Specifically, the 'non-restricted' (50-minute expiration) option in the urgency treatment offers both a higher payment for completing the typing task and a window of opportunity to do so that is a strict superset of that offered by the 'restricted' (5-minute expiration) alternative. A subject can complete the typing task within the first 5 minutes of making their choice, regardless of whether its availability expires in 5 or 50 minutes. However, in the latter case, subjects also have the option to complete it after the first 5 minutes (which they do not in the former). Thus, the low bonus option is weakly dominated in the control condition and strictly dominated in the urgency treatment.[3]

Despite the dominated nature of the urgent option, Zhu et al. (2018) observe the striking result that while the vast majority of subjects unsurprisingly choose the higher-payoff option in the control condition (just 7% choose the low-payoff option), a large share (48%) of subjects choose the low-payoff option in the urgency treatment.[4] The authors argue that the underlying mechanism behind the mere urgency effect is the *attention-based psychological account*: *"The limited time frame embedded in an urgent task is a salient restriction in the local decision context, and it elicits attention, diverting focus away from the magnitudes of task outcomes (e.g., payoffs)."* (Zhu et al., 2018, p.675). This phenomenon is related to a growing literature that explores how *choice restrictions* draw attention and create attentional neglect relative to unrestricted domains (Botti et al., 2008; Roux et al., 2015; Kristofferson et al., 2016; Mehta & Zhu, 2015; Cannon et al., 2019; Hamilton et al., 2019a; Johar et al., 2015; Zhu & Ratner, 2015; Zhu et al., 2018; Berlyne, 1969; Pribram & McGuinness, 1975; Mullainathan & Shafir, 2013; Shah et al., 2012; Brendl et al., 2003; Cialdini, 1993).[5] A crucial element of the mere urgency effect theorisation is that the heightened attentional focus caused by the choice restriction results in a subsequent urge to pursue that which is restricted (Atkinson, 1953; Zeigarnik, 1938; Finucane et al., 2000; Kahneman & Frederick, 2002; Loewenstein et al., 2001; Stanovich & West, 2000; Zur & Breznitz, 1981; Svenson & Edland, 1987; Svenson et al., 1990).[6]

The mere urgency effect has the potential to have widespread and valuable implications for the use of time in organisations and daily life. However, given that this is a newly discovered

[2]A note on terminology: throughout this paper, we make reference to 'completion window', 'availability' and 'expiration' interchangeably —all referring to the time that any given task is available, but not the duration of the task (which is always fixed at 3-minutes).

[3]This is under the assumption that people prefer: (i) more money than less, and (ii) less restricted time than more.

[4]These results refer to Experiment IIB of Zhu et al. (2018).

[5]Botti et al. (2008, p.185) define a choice restriction as *"...any internally or externally imposed boundary condition that limits and/or confines choices."*

[6]For more detail surrounding the rationale behind the attention-based account and related literature, see Zhu et al. (2018).

phenomenon, with essentially one paper for its evidential support, the mere urgency effect demands further investigation. In this paper, we attempt to replicate the mere urgency effect, and conditional on doing so, see if it remains robust to a variety of changes in experimental design —to rule out confounds that might explain the observed behaviour. We followed a three-phase sequential replication methodology: (i) re-analysed the data from the original study, (ii) re-ran the experiment as in Table 1.1 following its exact procedures with a new subject pool, and (iii) probed at the robustness of the effect (see Levitt & List 2009; Czibor et al. 2019; and Alempaki et al. (2019)). In phase (iii), our strategy had two parts. First, we conducted additional replications that attempted to identify, and isolate confounds in the original design that might explain the dominance violation —factors that we believe are distinct from an 'urgency' effect. Second, we employed a 'fresh-eyes' approach, where we constructed an entirely new set of experimental instructions with the aim of minimising scope for misunderstanding by subjects. Ultimately, our core aim in the third phase was to provide what we thought were the 'fairest' tests for the existence of the mere urgency effect. Taken together, the results from four replication studies, show that while the mere urgency effect attenuates as we remove hypothesised confounds in the original experimental instructions, it proves to be a robust phenomenon. There is also strong evidence underpinning attention as a significant correlate of choice behaviour. While this chapter contributes to the existence and robustness of the mere urgency effect, it also adds to current discourse surrounding the importance of conducting independent and in-depth replication studies.

The rest of this chapter is organised as follows. Section 1.2 outlines the literature on dominance violations which has important relevance to the mere urgency effect. Sections 1.3 and 1.4 provide the experimental designs, procedures, and results across the three-phase sequential replication process. Section 1.5 provides meta-analytic estimates of effect sizes. Section 1.6 provides an overall discussion of the results and provides areas for future research avenues. Section 1.7 concludes.

1.2 Literature surrounding dominance violations

A core assumption inherent in a large body of economic literature is that rational agents will choose dominant options instead of dominated alternatives (Kourouxous & Bauer, 2019). Tversky & Kahneman (1986, p.253) describe dominance as *"...perhaps the most obvious principle of rational choice"*. However, a substantial research effort over the years has documented many instances where dominant alternatives are not chosen. Kourouxous & Bauer (2019) provide a comprehensive review, stating amongst others that the leading causes of such violations in rational choice are framing, reference points, certainty effects, bounded rationality, and emotional responses.[7] Perhaps most relevant to the current research are *framing effects* and *emotional responses.* Kourouxous & Bauer (2019, p.210) document *"...that framing, i.e., the description and presentation of a decision problem, strongly influences the decision-maker's choices."* That is, the way outcomes are described and alternatives presented, can influence a decision-maker into choosing a dominated option. Tversky & Kahneman (1986, p.265) report that people find dominance to be a compelling reason to choose an option, although when the dominance is *"masked by a frame"*,

[7]While much of the literature on dominance violations concerns risk, we believe that an analogous idea of dominance applies in our context of payoffs and time.

then they might be inclined to choose the inferior option.

Problems 1 and 2 in Table 1.2 are examples of how the same choice between two lotteries can be framed differently. Each lottery is characterised by the percentage of different coloured marbles in a box and the amount of money subjects can win or lose, depends on the randomly drawn colour. Presented as in Problem 1, it is quite easy to see that option B dominates option A. However, Problem 2 presents an identical reformulation of the decision problem (combining red and green in option B and yellow and blue in option A). While reformulating the problem simplifies the decision (by giving fewer options to choose from), it masks the transparency of the dominance. Indeed, Tversky & Kahneman (1986) find that while 100% of subjects choose option B in Problem 1, just 42% choose option D in Problem 2. This result is very similar to the well-known *isolation effect* studied by Tversky & Kahneman (1985), where subjects fail to correctly combine multiple lotteries into one lottery to identify the dominant option. Nevertheless, violations of dominance attributed to framing effects are widespread in the literature (see Kahneman & Tversky (1981); Tversky & Kahneman (1985, 1989); Rabin & Weizsäcker (2009); Li (2001); Wang (1996); Birnbaum & Bahra (2007); Starmer (1999); and Yamagishi (1997)).[8]

Problem 1	Option A	90% white $0	6% red win $45	1% green win $30	1% blue lose $15	2% yellow lose $15
	Option B	90% white $0	6% red win $45	1% green win $45	1% blue lose $10	2% yellow lose $15
Problem 2	Option C	90% white $0	6% red win $45	1% green win $30	3% yellow lose $15	
	Option D	90% white $0	7% red win $45	1% green lose $10	2% yellow lose $15	

Table 1.2: Transparency of dominance violation

Source. Decision problem from Tversky & Kahneman (1986).

Another essential driving mechanism behind dominated choice is emotions. Kourouxous & Bauer (2019, p.228) discuss that emotional responses refer to instances where the decision-maker is influenced by *"...emotions and other human traits"* that trigger ad hoc heuristics or attentional shifts that impede decisions (Sunstein & Zeckhauser, 2011; Sunstein, 2003; Chanel & Chichilnisky, 2009, 2013; Arrow, 1966; Wickham, 2003; Johnson et al., 1993). For example, Johnson et al. (1993) observes that people often place greater values on options that represent probability subsets of other alternatives. When asking subjects to assign a value to various life insurance policies, they find that policies covering terrorist attacks and other mechanical failures are priced significantly higher than policies that cover *any* cause of death. Since the former is a subset of the latter, this choice represents a dominance violation. The authors attribute this result to frequent media coverage of such events while flying, triggering an emotional response from the decision-maker.

The literature demonstrates that dominance violations, to a certain degree, depend on the structure of the decision-environment. This has important implications in the context of the mere

[8]For a more comprehensive overview of the literature and violations of dominance on the grounds of framing, see Kourouxous & Bauer (2019).

urgency effect and should be considered carefully. The framing of attributes at the moment of choice (i.e., the transparency of the dominated nature) and the extent to which 'urgency' triggers an emotional response both have an essential role to play in the propensity of dominated choice. These features also interact with attention, which can be shifted to some aspects of the decision-environment depending on how and what information is presented.

1.3 Can we replicate the effect?

1.3.1 The importance of replication

While replication is encouraged in the academic realm, it has struggled to gain much traction and remains an ideal to be *"...professed but not practised"* (Andreoli-Versbach & Mueller-Langer, 2014, p.1621). For example, just 0.1% of articles in the top-50 economic journals between 1974 and 2014 are replication studies (Mueller-Langer et al., 2019). However, interest in reproducibility has grown in recent years, mainly because the robustness of results in some fields such as psychology, neuroscience and genetics are increasingly being scrutinised (Camerer et al., 2016, 2018; Open Science Collaboration, 2015; Klein et al., 2014; Czibor et al., 2019). For example, the *Reproducibility Project: Psychology* replicate 100 studies published in three top psychology journals and find that just 36% of the studies replicate in the same direction as the original effect sizes (Open Science Collaboration, 2015).

In economics, while not as low, the replication rate is still somewhat concerning (Duvendack et al., 2017). Camerer et al. (2016) conduct the first systematic review of laboratory experiments in economics. They replicate 18 studies published in the *Quarterly Journal of Economics* and *American Economic Review* and find that 61% (11 out of the 18) produce the same effect size as the original studies. While these large-scale replication agendas examine whether findings replicate across a wide range of studies, describing whether there is a general issue with reproducibility, they do not as Shah et al. (2019, p.6) put it *"...permit a deeper dive into individual projects"*. In line with this view, the focus of this research is to provide an independent and focused deep dive into a novel phenomenon, the mere urgency effect. To borrow more words from Shah et al. (2019, p.6), replication *"...provides an important channel by which we continue to ensure that we are building our science on a foundation of solid results."*

1.3.2 Reanalysis of original data and pure replication

Experimental design and procedures

Zhu et al. (2018) conducted five experiments to motivate the existence, underlying mechanisms as well as attenuating and exacerbating factors surrounding the mere urgency effect. After careful consideration, we chose their third experiment, Experiment IIB (EIIB), as we believe it best encompasses what the authors were after: to show not only the existence of the mere urgency effect but also to provide evidence for the attention-based psychological account. Moreover, in almost all their other experiments, they employ some close variant of the design employed in EIIB. To conduct Replication I (RI), we contacted the authors to obtain their instructions and data from Experiment IIB (EIIB).[9] Figure 1.1 provides a screenshot generated from running Zhu et al.'s (2018) Qualtrics code of the experimental instructions for the control and urgency conditions that subjects see near moment of choice. Note that each outlined screenshot represents a new page that subjects enter, once they have clicked the 'continue' button on the screen.

Analogous to EIIB, RI was conducted on Amazon Mechanical Turk (MTurk) (US sample) and coded on Qualtrics. The experiment consisted of three parts. First, contractual workers on MTurk performed an attention check question (ACQ), where they needed to demonstrate that they were paying attention to the instructions.[10] Those who failed to do this were excluded from the study and were not paid. The remaining subjects then moved on to the actual instructions, which explained the nature of the experimental procedure and options. At this stage, they were also presented with *ex-ante* comprehension questions about payoffs, availability, and task features. They could only proceed to the next stage after answering these questions (they were not excluded if they answered incorrectly). Second, MTurk workers made their decision and subsequently worked on the typing task. Third, subjects responded to a series of *ex-post* questions about reasons for task choice, attention, comprehension, and general demographics. Specifically, subjects were asked: an open-ended question about why they chose the task that they did, how much attention they paid to bonus amounts and task availability, their ability to recall bonus and task expiration, their age, gender, income, and education, and any other comments they wanted to make. There were other additional questions that we asked subjects, however, for brevity, we do not state these here. Section A.1.1 in the Appendix provides a comprehensive set of instructions and questions used.

The data for RI was collected in April 2019 in a single session within the original study hours (10h00-15h00 US time). We excluded individuals who participated in EIIB to rule out possible learning effects from one study to another. In line with best practise in the replication literature, we conducted a series of *a-priori* power calculations to obtain at-least 90% power (see Section A.2 in the Appendix). In total, 400 contractual workers completed the study for a fixed fee of $0.50 plus a bonus of either $0.12 or $0.16, depending on which task they chose.[11] Subjects took an average of 11.67 minutes to complete the study (earning $3/hr).[12] See Table A1 in the Appendix for summary statistics of EIIB and RI.

Reanalysis and pure replication: Results

Although the urgent unimportant task (5-minute availability and $0.12 bonus) is dominated by the non-urgent important task (50-minute availability and $0.16 bonus), according to the mere urgency effect, it is expected that more contractual workers will choose the unimportant task under a smaller task availability —that is, the dominated option will be chosen more frequently

[9]A huge thank you to Meng Zhu and her team. Meng was very helpful in disseminating any resources that we requested (i.e, Qualtrics code, data) and responded quickly to any clarifying questions. We cannot thank you enough.

[10]Specifically, the ACQ required them to ignore a series of sports items listed and instead input the word 'typing' in a blank box provided at the bottom of the same page. Such attention checks are common in online studies as they filter out inattentive subjects, thus improving data quality. We also only recruit MTurk workers who have a high reputation with above 95% approval ratings (Sheehan, 2018; Oppenheimer et al., 2009; Paolacci et al., 2010).

[11]In EIIB, 200 contractual workers took an average of 9.35 minutes. The longer duration in our replication is because of the additional *ex-post* questions that we added.

[12]While the hourly wage may seem low, Hara et al. (2018) recorded 3.8 million HITs on MTurk and show that median hourly wage is $2/h.

in the urgency than in the control condition. The top panel in Figure 1.2 shows the dominated choice share between treatments and across experiments. 'Experiment IIB' (EIIB) refers to the our reanalysis of the original data provided by Zhu et al. (2018). We conduct the same statistical analysis as the authors, the results of which are shown in Table 1.3 (with their accompanying

(A) Experiment IIB (Control condition)

(B) Experiment IIB (Urgency condition)

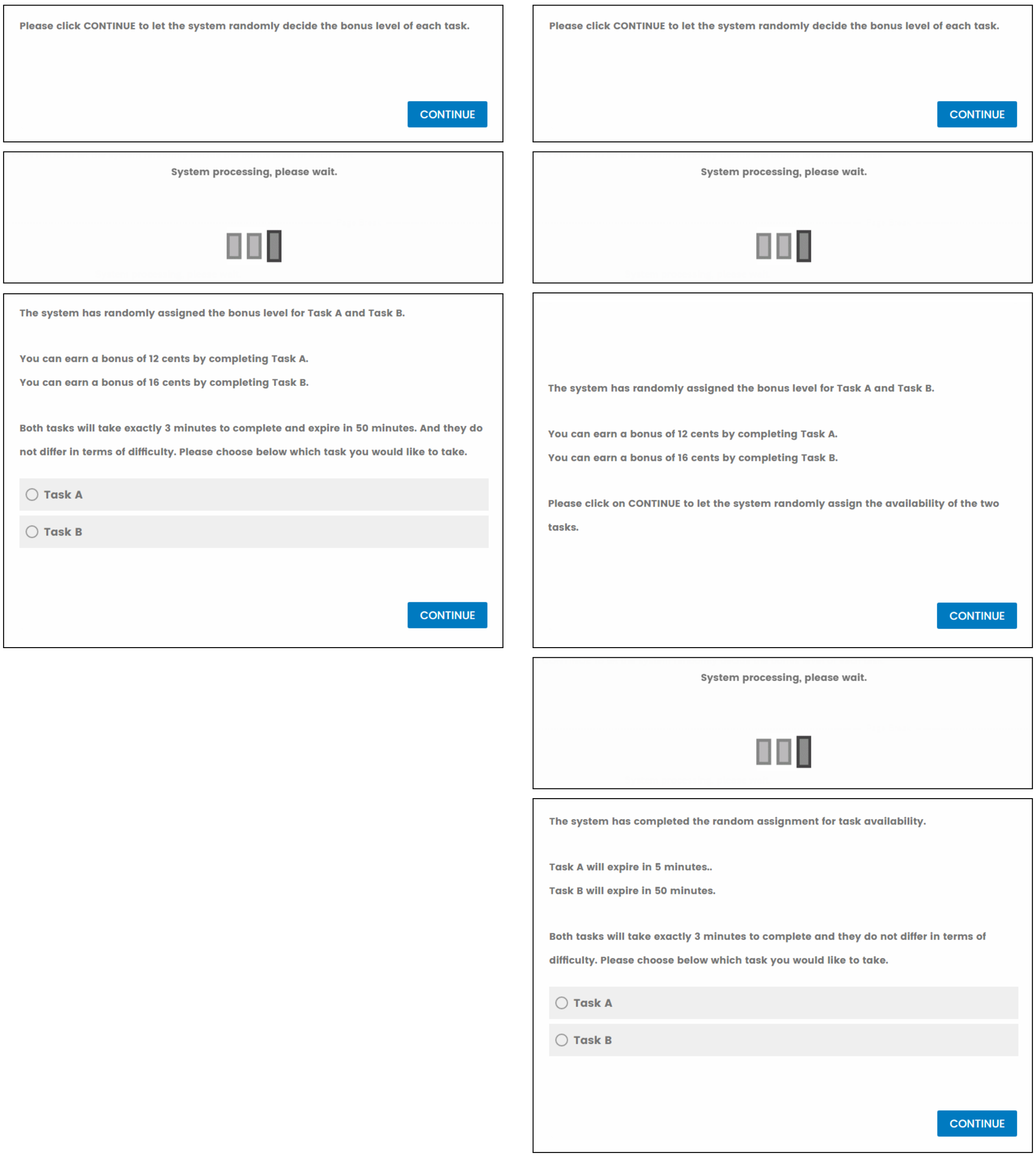
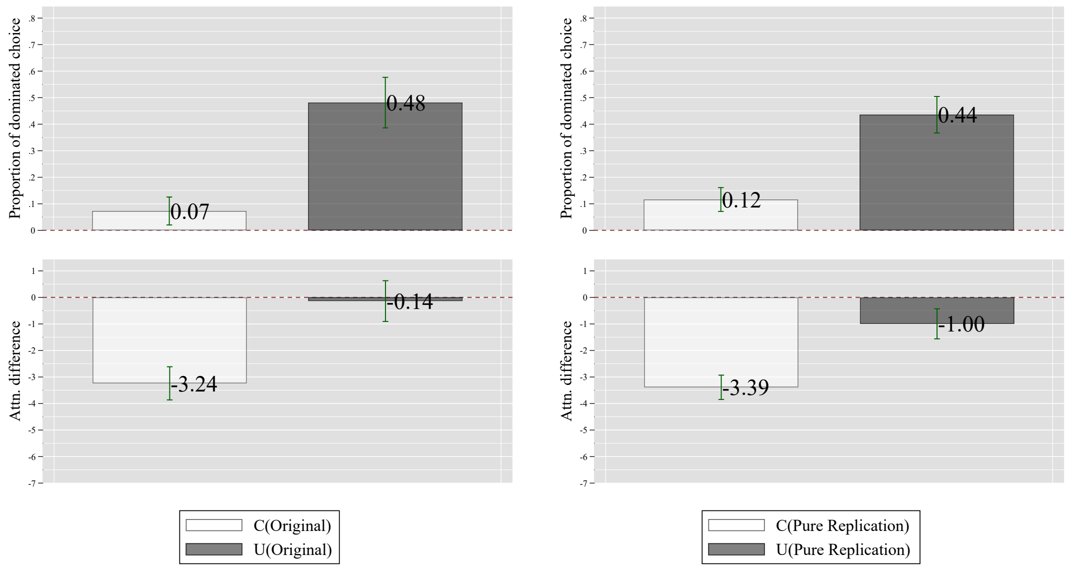


Figure 1.1: Experimental instructions near point of decision (Experiment IIB)

Notes. These are the experimental instructions near the decision point for Experiment IIB from Zhu et al. (2018). The information about the study design is provided in Table 1.1. One additional note is that the tasks for either condition were not actually randomised on screen (i.e., the 'system randomiser' was not real). That is, Task A was always the dominated option and presented first on the decision screen, while Task B was always the dominant option and presented second on screen. We discuss the implication of this in Section 1.4.

p-values and significance levels).[13] This reanalysis confirms that there is indeed a significantly lower proportion of subjects who chose the dominated option in the control condition relative to the urgency with the exact same choice proportions as in Zhu et al. (2018) (7.3% and 48.1%).[14] 'Replication I' (RI) refers to the results from our first replication. The results show that the dominated choice proportions are 11.6% and 43.6% respectively, indicating comparable results to Zhu et al. (2018).

The bottom panel in Figure 1.2 shows the mediating role of relative attention. In the same manner as Zhu et al. (2018), a composite measure of relative attentional focus is constructed by subtracting the *ex-post* measures of attention to task expiration from bonus amounts. Since both measures are on a scale from 1 to 8, the relative attention measure ranges from -7 to $+7$, with higher values illustrating relatively more attention to timings than payments.[15] The results for EIIB and RI reaffirm that there is relatively more attention to bonus amounts in the control condition compared to the urgency treatments. Moreover, mediation analysis shows that relative attentional focus significantly mediates the effect of urgency on task choice in both EIIB ($\beta = 1.25$, $SE = 0.34$, 95% CI = $[0.70, 2.04]$) and RI ($\beta = 1.06$, $SE = 0.23$, 95% CI = $[0.69, 1.58]$).[16] These results



(a) Experiment IIB

(b) Replication I

Figure 1.2: Dominated choice share between treatments (EIIB and RI)

Source. Data for EIIB provided by Zhu et al. (2018).
Notes. Figure 1.2a refers to data from EIIB, while Figure 1.2b data from RI. The top panel in both figures represents dominated choice proportions across the control condition and urgency treatment for EIIB and RI, respectively. The bottom panel in both figures represents the corresponding relative attentional focus across the control condition and urgency treatments for EIIB and RI. Attn. difference (-7 to $+7$) = Attn. expiration (1 to 8) - Attn bonus amounts (1 to 8). Whiskers represent the 95% confidence intervals.

[13]For the relative attentional focus measure, we chose non-parametric two-sample Wilcoxon rank-sum tests instead of t-tests. This is because normality was rejected by a Shapiro-wilk test ($p < 0.01$) for both EIIB and RI.

[14]We should note that we did not expect any differences in our reanalysis results relative to Zhu et al. (2018). We merely consider it good practice to do so in a replication study.

[15]Attention to expiration and bonus amounts are negatively correlated with one-another in EIIB (Spearman's correlation rank, $\rho = -0.25$, $p < 0.01$) and RI (Spearman's correlation rank, $\rho = -0.30$, $p < 0.01$).

[16]This mediation analysis includes a single mediator model involving a binary dependent variable. It produces

Treatments	Choice	Attention
C(Original) vs. U(Original)	0.000***	0.000***
C(Pure Replication) vs. U(Pure Replication)	0.000***	0.000***
C(Original) vs. C(Pure Replication)	0.251	0.705
U(Original) vs. U(Pure Replication)	0.440	0.079*

Table 1.3: Overview of treatment comparisons (Experiment IIB and Replication I)

Notes. * p < 0.1, ** p < 0.05, *** p < 0.01. All values reported are p-values from χ^2 tests (Choice column) and two-sample Wilcoxon rank-sum tests (Attention column).

provide clear evidence that the mere urgency effect can be replicated in the original paradigm that was used to test the effect.

1.3.3 Exploring reasons for choice

Table 1.4 provides a breakdown of five of the most common reasons provided (with *'Other'* now representing the four additional categories not mentioned) for dominated and dominating choice across the control and urgency treatments pooled together. More specifically, the data from EIIB and RI have been pooled together, such that C(Original) + C(Pure Replication) = C(Pooled), and U(Original) + U(Pure Replication) = U(Pooled).[17] Note that each dominating and dominated disaggregation across control and urgency sums to 100%. If we first examine the pooled control conditions, we can see that most subjects who made a dominating choice stated that they paid attention to bonus amounts (90.5%). The small minority of subjects who chose the dominated option were more likely to mention that they just picked the first option (42.3%) or to provide another reason (23.1%). Moving to the pooled urgency treatments, for those who picked the dominating option, we can see the equivalently high prevalence of subjects who stated that they paid attention to bonus amounts (78.2%). However, for those in the dominated choice, there is a negligible minority who stated bonus amounts (3.8%), with the highest numbers expressing that they focused on task availability (32.60%), feared waiting in the task (28.8%), or provided some other reason for choice (25.0%). Broadly speaking, these results are consistent with the attention-based account, with more people stating bonus amounts as reasons for task choice in the control and task availability in the urgency. That said, this analysis and examination of the instructions near point of decision uncovers a few points of concern, which can broadly be grouped into two categories: (i) ones that likely do not represent a confound in the original instructions (but which we consider nonideal features of it), and (ii) ones that might represent confounds in the original instructions.

There is one feature of the original instructions that we felt was nonideal, but is something that could easily be dealt with. This refers to the non-negligible number of subjects who stated that they were picking the first option. As shown in Figure 1.1, in the original experimental design, the option labelled 'Task A' was always the dominated option and ordered first on the decision screen. Research highlights a tendency for individuals to choose the first option on a list

output consistent with that from Hayes (2017). Number of bootstrap replications: 2,000.

[17] This way, we have pooled across EIIB and RI, while keeping treatments separate. The data has been pooled for simplicity. While there are some differences between EIIB and RI, the overarching insights are comparable. See Figure A4 in the Appendix for the results disaggregated across EIIB and RI for all nine categories.

	C(Pooled)		U(Pooled)	
	Choice			
Reason for choice	Dominating	Dominated	Dominating	Dominated
Bonus Amounts	229 *90.51%*	2 *7.69%*	129 *78.18%*	5 *3.79%*
Task Availability	3 *1.19%*	2 *7.69%*	13 *7.88%*	43 *32.58%*
Difficulty Differed	2 *0.79%*	4 *15.38%*	1 *0.61%*	3 *2.27%*
Fear of Waiting	3 *1.19%*	1 *3.85%*	1 *0.61%*	38 *28.79%*
First Option	1 *0.40%*	11 *42.31%*	0 *0%*	10 *7.58%*
Other	15 *5.93%*	6 *23.08%*	21 *12.73%*	33 *25%*
Total	253 *100%*	26 *100%*	165 *100%*	132 *100%*

Table 1.4: Reasons for choice (Experiment IIB and Replication I, pooled)

Notes. The data from EIIB and RI have been pooled together, such that C(Original) + C(Pure Replication) = C(Pooled), and U(Original) + U(Pure Replication) = U(Pooled). χ^2 tests for multiple independent groups conducted: For C(Pooled), dominating vs. dominated ($p < 0.01$); For U(Pooled), dominating vs. dominated ($p < 0.01$); Across C(Pooled) and U(Pooled), dominating vs. dominating ($p < 0.01$); Across C(Pooled) and U(Pooled), dominated vs. dominated ($p < 0.01$). χ^2 tests for two independent groups conducted: C(Pooled), dominating (Fear of waiting) vs. dominated (Fear of waiting) ($p = 0.28$); For U(Pooled), dominating (Fear of waiting) vs. dominated (Fear of waiting) ($p < 0.01$); For C(Pooled), dominating (First option) vs. dominated (First option) ($p < 0.01$); For U(Pooled), dominating (First option) vs. dominated (First option) ($p < 0.01$). These tests are summarised in Tables A5, A6 and A7 in the Appendix. Note that the data generating process regarding reasons for choice differed between EIIB and RI. In EIIB, we manually coded open-ended responses into a series of nine reasons for choice. In RI, subjects chose one of the nine reasons that most closely resembled their choice from a list. Results are pooled between EIIB and RI since they yield similar insights.

(Mantonakis et al., 2009; Anderson, 1973; Asch, 1946; Hogarth & Einhorn, 1992). Observing this behaviour in the current experimental environment highlights that the proportion of dominated choice may have been artificially inflated in the control and urgency conditions. While this would not necessarily constitute a confound specific to the urgency treatment, it is still something that ought to be considered. In Zhu et al.'s (2018) defence, as shown in Figure 1.1, they do explain to subjects that attributes were randomised using a 'randomiser' on screen. However, we feel that this relies too heavily on the assumption that telling subjects attributes are randomised will change their behaviour away from their tendency to choose the first option. A more robust strategy is to *actually* randomise the attributes attached to each option on screen.

There are two features that we thought may represent possible confounds in the original instructions: (i) subjects' *misunderstanding* when they can leave, and (ii) *saliency* of attributes at the moment of choice. Regarding possible misunderstanding, we were worried about subjects thinking they might have to wait for the total duration of the expiry period (even if they had finished their typing task). Nowhere in the original instructions did it explicitly let subjects know they would not need to wait for the entire duration. This may have introduced one confound in the urgency treatment since the results may reflect subjects basing their decision on the opportunity cost of choosing to wait for 5 versus 50 minutes. With this misperception, it could have been 'rational' for a subject to choose the urgent unimportant task to economise the total amount of time spent in the experiment. Regarding the saliency of the dominance, we were concerned about the relative prominence of availability versus payoffs at the moment of choice. Specifically, in the urgency treatment (see Figure 1.1), subjects first saw the payoff magnitudes of the tasks and then, on a separate screen (where their decision is made), the 5- and 50-minute task expiration (i.e., payoffs were no longer present on screen). This feature was also coupled with the fact that the 'randomiser' also split this information.[18] Therefore, there is a possibility that individuals may have pursued the dominated task not because the spurious time pressure led them to prioritise time over payoffs in allocating their attention but simply because payoff information was no longer visible at the moment of choice. Thus, the payoff information may have been forgotten (or received less attention) because it was no longer on screen, rather than because of any perception of time-scarcity.

Both of these issues surrounding misunderstanding and saliency may have 'masked' the dominated nature of the options in the experiment and are related to the general notion surrounding 'transparency of dominance' discussed in Section 1.2. Zhu et al. (2018, p.676) state that the experiment has the feature that *"...the restricted option is **clearly** dominated by the non-restricted alternative"* [emphasis added]. We bold the word *"clearly"* because the transparency of the dominance is under the assumption that the decision-maker has all the information readily available for them to make an informed choice at the moment of decision. While we agree that the urgent option is objectively better than the non-urgent option, we were not yet convinced that this feature was clear enough for subjects in the experiment.

1.4 Probing at the robustness of the effect

In this section, we move to the third phase of our sequential replication process. Here, we conducted three more replications: Replications II, III, and IV (RII, RIII, and RIV), all designed to control for the issues discussed in the previous section. The core aim of these additional replications were essentially for us to provide what we thought was the 'fairest' test of the mere urgency effect. While the core objective was similar, the methodology followed in RII and RIII was different when compared to RIV. Specifically, in RII and RIII, we purposefully made iterative changes to the original experimental instructions to identify and isolate each of the possible possible confounds suggested above.[19] In RIV, on the other hand, we employed a more holistic approach, which involved a complete rewriting of the instructions. The rationale for employing these distinct, but complementary, approaches is discussed below.

[18] Each individual screen shown in Figure 1.1 was separated by another screen that randomised attributes to do with timing and payoffs.

[19] The rationale for the iterative changes is to make the necessary adjustments to identify confounds while holding constant the experimental paradigm used to test the effect.

1.4.1 Identifying and isolating potential confounds

Replications II & III: Experimental designs

In this section, we discuss RII and RIII, where we designed replications that purposefully made iterative changes to the original experimental instructions to identify and isolate possible confounds. The designs of these replications are shown in Table 1.5. To facilitate understanding of the changes we made in each, the table should be read in conjunction with Figure 1.1, showing the original experimental instructions employed by Zhu et al. (2018). For the full set of instructions, consult Sections A.1.2 and A.1.3 in the Appendix.

In both RII and RIII, we included a second and third replication of the original instructions. The rationale behind this was that we wanted to rule out any possible variation between experimental sessions, allowing us to neatly compare our treatment manipulations against the mere urgency effect benchmark. Another benefit is that we could examine the extent to which picking the first option plays a role in the original effect size, since we randomised which task is dominated in all additional treatment manipulations. RII consisted of two additional experimental manipulations: (i) 'Leave Highlight', and (ii) 'Saliency'. Specifically, Leave Highlight points

Replications	**Changes to original EIIB design from Zhu et al. (2018)**
Replication II	See Section A.1.2 in the Appendix for instructions.
Baseline condition: Control (Replication)	Randomise order (and labels) of tasks.
Treatment 1: Urgency (Replication)	Randomise order (and labels) of tasks.
Treatment 2: Urgency (Leave Highlight)	Highlight at moment of choice that subjects can leave by stating: *'Note: the task expiry times only affect when the tasks are available for you to complete (not how long they will take), Whichever task you choose, it will only take 3-minutes to complete, and then you can move on'*; and randomise order (and labels) of tasks.
Treatment 3: Urgency (Saliency)	Place information about (i) task availability and (ii) bonus amounts on the same screen where the decision is made to make information equally salient at the moment of choice; and randomise order (and labels) of tasks.
Replication III	See Section A.1.3 in the Appendix for instructions.
Baseline condition: Control (Replication)	Randomise order (and labels) of tasks.
Treatment 1: Urgency (Replication)	Randomise order (and labels) of tasks.
Treatment 2: Urgency (Combined, Payoffs)	Combine 'Leave Highlight' and 'Saliency' together, showing task bonus amounts first and task availability second on the decision screen; and randomise order (and labels) of tasks.
Treatment 3: Urgency (Combined, Availability)	Combine 'Leave Highlight' and 'Saliency' together, showing task availability first and task bonus amounts second on the decision screen; and randomise order (and labels) of tasks.

Table 1.5: Experimental designs (Replications II and III)

Notes. All between-subject designs. Within each replication, subjects have an equal probability of assignment. Randomisation of dominated options means that Tasks A & B are the dominated option with 0.5 probability (i.e., have a bonus of $0.12 and an availability of 5 and 50 minutes in the urgency and control conditions, respectively). Replications II, III and IV are *2x5*, *2x4*, *2x2* experimental designs, respectively.

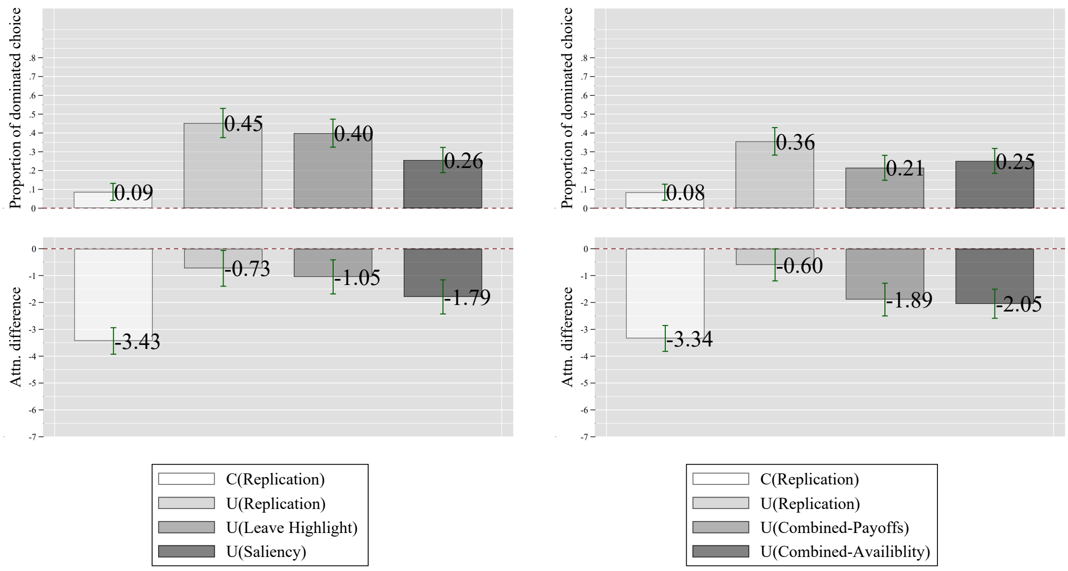
out to individuals that they can leave the experiment (i.e., they do not need to wait the full 50-minute period), and Saliency puts all information on the same screen where the decision is made (increasing the equality in salience between bonus amounts and availability). In RIII, the two additional treatment manipulations, 'Combined-Payoffs' and 'Combined-Availability', are a combination of Leave Highlight and Saliency, where either payoffs or availability are presented first within the screen, where the decision is made. The motivation behind this is that RII provides information about the individual effects of either highlighting that one can leave as well as putting information on the same screen, but not the combination of the two together. Further, there is a possibility that reversing the order of bonus amounts and availability within the decision screen may also have a notable impact on choice.

Thus, Leave Highlight, Saliency, Combined-Payoffs and Combined-Availability attempt to individually and jointly remove possible confounds from the original set of instructions, such as (i) understanding that one could leave after task completion, and (ii) the relative saliency of each attribute on the decision-screen —both of which challenge the notion that the restricted option is 'clearly' dominated by the unrestricted alternative. Therefore, if we observe that the proportion of dominated choice significantly attenuates to the same levels as the control condition in any of the treatments (Leave Highlight, Saliency, Combined-Payoffs and Combined-Availability), then this would challenge the existence of the mere urgency effect. This is because it would be explained by a simple misunderstanding that subjects think they have to wait the full 50 minutes in the task, and/or because they do not have all information available to them to make an informed choice at the moment of decision. On the other hand, if we observe that the effect survives under these treatments, then this would provide strong evidence for the existence of the effect. These replications underpin our main objective, which is to provide what we feel is the fairest test of the mere urgency effect within this experimental paradigm. The Saliency, Combined-Payoffs and Combined-Availability treatments in particular provide what we consider as 'presentational balance' and do not privilege 'payoff information' over 'timing information' (or vice versa) at the point of decision.

As previously, the two replications were conducted on MTurk (US sample) and coded on Qualtrics. We ran a single session for each in June (RII) and October (RIII) of 2019 within the original study hours and excluded individuals who participated in Zhu et al.'s (2018) EIIB or any former replications. The sample sizes were informed through a priori power analysis to obtain at-least 90% power (see Figure A1 in the Appendix): RII ($N = 808$); and RIII ($N = 647$). The average completion times were: 12.37 (RII) and 12.34 (RIII) earning $3/hr. Tables A2 and A3 in the Appendix provides summary statistics for RII and RIII, respectively.

Replications II & III: Results and discussion

Figure 1.3 provides the results for the share of dominated choice (top panel) across the treatments in RII and RIII. We also provide information for how relative attentional focus changes (bottom panel) across the treatments as this is central in Zhu et al.'s (2018) paper. That said, the focus of discussion in this section will be surrounding choice, and we will provide a more in-depth and interpretative discussion surrounding attentional focus later in Section 1.6.1. Table 1.6 accompanies these results with the p-values and their significance levels from several statistical tests. Moving to choice behaviour, we can see that in RII, the proportion of dominated choice in the control and urgency conditions are 8.7% and 45.3% respectively, confirming a second straightforward replication (after randomising which task is dominated) with the original EIIB instructions from Zhu et al. (2018). Figure A2 in the Appendix, provides the dominated choice shares disaggregated across which task was dominated, highlighting a significant order effect and underpinning the importance of randomisation. This additional replication also allows us to observe the extent to which dominated choice attenuates in the treatment manipulations where we attempt to remove confounds that might be driving up the proportion of dominated choice. In Leave Highlight, there is no significant attenuation in dominated choice after highlighting to subjects that they can leave before the 50 minutes have passed (45.3 vs. 39.9%). However, in Saliency, which places all information on the same screen, we see a significant attenuation in the share of dominated choice (45.3 vs. 25.6%). Despite this significant attenuation, the dominated choice share remains significantly higher than the control condition. In RIII, we see a third straightforward replication of the original mere urgency effect with task randomisation (8.5 vs. 35.5%).[20] In both Combined-Payoffs and Combined-Availability, a combination of highlighting that individuals can leave and putting all information on the same screen (where payoffs and availability are randomised), there is a significant attenuation in the share of dominated choice in relation to the urgency condition. That said these proportions remain significantly higher than



(a) Replication II

(b) Replication III

Figure 1.3: Dominated choice shares across conditions (RII and RIII)

Notes. Figure 1.3a refers to data from RII, while Figure 1.3b data from RIII. The top panel in both figures represents dominated choice proportions across the control condition and urgency treatment for RII and RIII. The bottom panel in both figures represents the corresponding relative attentional focus across the control condition and urgency treatments for RII and RIII. Attn. difference (−7 to +7) = Attn. expiration (1 to 8) - Attn bonus amounts (1 to 8). Whiskers represent the 95% confidence intervals.

[20]Notice that in the U(Replication), 35.5% is significantly lower than 45.3%. This heterogeneity serves to reinforce our strategy of conducting a within replication of the original instructions as a benchmark to compare the other treatment manipulations.

	Treatments	Choice	Attention
Replication II	C(Replication) vs. U(Replication)	0.000***	0.000***
	C(Replication) vs. U(Leave Highlight)	0.000***	0.000***
	C(Replication) vs. U(Saliency)	0.000***	0.001***
	U(Replication) vs. U(Leave Highlight)	0.323	0.507
	U(Replication) vs. U(Saliency)	0.000***	0.017**
	U(Leave Highlight) vs. U(Saliency)	0.006***	0.075*
Replication III	C(Replication) vs. U(Replication)	0.000***	0.000***
	C(Replication) vs. U(Combined-Payoffs)	0.001***	0.001***
	C(Replication) vs. U(Combined-Availability)	0.000***	0.001***
	U(Replication) vs. U(Combined-Payoffs)	0.006***	0.005***
	U(Replication) vs. U(Combined-Availability)	0.039**	0.008***
	U(Combined-Payoffs) vs. U(Combined-Availability)	0.441	0.714

Table 1.6: Overview of treatment comparisons (Replications II and III)

Notes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All values reported are p-values from χ^2 tests (Choice column) and two-sample Wilcoxon rank-sum tests (Attention column).

in the control condition. Further, the results in the two combined treatments indicate that there is no difference between randomising the order of payoffs and availability within screen.

For relative attentional focus, while we provide a more in interpretative discussion later, the core results are presented here. What is immediately clear in Figure 1.3 is how the relative attention measure tracks the dominated choice shares across the treatments. Specifically, the higher the share of dominated choice in a treatment, the more relative attention is placed on expiration times than bonus amounts. To illustrate this, there is significantly more attention placed on expiration time than payments in the control condition relative to any of the treatments in RII and RIII. Moreover, as the share of dominated choice decreases, relative attention shifts toward more attention on payments than expiration times. This is exemplified by the significantly lower values of relative attention in Saliency as compared with the original urgency treatment. Similarly, we see that both Combined-Payoffs and Combined-Availability exhibit lower levels of relative attention when compared to the original urgency treatments. In the same manner as before, we conduct mediation analysis, showing that relative attentional focus significantly mediates the effect of urgency on task choice.[21] These results underpin the fact that relative attention is an important correlate of choice behaviour.

Taken together, the above results show that when presenting an arguably 'fairer' test of the mere urgency effect, by highlighting to subjects that they can leave and placing all information available for rational choice at the moment of decision, although the effect attenuates, it survives. Moreover, the results also underpin that relative attentional focus is an important correlate of choice behaviour. That said, there were still other concerns that we wanted to address before concluding that the mere urgency effect is a robust phenomenon.

More specifically, we were concerned that subjects may be making inferences about differ-

[21] Mediation results: RII —Urgency ($\beta = 0.78$, $SE = 0.18$, 95% CI = $[0.47, 1.2]$), Leave Highlight ($\beta = 0.38$, $SE = 0.09$, 95% CI = $[0.23, 0.58]$), Saliency ($\beta = 0.24$, $SE = 0.07$, 95% CI = $[0.12, 0.41]$); RIII —Urgency ($\beta = 1.01$, $SE = 0.20$, 95% CI = $[0.70, 1.48]$), Combined-Payoffs ($\beta = 0.12$, $SE = 0.04$, 95% CI = $[0.06, 0.22]$), Combined-Availability ($\beta = 0.11$, $SE = 0.04$, 95% CI = $[0.05, 0.19]$)). As before, each mediation involved $2,000$ bootstrap replications.

ences across the tasks. That is, we thought that the task with less time allotted to it in the urgent treatment may be perceived as easier by subjects and therefore drive the choice behaviour.[22] The original design asks subjects to choose between two tasks, 'Task A' and 'Task B', which in and of itself implies two *distinct* tasks. However, subjects are actually choosing between an identical task, in terms of work required, that varies in attributes attached to it: (i) bonus amounts (ii) and task availability. The next concern is a return to the issue of subjects possibly misunderstanding time constraints. Previously, we attempted to highlight to individuals in Leave Highlight that they would not need to wait for the full 50-minute duration, and observed a negligible decrease in the proportion of dominated choice. While we felt the additional wording in this treatment (see Table 1.5) was clear, there were still a number of subjects who stated 'Fear of waiting' as a reason behind dominated choice in RII and RIII (see Figures A5 and A6 in the Appendix). Given this, what we felt may still be lacking is a clear distinction between 'task duration' (how long the task takes) and 'task availability' (how much time they have to complete the task in) as a salient feature of the decision-environment. Our last concern relates to the 'randomiser'. One thing that we previously discussed and did not address in RII and RIII was the idea that the randomiser itself may have been a source of confusion. For example, if subjects misunderstand and believe their choice does not make a difference because *all* attributes are randomised across each decision-screen, then this represents a possible confound.[23]

1.4.2 Rewriting instructions from scratch

Replication IV: Experimental design

Our key aim for Replication IV (RIV) was to attempt to address the concerns mentioned in the previous section. However, we had a slightly different methodological approach in this replication relative to the previous ones. That is, we employed a 'fresh eyes' approach by creating from scratch a new set of experimental instructions (for both control and urgency conditions) that did not take the original ones as their starting-point and were intended to minimise scope for possible confounds. Among the benefits we envisaged were to (i) make clearer that the typing tasks do not differ between options, (ii) reduce scope for misunderstanding of time constraints, and (iii) remove arguably redundant features of the original design. Since we decided to completely rewrite the instructions from scratch, all of these changes were essentially 'baked' into one treatment and control condition, where there is no within replication of the original mere urgency effect. The rationale behind this was that, at this stage, we were more interested in the overall mere urgency effect size, rather than individual contributions towards it. Specifically, we wanted to construct a set of experimental instructions that were our best attempt at removing any nonideal features and possible confounds that may be present in the original design. Thus, this treatment is a marked step toward the generalisation end of a spectrum away from the straight replications and iterative changes in RI, RII and RIII.

Figure 1.4 provides the experimental instructions at point of decision for RIV. For the full-set of instructions, consult Section A.1.4 in the Appendix. These instructions should be compared

[22]There is research that shows humans associate work that has a longer deadline as more difficult (Zhu et al., 2019; Goswami & Urminsky, 2015).

[23]This could be specific to the urgency treatment because it saw two attributes randomised (bonus amounts and task availability), where the control only saw one attribute randomised (bonus amounts).

against the original instructions in Figure 1.1. While, we completely re-wrote the instructions, we kept the lessons learned from the changes in RII and RIII. That is, in addition to randomising which option was dominated and putting all information available for choice at moment of decision, we broadly made three-key additional changes.[24] The first is that we changed the choice between two 'tasks' to a choice between two 'options' to highlight that subjects are in fact choosing between two options of the same underlying task, with different attributes attached to them. The second is that we wanted to make as clear as possible the distinction between 'task duration' and 'task availability'. This involved changing the concept of 'task availability' to 'when you must start'. Within the 'when you must start' information, we included the text *'Any time from now to 47 minutes'* or *'Any time from now to 2 minutes'* depending on treatment. This information along with bonus amounts was placed into a table since we felt this was the best possible way to clearly distinguish between the various attributes. We also included a short paragraph of text underneath the table that read the following: *'Note. A 3-minute countdown timer begins when you start the task, meaning that it is IMPOSSIBLE for the task to take you longer than 3-minutes. The timer will be visible to you throughout the task. Once you have finished the task, you can move on.'* The purpose of this text was to further emphasise that the task would only take 3 minutes and there was a window of time in which an individual had to complete it. The third is that we removed any mention of the randomiser to subjects while continuing to randomise which op-

(A) Replication IV (Control condition) (B) Replication IV (Urgency condition)

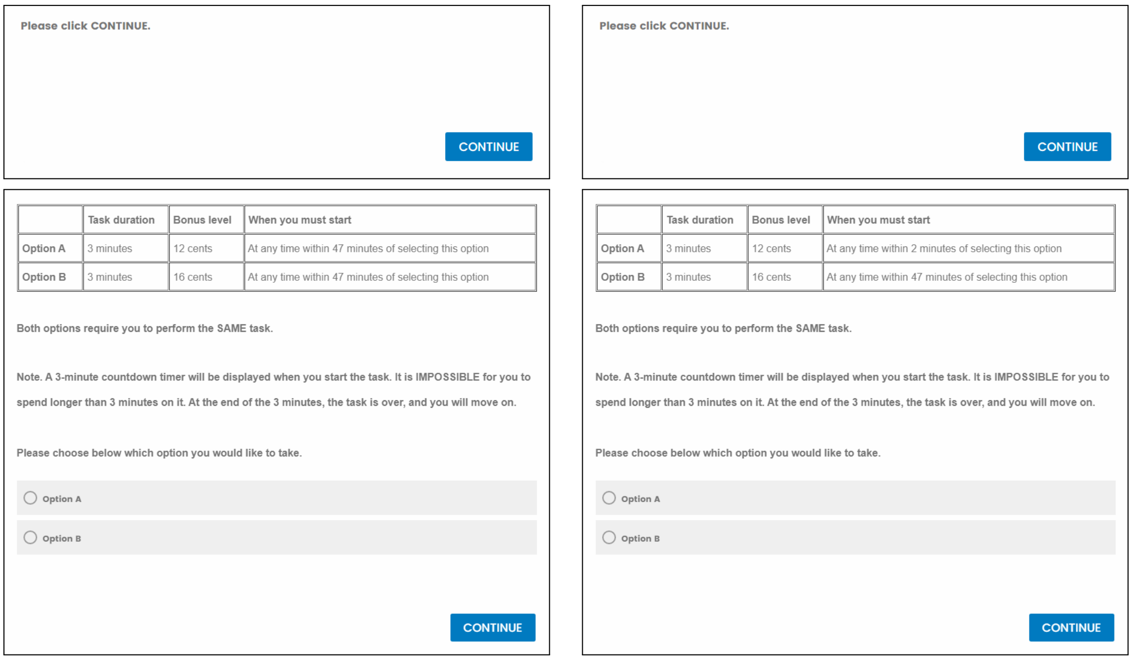


Figure 1.4: New instructions at point of decision (Replication IV)

Notes. These are the experimental instructions at the decision point for our 'Fresh-eyes' experiment (RIV). The dominated task was randomised. That is, 'Option A' was either 12 or 16 cents with 0.5 probability (with the respective start times across the two conditions), and similarly for 'Option B'.

[24]The closest direct comparison for this replication study is the Combined-Payoffs treatment implemented in RIII, since this is where we saw the least amount of dominated choice relative to the control condition.

tion was dominated to avoid any order effects. This design was our best attempt at creating the cleanest set of experimental instructions that we could. Notice that in terms of informational balance across the control and urgency conditions, the only information that changes is the 'when you must start' for the low-payoff option, which amounts to a difference of one number ('47' vs. '2'). This was intentional as we wanted to have the closest consistency between the control and urgency treatment as possible.

As previously, RIV was programmed on Qualtrics and we used subjects from MTurk (US sample). We employed all the same procedures as in previous experiments. The data was collected in January 2020, with a total sample size of $N = 315$ informed through *a-priori* power analysis (see Section A.2 in the Appendix). The average completion time was 12.77 minutes (earning $3/hr). Table A4 in the Appendix provides the summary statistics.

Replication IV: Results and discussion

Figure 1.5 provides the experimental results for RIV. Again, consistent with the mere urgency effect, there is a significantly higher proportion of dominated choices in the urgency treatment relative to the control condition with choice shares 11.3% and 35.9%, respectively (χ^2 test, $p < 0.01$). Compared against Combined-Payoffs (our treatment that saw the lowest dominated choice share) in Figure 1.3, there is a significantly higher proportion of dominated choice in the Fresh-Eyes treatment (χ^2 test, $p < 0.01$). This result is quite surprising and contrary to what we would have expected. While it is not possible to know what exactly drove this result, one possibility

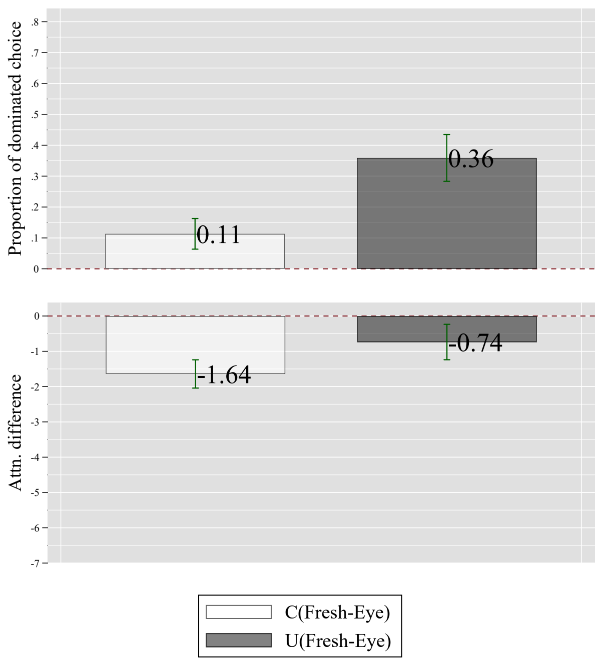


Figure 1.5: Dominated choice shares across conditions (Replication IV)

Notes. The top panel represents dominated choice proportions across the control condition and urgency treatment for RIV. The bottom panel represents the corresponding relative attentional focus across the control condition and urgency treatments for RIV. Attn. difference (-7 to $+7$) = Attn. expiration (1 to 8) - Attn bonus amounts (1 to 8). Whiskers represent the 95% confidence intervals.

is that it was simply due to random variation across experiments. Another is that the changes from 'task availability' to 'when you must start' may have increased the perceived urgency of the restricted option, since the relative time distinction went from 50 vs. 5 minutes (a ratio of 10:1) to 47 vs. 2 minutes (a ratio of 23.5:1). Nevertheless, these results do underpin the common thread that across all replications, the mere urgency remains robust. As before, we also provide the relative attentional focus, that shows subjects in the control treatment paid relatively less attention on timing than payments when compared to the urgency treatment (Wilcoxon rank-sum test, $p < 0.01$). We conduct mediation analysis, showing that relative attentional focus significantly mediates the effect of urgency on task choice.[25] We now turn to an econometric framework to investigate how the effect size of the mere urgency effect evolves throughout our replications while controlling for demographics, task ordering effects, and comprehension across all sessions.

1.5 Meta-analytic estimates of effect sizes

This section provides meta-analytic estimates of effect sizes across all data from EIIB, RI, RII, RIII, and RIV. This exercise is important as we can run several regressions to understand how the magnitude of the mere urgency effect changes across the treatments. Moreover, they allow us to control for other characteristics correlated with dominated choice. Figure 1.6 represents eleven separate marginal effects probit regressions (see Table A8 and A9 in the Appendix for the output from each individual specification for choice and attention, respectively). For ease of interpretation, the general model estimated is: $P(y = 1|x) = G(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) = G(\beta_0 + x\beta)$.[26] Where y is the dependent variable consisting of Choice (1=Dominated, 0=Dominating), and Attention (1=Attention to time > Attention to bonus, 0=Otherwise); and x represents a set of explanatory variables (i.e., treatment manipulations, (socio-)demographics, and comprehension). The probit model is estimated using maximum likelihood estimation. The figure captures how the probability of (i) making a dominated choice and (ii) relative attentional focus changes across the treatments. Note that the base categories for each specification are set to zero (by definition).

Panel A of Figure 1.6 shows that in EIIB of Zhu et al. (2018), the probability of choosing the dominated option is 43.5% higher in the urgency treatment relative to the control. Throughout the replication studies, we see a significant attenuation in the effect size (see Table A12 in the Appendix where the urgency replications are the base category), reinforcing the point that we find at least one confounding factor in the original design that significantly contributes toward the original effect size. More specifically, when placing all information to do with bonus amounts and availability on the same screen where the decision is made (which we believe to be the fairest test of the mere urgency effect), we see a significant attenuation in dominated choice (the attenuation is irrespective of the order of bonus amount and availability on the same screen). However, since all effect sizes remain significant in all specifications, the mere urgency effect remains robust to removing such factors.

In the pooled specification (where all data from the original study and subsequent replica-

[25]Mediation results: RIV —Fresh-Eye ($\beta = 0.37$, $SE = 0.15$, 95% CI = $[0.11, 0.72]$). The mediation involved 2,000 bootstrap replications.

[26]See Wooldridge (2015) for more details on the estimation technique.

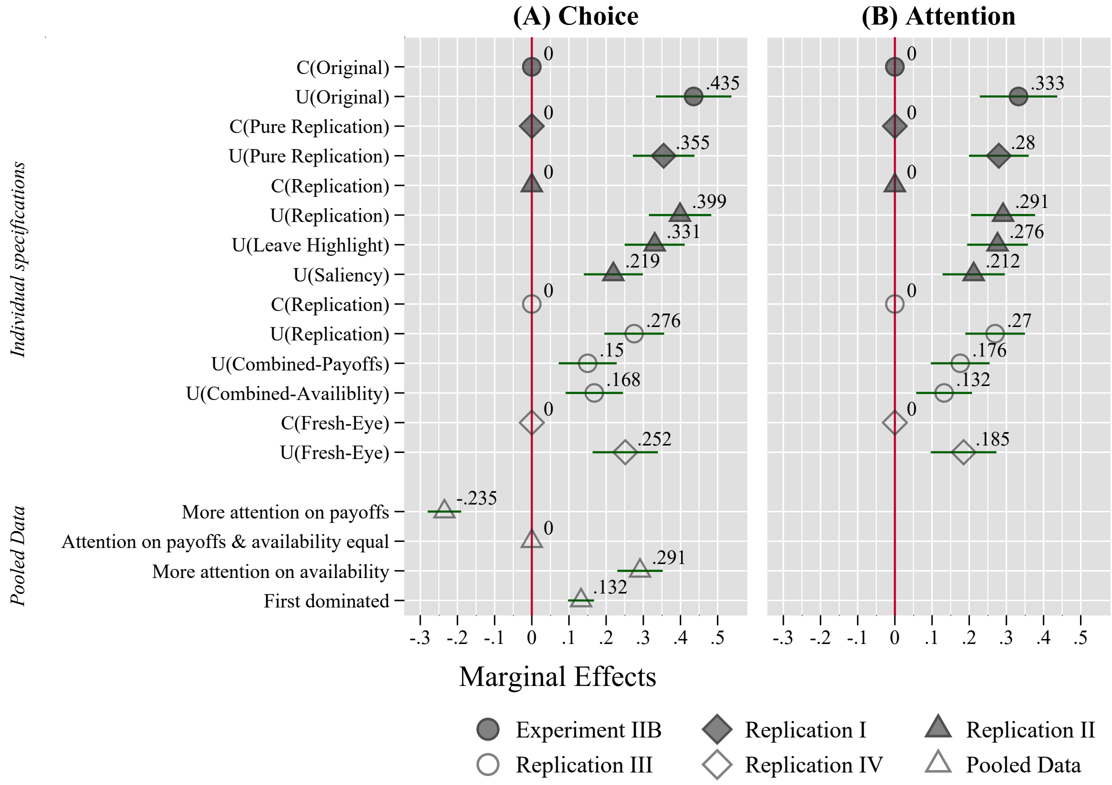


Figure 1.6: Meta-analytic estimates of effect sizes

Notes. Dependent variables: Panel A - Choice (1=Dominated, 0=Dominating), and Panel B - Attention (1=Attention to time > Attention to bonus, 0=Otherwise). Each specification refers to the urgency treatments relative to the control condition in each of the four experiments (EIIB, RI, RII, RIII and RIV). Panel A corresponds to Table A8, while Panel B corresponds to Table A9 in the Appendix. The coefficients for EIIB, RI, RII, RIII, RIV and Pooled Data correspond to separate specifications (columns) in Tables A8 and A9. For Table A8: EIIB (column 1), RI (column 2), RII (column 3), RIII (column 4), RIV (column 5), and Pooled Data (column 6). For Table A9: EIIB (column 1), RI (column 2), RII (column 3), RIII (column 4), and RIV (column 5). Additional controls in all specifications: gender, age, income, and comprehension (measured by ability to recall bonus amounts and availability). The whiskers represent the 95% confidence interval. For those interested, we also investigate differences in start times (see Table A14 in the Appendix), and correlates of choice (see Table A15 in the Appendix). The results of this ancillary analysis show no differences in start times across the treatments or choice and that gender, ability to recall bonus amounts, perceived busyness, English ability and patience are all significant correlates of choice. The busyness results align with those from Zhu et al.'s (2018) Experiment IV, where they find that perceived busyness is associated with higher rates of dominated choice.

tions is pooled), we see strong evidence for the tendency of individuals to choose the first option, highlighting the importance of randomising which option is the dominated. To elaborate, the probability of choosing the dominated option is 13.2% higher when the dominated option is displayed first. The pooled specification also sheds light on the attention-based psychological account. Specifically, if an individual stated that they paid more attention to availability than bonus amounts, they were 29.1% more likely to choose the dominated option. Conversely, if an individual stated that they paid more attention to the bonus amounts than availability, they were 23.5% less likely to choose the dominated option. The key rationale is that the limited time frame embedded in the urgency treatment increases attentional focus on the availability and leads to a subsequent urge to pursue the spuriously urgent dominated task. Therefore, it is also interesting to examine whether the urgency treatments are correlated with differing attentional focuses on availability relative to bonus amounts. Panel B of Figure 1.6 shows clear evidence that there is a higher attentional focus on time relative to payoffs in the various urgency manipulations. For example, in EIIB, an individual in the urgency treatment is 33.3% more likely to state that they paid greater attention to availability relative to bonus amounts than someone in the control condition. Notice that the magnitude of the effect sizes in both figures follow closely to one another, indicating that as one attenuates the mere urgency effect, the relative attentional focus also shifts towards bonus amounts rather than availability.[27]

As one final test of robustness, we use our measure of reasons provided for choice discussed in Section 1.3.3, and drop individuals who stated reasons that are arguably not consistent with an 'urgency effect'. Specifically, we dropped individuals who stated that: (i) the difficulty between the tasks differed, (ii) feared waiting 50 minutes, (iii) picked the first option, (iv) chose randomly, or (v) provided some other reason for choice.[28] We then conduct the same econometric analysis as before (see Table A10 and A11 in the Appendix for these estimations), and find analogous results to Figure 1.6 (albeit with smaller effect sizes).

1.6 Discussion and areas for future research

1.6.1 The attention-based psychological account

As discussed in previous sections, it is essential to provide a more interpretative discussion surrounding the attention-based account as the proposed mechanism behind the mere urgency effect. In each of our replication studies, subjects consistently pay more attention to timing than payoffs when comparing our control conditions relative to the urgency treatments. This fact is very apparent in RIV, which saw full informational balance across the two conditions, barring the spurious urgency implemented by changing one number in the instructions at the point of decision (see Figure 1.4). These results provide compelling evidence for the attention-based account as the key underlying mechanism behind choice behaviour. Nevertheless, what is curious is that attention also varies irrespective of whether we manipulate the level of *urgency* within the decision-environment. Across all our urgency treatments, we technically do not vary the level of spurious urgency (the dominated task is always characterised by 5 minutes of availability and the dominant 50 minutes). Yet, we observe that the level of relative attentional focus moves in sync with the proportion of dominated choice. These results suggest that an important consideration is not just that spurious urgency in and of itself elicits attention and drives choice, but it is also the relative prominence of it in the decision-environment that matters.

To test this, we ran an additional treatment in RII that directly set out to 'bias' attention towards payments in the decision-environment. This treatment, named 'Reversal', presented task availability first, and then on a separate screen, where the decision is made, the bonus amounts (see Section A.1.2 in the Appendix).[29] In this sense, while the level of spurious urgency was

[27]We also conduct several ordinary least squares regressions, where we interact the treatment with attentional focus. These results are provided in Table A13 in the Appendix. They show significant interaction effects between the treatments, along with significant main effects of the treatments and attentional focus.

[28]This implies that we kept individuals who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful. We kept (iii) and (iv) because these are plausibly related to a sense of urgency.

[29]This treatment serves as a quasi-replication of Zhu et al.'s (2018) Experiment III, where the authors showed that

technically the same as in the previous treatments, we privileged monetary information at the moment of choice. The results from this additional treatment show that the share of dominated choice, as well as attention, is fully attenuated to the same level as the control condition (see Figure A8, as well as Tables A8 and A9 in the Appendix). These results highlight that the way information is presented within the decision-environment is also a crucial ingredient surrounding the propensity of dominated choice. Nevertheless, these results underpin that attention is the key mechanism underlying the mere urgency effect. Specifically, there is a direct tension between the two orthogonal attributes of bonus amounts and expiration time. Subjects who are relatively more focused on temporal attributes (disregarding pecuniary ones) are likely to be associated with higher propensities of dominated choice. Thus, any decision-environment that elicits greater relative attention on attributes to do with time at the neglect of payoffs should result in higher levels of dominated choice. Since the original Zhu et al. (2018) design places information about timings at the moment of choice (with payment information no longer on screen), it is no surprise that this resulted in the highest prevalence of dominated choice. In our replications, when we provide equal weight on both attributes or more weight on payments, we see a significant attenuation in dominated choice. These results showcase that saliency of information in the decision-environment matters for rational choice, corroborating the literature surrounding dominance violations.

1.6.2 Other critiques and comments

We should also bring attention to other possible critiques and explanations of the mere urgency effect, many of which have been carefully discussed and dealt with by Zhu et al. (2018). Specifically, does the effect map out to other contexts with differing subject pools, does it persist over time, can it be attributed to a simple misunderstanding or confusion, and is it the result of a commitment device? The authors conducted several experiments addressing some of these issues. Specifically, in Experiment I, they conducted a laboratory experiment with students in a different environment with an alternative sample, finding comparable results. In Experiment IV, in a repeated choice environment, they found that it persists over time.[30] Regarding comprehension, in addition to ensuring highly attentive subjects through the ACQ and 95% HIT approval rates in all experiments, subjects were tested on experimental procedures and options at the beginning and end of the experiment. Moreover, the effect persists even after removing any alternative reasons for choice that we considered not related to an urgency effect. These reasons indicate that confusion or misunderstanding is unlikely to drive the results. Concerning the last explanation, it could be argued that individuals treat the dominated option as a commitment device, in which case it might be rational to choose it. It seems reasonable to assume that of those who need the costly commitment device, there will be some who know they need it and others who do not know they need it. If this is the case, we should observe several subjects who choose

dominated choice was significantly attenuated by reminding subjects about payoffs at the moment of choice. While this treatment is very similar, it is different in the way it makes bonus amounts salient. More specifically, we reverse the attributes on different screens, while Zhu et al. (2018) do not reverse the attributes and instead emphasise bonuses at the moment of choice with a yellow highlighter for a second time. We chose not to present this treatment in the previous results as it is in a different spirit to the others that investigate confounds within the original design.

[30] That said, these experiments are still arguably within the confines of a very specific experimental paradigm used to test the effect.

the dominating option and start after 2 minutes. However, only 13 (0.6%) individuals across all experiments started after 2 minutes. Nevertheless, we cannot rule this possibility out entirely.

1.6.3 Areas for future research

There are several exciting areas for future research. First, it would be interesting to explore the extent to which the observed behaviour might result from a 'commitment device'. Second, it would be interesting to explore how this effect maps out to contexts other than the arguably narrow experimental paradigm used to test it. This could be in the form of moving away from a one-shot dichotomous choice to a more continuous time-use case or conducting a more conceptual replication. Third, regarding the underlying mechanism of the attention-based psychological account, future research could examine how different levels of spurious urgency (while holding information on the decision-screen constant) affect relative attentional focus and choice. Relatedly, it would be helpful to measure attention more objectively through a mouse- or eye-tracking methodology (Johnson et al., 2008; Reutskaja et al., 2011; Schulte-Mecklenbeck et al., 2017). Such research would provide information about what features subjects consider when making their decisions and how long they spend considering these features. This would be a step in the right direction for a more objective test of the attention-based psychological account rather than using a subject's subjective attentional recall to motivate it.

1.7 Conclusion

This chapter investigates a novel phenomenon known as the mere urgency effect (Zhu et al., 2018), which posits that individuals are more likely to perform unimportant tasks over important tasks when unimportant tasks are characterised by spurious time pressure. In our research, we conducted four replication studies to see if the effect remains robust to a variety of changes in experimental design. We followed a three-phase sequential replication methodology: (i) re-analysed the data from the original study, (ii) re-ran the experiment following exact procedures with a new subject pool, and (iii) probed at the robustness of the effect. After four replication experiments with essentially eight new treatments aimed to replicate and stress-test the mere urgency effect ($N = 2,003$), while we observe a significant attenuation in the effect-size after ruling out hypothesised confounds, our results provide compelling evidence for the existence and robustness of it.

More specifically, the effect remains robust to (i) removing task/option-order effects (RII, RIII, RIV), (ii) placing all information relevant for informed choice on the same page where the decision is made (RII, RIII, RIV), (iii) explaining to subjects that they do not need to wait the full duration of time available to complete the task (RII, RIII, RIV), (iv) highlighting that subjects are choosing between the same task with different attributes instead (RIV), (v) ensuring distinction between task duration/availability (RIV), and lastly (vi) removing other arguably superfluous information about the decision-environment (RIV). Moreover, these results also hold econometrically, after controlling for several factors correlated with choice, and when we restrict the sample, excluding subjects who provide reasons that highlight misunderstanding or random choice. We also provide further support for the attention-based psychological account as the mechanism driving the effect.

In line with Zhu et al. (2018), we believe that the effect has important theoretical and practical implications for how choice restrictions impact consumer judgement and decision-making. An individual going through the motions of life, pursuing tasks that are characterised by a sense of urgency, may often be surprised when something important has crept up on them and slipped away.

CHAPTER 2

Time & Bracketing —evidence for suboptimal time allocation?

Abstract

Substantial literature shows the importance of *choice bracketing* to contexts that involve *interdependent* decisions. However, little is known about bracketing within the context of *time allocation*. This is surprising given the relevance of bracketing to a wide-class of time allocation problems and that such decisions are collectively amongst the most important choices an individual ever makes. In this research, we set out to answer two questions. Does narrow bracketing in time allocation exist? And if so, how robust is it? We develop a simple theoretical framework that, under certain conditions, predicts a separation between broad and narrow bracketing behaviour. We operationalise this framework in an experimental design. Across three studies, our results find surprisingly high levels of narrow bracketing and that it remains remarkably persistent. We unpick the underlying motivations for bracketing in our set-up, finding that payoff considerations are not the only factor underlying choice behaviour. Moreover, we find that comprehension, attention, cognitive ability, and certain demographics significantly correlate with suboptimal time-use.

2.1 Introduction

"The consequences of choices can rarely be fully appreciated in isolation. Even seemingly trivial decisions, such as whether or not to indulge in desert, save small amounts of money, or purchase lottery tickets, can have profound cumulative effects on our physical and material well-being. When we make choices without considering these effects, we ... make a series of local choices that each appear to be advantageous but which collectively lead to a bad global outcome."

—Read, Loewenstein, and Rabin (1999b, p.171)

A broad experimental and theoretical literature shows how choice bracketing permeates many critical economic decision environments, from decisions under risk to intertemporal choice, consumption, trade, investment, and labour supply. Read et al. (1999b) argue that broad bracketing —having an eye to the combined global consequences of a sequence of choices —generally leads to superior decision making to that of narrow bracketing —choices made with an eye to the local consequences of one or a few decisions. While research on choice bracketing illuminates the importance of bracketing in general any context that involves multiple interdependent decisions —arguably most of the choices we face in life —there is little research, to our knowledge, that investigates time allocation within the framework of choice bracketing.

This chapter and the next is based on joint work with Robin Cubitt and Chris Starmer. We would like to thank all CeDEx Brownbag members, subjects at the 2020/21 NIBS workshop as well as the 2022 CCC meeting for their helpful feedback. We also thank friends and colleagues at CeDEx, NIBS and other institutions for their invaluable insights. We express deep gratitude to NIBS for funding this research. NIBS project code: RA1175. The studies reported in this chapter and the next were approved by the Nottingham School of Economics' Research Ethics Committee.

In this research, we study the psychological sources of suboptimal time allocation. This is important since time allocation decisions are collectively among the most crucial decisions an individual ever makes. How long to stay in education, which attainable jobs to take, whether to take a career break for child care, when to retire, whether to engage in frequent exercise etc., are all decisions that have huge impacts on our material and psychological well-being —the outcomes of which can be affected by how we bracket. For example, the extent to which an individual chooses to exercise regularly might depend on their ability to recognise the accumulated benefits from running multiple times a week for an entire year (broad-frame) as opposed to the arguably trivial health benefits from a single run (narrow-frame).[1]

To illustrate more precisely the distinction between narrow and broad bracketing and their implications for time allocation, let us consider another example. Suppose one has five working days to complete a set of tasks, Monday to Friday. Broad bracketing is making one decision over the allocation of all five days. On the other hand, a simple illustrative form of narrow bracketing is making five separate decisions, one each morning, about how to spend the coming day.[2] The critical question is: Does it matter which form of bracketing is used for how the week is spent? The answer to this is not necessarily. For example, if there is only one sensible sequence to complete the tasks and they will take all week, then both decision-frames are comparable. However, they might lead to different outcomes when new information unfolds during the week, effects of cognitive depletion vary across the tasks, or decision-making is impacted by present bias more severely when multiple separate decisions are made.

Our focus in this research is motivated by a different set of considerations, analogous to a case where there are *interdependencies* between days of the week that a narrow bracketer might fail to consider, but a broad bracketer does not. To elaborate, if what is done on Monday affects how time is best spent or how productive an agent is on Friday, then broad bracketing is generally superior to narrow bracketing. This is interesting because such interdependencies across periods are pervasive in real time use problems where there is learning-by-doing or any other source of non-linearity in returns from an activity. Moreover, this problem shows an interesting parallel to portfolio allocation, where there are interactions across asset classes in risk and returns that a narrow bracketer neglects and a broad bracketer does not (Samuelson, 1963; Tversky & Kahneman, 1985; Benartzi & Thaler, 1995; Thaler et al., 1997; Gneezy & Potters, 1997).

In this research, we set out to empirically measure the extent of narrow and broad bracketing in a controlled experimental environment. As an initial step, we construct a simple theoretical framework in which different time use emerges from broad and narrow bracketed decision-making. To elaborate, we create an environment consisting of two working activities, A and B, that vary along two dimensions: (i) the time that can be spent on them and (ii) their income-generating properties. Specifically, activity A is available for a time period that is strictly less than activity B and displays constant returns to time spent working on it, while activity B displays increasing returns to time spent. The activities have the property that the initial returns

[1] This refers to the binary decision to exercise or not. If choices are made one day at a time, the expected pain of exercise might seem to outweigh the perceived health benefits —so not exercising might be viewed as the obvious choice. However, now suppose the decision to exercise is made more broadly. In that case, the expected benefits of exercising might outweigh the perceived costs, especially if these costs decrease over time as one gets fitter.

[2] Note that this is one form of narrow-bracketing. Other forms might involve making only a few decisions throughout the week or multiple decisions daily, amongst others.

to time spent on activity A are higher than activity B. However, the returns to activity B quickly surpass the returns to activity A at higher levels of time spent working. Thus, the rational allocation of time under payoff maximisation is for subjects to spend all time on activity B and none on activity A.

From this perspective, activity A can be viewed as a distractor task, to which subjects should allocate no time at all. Our theory predicts, under certain conditions, that broad bracketers spend all available time on activity B, while narrow bracketers spend as much time on activity A as is allowed. This is because narrow bracketers fail to account for the interdependencies that exist in the decision problem —that is, time spent on activity A crowds out the higher earnings that could be achieved by spending all time on activity B. One interpretation of these distinct payoff functions is that there is a learning-by-doing in activity B that is missing in activity A. The benefits of working on something in life might only be felt later. For example, it takes years to reap the benefits of playing a musical instrument, learning another language, engaging in regular exercise or growing a relationship.[3]

Our experimental investigations involve three separate studies designed to measure the extent of narrow bracketing in time-use and, conditional on observing it, testing its robustness.

Specifically, in Study 1, we operationalise the theoretical framework in an experimental design to investigate whether choice bracketing in time allocation exists. The results show surprisingly high levels of suboptimal time allocation behaviour, consistent with narrow bracketing. In line with our theoretical framework, we observe a strong separation in behaviour types, with subjects either narrow or broad bracketing, with little behaviour consistent with partial narrow bracketing. Given the significant prevalence of suboptimal time-use behaviour that we observe, in Study 2, we investigate the sensitivity of this result by manipulating the returns from the activities. The results reveal persistently high levels of behaviour consistent with narrow bracketing, providing robust evidence for poor time-use in our environment. However, they do suggest non-monetary considerations are also important. Consequently, in Study 3, we explore the boundary conditions for suboptimal behaviour. The results highlight that while financial considerations are an important determining factor of choice (individuals are lured by the higher initial returns of activity A, ignoring how this impacts their overall earnings), other non-monetary dimensions also play an essential role (i.e., the restriction placed on activity A).

Finally, to achieve maximal statistical power to investigate choice behaviour, we pool the data across all three studies and find that comprehension, attention, demographics and cognitive ability are all significant correlates of time-use. Taken together, our three studies provide strong evidence for high volumes of suboptimal time allocation behaviour. Narrow bracketing not only appears ubiquitous in time allocation decisions but is also surprisingly robust. To the best of our knowledge, these results are a first of their kind in the literature. They offer an insight into the efficiency losses that may arise when an individual fails to account for the interdependencies in their time-use decisions.

The rest of the chapter is organised as follows. Section 2.2 outlines the related literature surrounding choice bracketing. Section 2.3 presents the theoretical framework and the main

[3]Of course, since these decisions contain temporal features (often spanning many years), present-bias plays an important role. However, our experimental environment attempts to strip out such motivations, focussing on situations where decision-makers might take insufficient account of the interdependencies that arise in their choices.

theoretical results. Section 2.4 outlines the general experimental design. Sections 2.5, 2.6, and 2.7 present the results across the three studies —showing the existence and robustness of suboptimal time-use behaviour. Section 2.8 pools the data and provides an exploratory analysis into the correlates of time allocation. Section 2.9 concludes.

2.2 Choice bracketing literature

A recent literature review by Vorjohann (2020) shows impressively that there is a vast experimental and empirical literature on choice bracketing. This spans many crucial economic settings such as labour supply (Fallucchi & Kaufmann, 2021; Camerer et al., 1997), finance and investment (Barberis et al., 2006; Benartzi & Thaler, 1995; Kumar & Lim, 2008; Thaler et al., 1997; Gneezy & Potters, 1997; Choi et al., 2009; Brown et al., 2008), trade (Kahneman et al., 1990), consumption (Simonson, 1990; Simonson & Winer, 1992; Abeler & Marklein, 2017; Read & Loewenstein, 1995), choice under risk (Kahneman & Tversky, 1981; Rabin & Weizsäcker, 2009), intertemporal choice (Koch & Nafziger, 2020; Andreoni et al., 2018; Read et al., 1999a), and procrastination (König & Kleinmann, 2004; Sabini, 1982). Read et al. (1999b, p.172) who first coined the choice bracketing terminology, highlight that the phenomenon has been studied under the guise of multiple synonymous names in earlier literature:

> *"Simonson (1990) used sequential and simultaneous choice; Kahneman & Lovallo (1993) used narrow and broad decision frames; Herrnstein & Prelec (1992a,b) used isolated and distributed choice; Rachlin (1995) used decision making based on acts and patterns; and Heyman (1996) used local and overall value functions… All of these researchers have used these terms in the way that we use choice bracketing —to distinguish between choices made with an eye to the local consequences of one or a few choices (narrow bracketing), or with an eye to the global consequences of many choices (broad bracketing)."*

Highlighting the prevalence of narrow bracketing in a recent study, Ellis & Freeman (2020) employ a novel experimental design to investigate the level of bracketing behaviours across three different contexts in portfolio allocations under risk, social allocations, and induced-utility shopping experiments. Across the three settings, they find that just 0-15% of subjects are consistent with broad bracketing, with 40-43% consistent with narrow bracketing and the rest with neither. Although the aforementioned literature on choice bracketing has shown to be central to understanding a large volume of human choice, with narrow bracketing being surprisingly common, there is not yet a parsimonious theory of why people bracket their choices. Read et al. (1999b) discuss four main elements that play a role in understanding narrow bracketing behaviour: (i) cognitive capacity limitations, (ii) cognitive inertia, (iii) pre-existing heuristics, and (iv) self-control.

First, the authors argue that *cognitive limitations* are an important determinant of narrow bracketing, referring to research on attention (Kahneman, 1973), analytic processing (Simon, 1957), memory (Baddeley, 1992), and perception (Miller, 1956). As bracketing becomes broader, so the complexity of the problem increases, making it increasingly taxing for an agent whose cognitive capacity is bounded. This is in line with more recent work on choice complexity (Stracke et al., 2017) and cognitive ability (Abeler & Marklein, 2017) as factors that influence bracketing behaviour. Second, Read et al. (1999b) stress the importance of *cognitive inertia*, stating that many effects of bracketing are simply because people deal with problems in the manner presented to them. This means when choices are presented one at a time, decision-makers may be more likely to bracket narrowly. This has been illustrated in the domain of gambles (Redelmeier & Tversky, 1992; Keren & Wagenaar, 1987) and offers a possible explanation for the embedding effect (Kahneman & Knetsch, 1992).[4] Third, bracketing behaviour may also be determined by *pre-existing heuristics*. For example, it is common practice to segment the week into the 'work-week' and the 'weekend'. While this is a norm used for good reasons, Read et al. (1999b) argue that it can affect the way we bracket, drawing on their research showing how people order more unhealthy desserts when their decision is narrowly framed into workdays and weekends. Fourth, people may narrow bracket to overcome problems of *self-control* (Shefrin & Thaler, 1988; Fudenberg & Levine, 2006; Koch & Nafziger, 2016, 2020). Contrasting our set-up, this is one situation where narrow bracketing can facilitate optimal choice. For example, one might be more likely to complete 10,000 steps of walking per day when setting this as a daily goal instead of 70,000 steps for the whole week. This kind of behaviour might explain why some taxi drivers implement a daily income target (Camerer et al., 1997; Dupas et al., 2014). The rationale is that if they employ a weekly target, they might stop work early on any given day, promising themselves that they will make up for it on a following one.

As these different examples show, there are broadly two complementary theories concerning narrow bracketing. On the one hand, narrow bracketing is considered as a choice error, explained by cognitive limitations, cognitive inertia and pre-existing heuristics (Rabin & Weizsäcker, 2009). On the other hand, it is argued to be beneficial to achieve self-control, as explained by *motivational bracketing*. This term refers to a present-biased individual who may bracket their goals narrowly for motivational reasons. In several laboratory experiments, Koch & Nafziger (2019) attempt to understand correlates of narrow bracketing and investigate whether there is a parsimonious model of bracketing or whether choice error and motivational bracketing are distinct phenomena. Overall, the authors find mixed support for the unifying theory of choice bracketing and problems of self-control. They find that only a few choice bracketing phenomena examined correlate with cognitive skills, not lending credence to the cognitive capacity limitations perspective of choice bracketing. While in line with findings from Koch & Nafziger (2019), this contrasts work by Hilgers & Wibral (2014), Stracke et al. (2015), and Rabin & Weizsäcker (2009) who find evidence to support narrow bracketing being related to cognitive capacity and complexity of the decision environment. Taken together, these studies suggest that certain correlates of narrow bracketing are context dependent and require more empirical investigation.

Another behaviour that is highly relevant to our research and falls under the umbrella term of bracketing phenomena is known as *melioration* (Herrnstein & Prelec, 1991; Herrnstein et al., 1993). Herrnstein et al. (1993, p.150) define melioration as *"...the process of choosing the alternative among a set of alternatives which has the currently higher yield in utility"*. Analogous to choice bracketing, the process comes about because people often fail to consider the indirect

[4] The embedding effect refers to the finding that the willingness to pay for a particular good may vary *"...over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package"* (Kahneman & Knetsch, 1992, p.58). For example, people might be willing to pay the same price to clean one river in London instead of cleaning all the rivers in the United Kingdom (an example adapted from Read et al. (1999b)).

consequences of their actions (i.e., the consequences of current choices on future yields). Melioration can thus be thought of as involving an intrapersonal externality (or internality), occurring when a person underweighs or ignores the consequences of their decisions for themselves. Stillwell & Tunney (2009) discuss that internalities cause a general overinvestment in activities with decreasing average returns (e.g., addictive substances become less enjoyable with increased consumption), and underinvestment in activities with increasing average returns (e.g., practising a musical instrument becomes more enjoyable with increased training). Melioration is often studied with the canonical *Harvard Game* design (Herrnstein et al., 1993).[5]

In a similar vein to narrow bracketing, several attempts have been made to explain why melioration behaviour exists. Herrnstein et al. (1993) distinguish between cognitive and motivational reasons. Over several experimental manipulations, the authors find support for the cognitive perspective.[6] The more salient one makes the internality, the less likely individuals are to follow behaviour consistent with melioration (as opposed to maximisation). A wave of subsequent research has attempted to shift people away from melioration toward maximisation. Many are only successful when simplifying the decision environment and, consequently, reducing the cognitive complexity of the task —that is, providing hints, allowing for practice, or simplifying the decision problem (Herrnstein et al., 1993; Warry et al., 1999; Tunney, 2006; Herrnstein & Prelec, 1991; Stillwell & Tunney, 2009).[7] Taken together, this literature show that melioration behaviour can be overcome when decision problems are simplified, lending credence to the cognitive capacity limitations hypothesis.

While it is evident that extensive literature surrounds choice bracketing and how it permeates many decision environments, surprisingly, little research examines it within the context of allocating time. To our knowledge, this research offers the first contribution to the literature by empirically measuring bracketing behaviour and time allocation in a controlled experimental environment. While connected to the canonical melioration design, our set-up differs in several important ways. The most crucial of which is that our design is solely focussed on *real* time allocation decisions, where subjects must work on tasks to produce output and subsequent income. In fact, while Herrnstein et al. (1993) draws many parallels concerning melioration and time allocation, providing many use cases (i.e., the incessant 'workaholic'), their experimental design does not explicitly involve the allocation of a subject's 'time' while working.[8] This offers

[5]The game provides subjects with two alternatives (options A and B). The payoff from option A on any one trial is higher than the payoff from option B, but option B leads to a higher payoff in the long run. To maximise expected utility throughout the experiment, subjects need to realise that it is optimal for them to postpone a higher short-term reward (the meliorating option) in lieu of an even larger long-term reward (the maximising option). More specifically, a rational agent should choose the option that pays off the lower local amount, but higher global amount. Experimental evidence reliably finds that subjects tend to choose the option with the higher immediate payoff, despite it being the inferior choice (Stillwell & Tunney, 2009).

[6]They conduct a total of five experiments where, amongst other things, they vary the shapes of the rewards functions (parallel and crossing over), the type of reward itself (coin delay vs coin value) and the salience of the internality (averaging window, presence of an arrow, and explicit hints).

[7]Stillwell & Tunney (2009) note that the list of choice anomalies that diminish in appropriately structured learning environments is growing: preference reversals (Tunney, 2006), and problems in Bayesian reasoning (Aaron & Spivey-Knowlton, 1998; Gigerenzer & Hoffrage, 1995; Krauss & Wang, 2003; Friedman, 1998; Goodie & Fantino, 1999).

[8]Another point of distinction is that in the melioration design, subjects need to discover the incentives by exploring the options. In contrast, we describe the incentives fully before subjects begin work. As a result, the melioration design arguably conjoins the difference between short- and long-term payoff maximisation (i.e., the ability of the decision-maker to understand and act upon the internality present) with a component of learning rewards from prior experience (where these rewards interact with one another). Since we describe the rewards at the beginning and they

a significant opportunity to implement a design that explicitly targets time allocation. Moreover, we also contribute to the current discourse surrounding how bracketing behaviour correlates with specific measures surrounding cognition and demographics.

2.3 Theoretical framework

In this section, we set up a simple framework where there are marked differences between broad and narrow bracketing perspectives in time allocation. We are interested in constructing an environment that founds our experimental design, where some bracketing behaviour could take place —one in which a narrow bracketer might fail to account for the interdependencies in their choices, leading to suboptimal decisions. Thus, our model is not a fully general theory on time allocation (such as Becker (1965)) but one that provides the base for our empirical analysis of bracketing behaviour in time allocation.

2.3.1 The time allocation problem

Let there be an interval of time $[0, T]$ and two high-level activities A and B, each of which subdivides into $n \geq 1$ component activities, indexed i ($1 \leq n \leq \infty$). We take the resulting set of 2n component-activities to comprise a mutually exclusive and exhaustive set of ways of spending time, within the overall time interval. We refer to the high-level activities, denoted A and B, as *activities*, and the component activities that comprise them, denoted a_i and b_i, as *components.*[9] The overall time constraint is:

$$\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i = T$$

$$\forall i \in \{1, \ldots, n\},\ a_i \geq 0 \text{ and } b_i \geq 0$$

where a_i is time spent on component i of activity A, and b_i is time spent on component i of activity B.[10]

We now specify two features that separate activities A and B: (i) the maximal time that can be spent on them and (ii) their income-generating properties. These features are crucial for our analysis of the distinctions between narrow bracketing and broad bracketing.

i. **The maximal time that can be spent on an activity**

Although there is no upper limit on the proportion of the available time interval that may be spent on activity B, a tighter upper limit applies to activity A. Specifically, the maximum proportion of time that may be spent on activity A is μ, with $0 < \mu < 1$.

do not interact with one another, our focus is purely on the former —the extent to which people fail to account for interdependencies in their time allocation decisions.

[9] Although it would be more general to allow for more than two high-level activities or to let the number of components vary across high-level activities, we do not do this as it would complicate the theoretical expressions without adding extra insight.

[10] Non-negativity constraints are imposed because spending a negative amount of time is impossible. The summation condition follows from the mutual exclusivity and exhaustiveness properties of our specification of time use.

$$\sum_{i=1}^{n} a_i \leq \mu T$$

$$\sum_{i=1}^{n} b_i \leq T$$

In other words, of the total T, μT units of time are *unreserved* and can be spent on either activity, however, the remaining $(1 - \mu)T$ units are *reserved* for activity B.[11]

This structure provides a simple distinction that we formalise below, between a narrow bracketer —who considers reserved and unreserved time separately, ignoring any interdependencies between them; and a broad bracketer —who considers both together, allowing for any such interdependencies.

ii. **Income-generating properties**

When deciding on how to allocate time, we assume that the only objective of the agent is to maximise the income derived from the activities. This means that activities are neither valued nor disliked in themselves.[12] Income can be generated as follows. Total income (y) is the sum of incomes from the two activities (y_A and y_B, respectively). y_A depends on the output (f_A) of activity A, which depends in turn on $a_1, ...a_n$. Similarly, y_B depends on the output (f_B) of activity B, which depends in turn on $b_1, ...b_n$. We specify:

$$f_A = min\left\{\frac{ka_i}{p}, \ldots, \frac{ka_n}{p}\right\}$$

$$f_B = min\left\{\frac{kb_i}{p}, \ldots, \frac{kb_n}{p}\right\}$$

with $0 < k < \infty$ and p an integer ≥ 1

This specification implies symmetry across activities in how time in the components of the activity generates output of the activity. Moreover, it imposes two restrictions on that process. First, each component activity produces a component-output which is a multiple k of time spent in the component. The parameter k is interpretable as the agent's productivity in time use. Second, *each* unit of output in the activity requires p units of component-output from *every* component of the activity.[13] Accordingly, output generation in each activity has a form of weakest-link structure.[14]

We now specify how activity-income depends on activity-output. The key difference between activities A and B is their income generation:

[11]In Chapter 3 we introduce an experimental treatment where reserved and unreserved time are specific subintervals of the available time in the experiment. However, this is not required by the theoretical model, nor relevant to Chapter 2. In Chapter 2, time is reserved for activity B only in the sense that, at any point, if μT units of time have already been devoted to components of activity A, then the remaining time has to be used on components of B.

[12]We assume that the agent is only motivated by acquisition of money and suffers no disutility from effort or any (component or high-level) activity *per se*.

[13]To simplify the model, here we assume that k and p are constants. In Section 2.4.3, we discuss how our experimental design deals with possible heterogeneity in k across subjects, while Chapter 3 relaxes the assumption on p and n. To avoid repeating the theory later, we specify here that each is an integer ≥ 1.

[14]As an example, if activity A has two components, the lowest output on one of the components determines payoffs for activity A (and vice versa for activity B).

$$y_A = \alpha(f_A) = \alpha \; min\left\{\frac{ka_1}{p}, \ldots, \frac{ka_n}{p}\right\} = \alpha(\frac{k}{p}) min\{a_1, \ldots, a_n\}$$
$$y_B = \beta(f_B)^2 = \beta \; \left[min\left\{\frac{kb_1}{p}, \ldots, \frac{kb_n}{p}\right\}\right]^2 = \beta(\frac{k}{p})^2 \left[min\{b_1, \ldots, b_n\}\right]^2$$
$$0 < \beta < \alpha < \infty$$

That is, income from activity A is characterised by *linear returns* to output and income from B is characterised by *quadratic returns* to output.[15] As $\beta < \alpha$ this implies that at low levels of output, the marginal returns to activity A are higher than those of activity B. However, once a sufficiently high level of output is reached, the returns to activity B are higher.[16] The crucial feature of these functions is that, in terms of income generation, activity B displays a form of increasing returns to scale that is absent from activity A. As a result, time spent on activity A crowds out the *higher* earning potential of time spent on activity B, whereas time spent on activity B crowds out earning potential in activity A at a *constant* rate. We refer to this difference as an *enhanced crowding out effect* of activity A.[17]

To sum up, so far we have specified two differences between the activities: the time that can be spent on each activity (in effect, some time is reserved for activity B due to the limited time that can be spent on activity A); and that output from activity B translates into income quadratically, whereas activity A does so linearly. Given the two distinct features that separate activities A and B, how will the agent allocate their time? There are two issues: (i) how to allocate time across activities; and (ii) how to allocate it across components (of those activities in which any time is spent). Our focus is the former, but we begin with the latter as an analytical preliminary.

Within-activity time allocation

Lemma: *The optimal way to allocate time across the components of activity A (respectively activity B) is to do so equally. That is, any optimal allocation of time must satisfy the following: For all i, j ∈{1, ..., n}, (i)* $a_i = a_j$*; and (ii)* $b_i = b_j$*.*

The Lemma arises because the agent is an income-maximiser, and aggregate income is only affected by the division of time across the components of a given activity insofar as that division affects the output of the activity. Given this, the Lemma is an immediate implication of the weakest-link structure specified above and is part of our motivation for having the structure. For experimental reasons explained below, we want the within-activity allocation of time to be a real decision, but not one that competes in complexity with, or complicates, the cross-activity allocation of time, which is our focus.

---

[15] Note that the payoff functions are not the only functions that will allow us to separate behaviour types. Using a (i) linear versus convex payoff, (ii) convex versus concave, or (ii) stepwise functions yields similar results.

[16] As shown below, $f_B > \frac{\alpha}{2\beta}$ is 'high enough'.

[17] As mentioned in Section 2.1, one interpretation of the linear and quadratic forms is that there is a learning by doing in activity B that is absent from activity A. The benefits of working on something in life might only be felt later on. Some tasks might appear to have more rewards up-front, and others much later on, where the later-on benefits far outweigh short-run ones. For example, the returns to becoming a world-class pianist take years of training, and if one constantly switches different musical instruments, one will never become a Grandmaster. The same can be said for an important project. If one gets stuck on doing tons of little unimportant tasks (that might appear to have short-term benefits), then one might be doomed to never get the project completed.

Cross-activity time allocation in the face of bracketing

The Lemma greatly simplifies the overall time allocation problem by reducing it from a choice of $2n$ numbers to a choice of one number, namely the amount of time spent on (any) one component of activity B, defined as $\bar{b}$. In view of the Lemma, the choice of $\bar{b}$ also determines the total amount of time spent on activity B ($n\bar{b}$), the total amount of time spent on activity A ($T - n\bar{b}$), and therefore $\bar{a}$, which is the total amount of time spent on each component of activity A, given by $\frac{T-n\bar{b}}{n} = \bar{a}$. Crucially, all three terms are determined by $\bar{b}$. Hence, the overall optimisation problem reduces to choosing $\bar{b}$.

The Lemma leaves open how the agent allocates their time across activities, and hence, the value of $\bar{b}$. Its determination depends crucially on how the decision-maker considers the allocation of unreserved time, and hence on whether they are a broad or narrow bracketer.

Broad bracketing problem —BBP

Given the assumptions outlined above, a broad bracketer considers how to maximise overall income as a single time allocation problem. They adhere to the constraints that $\sum_{i=1}^{n} a_i \leq \mu T$ and $\sum_{i=1}^{n} b_i \leq T$, which create unreserved and reserved time, respectively, but do not separate the two types of time into two optimisation problems. By implication, they take into consideration how unreserved time is spent affects how remunerative is their reserved time for activity B.

More formally, given the Lemma above, the broad bracketing problem (BBP) is to choose $\bar{b} \in [\frac{(1-\mu)T}{n}, \frac{T}{n}]$ such that y is maximised:

$$y = \alpha(\frac{k}{p}) \left[\frac{T - n\bar{b}}{n}\right] + \beta(\frac{k}{p})^2 \bar{b}^2$$

The second derivative of the maximand is:

$$\frac{\partial y^2}{\partial \bar{b}^2} = 2\beta(\frac{k}{p})^2 > 0 \tag{2.1}$$

As the second derivative is always positive, there is no interior maximum.[18] Either $\bar{b} = \frac{(1-\mu)T}{n}$ or $\bar{b} = \frac{T}{n}$ solve the BBP. In the former case, the least amount of time possible is spent on activity B; in the latter case, the entire time is spent on activity B. Depending on parameters, either one of these could be the unique solution to the BBP.[19] When the BBP has the latter case as its corner solution, starting activity B early drives up its subsequent returns. It is crucial to this solution that doing activity B in unreserved time makes the reserved time more productive. A broad bracketer appreciates this, but an agent who thought separately about reserved and unreserved time would not.

Narrow bracketing problem —NBP

A narrow bracketer considers unreserved and reserved time separately. More specifically, given $\sum_{i=1}^{n} a_i \leq \mu T$ and $\sum_{i=1}^{n} b_i \leq T$ they consider: (i) how to allocate unreserved time (μT to maximise

[18] The first derivative is $\frac{\partial y}{\partial \bar{b}} = -\alpha(\frac{k}{p}) + 2\beta(\frac{k}{p})^2\bar{b}$.

[19] Note that the existence of corner solutions should not be surprising, given the presence of quadratic income generation in activity B.

income from unreserved time and then; (ii) how to allocate reserved time ($T - \mu T$), to maximise income from reserved time. The latter problem is trivial because the only available activity in reserved time is activity B. Therefore, the narrow bracketer's problem reduces to finding a solution for (i). Crucially, narrow bracketers simply discard what will happen in the reserved time when deciding how to divide the unreserved time between activities A and B.

A narrow bracketer concentrates on how to maximise their income from the μT units of unreserved time, taking as given their earnings from the time reserved for activity B. More formally, in view of the previous Lemma, the narrow bracketing problem (NBP) is to choose $\bar{b} \in [0, \frac{\mu T}{n}]$ to maximise y:

$$y = \alpha(\frac{k}{p}) \left[\frac{\mu T - n\bar{b}}{n} \right] + \beta(\frac{k}{p})^2 \bar{b}^2$$

Note that, although the maximand of the NBP is slightly different from that of BBP, its second derivative is unaffected by this difference and is still given by Equation 2.1. Accordingly, there is again no interior maximum and any solution must be a corner solution. There are two possibilities, depending on the parameters. Any solution is either a corner solution at $\bar{b} = 0$ or a corner solution at $\bar{b} = \frac{\mu T}{n}$. In the former, no unreserved time is spent on activity B; in the latter, all unreserved time is spent on activity B.

2.3.2 Broad and narrow bracketing juxtaposed

The key question suggested by the concepts of the last section is whether the form of bracketing adopted by the decision-maker affects how they spend time. Clearly, this is not possible for reserved time, as the constraints require at least $(1 - \mu)T$ units of time to be spent on activity B. But unreserved time could be spent differently across the two cases. We crystallise the most interesting form of this possibility in the following result.

Proposition: The following two statements are equivalent;

(i) Statement 1: $\bar{b} = \frac{T}{n}$ uniquely solves the Broad Bracket Problem (BBP), and $\bar{b} = 0$ uniquely solves the Narrow Bracket Problem (NBP).

(ii) Statement 2: $\frac{T}{n}(2 - \mu) > R > \frac{\mu T}{n}$, where $R \equiv \left(\frac{\alpha}{\beta}\right)\left(\frac{p}{k}\right)$. (Condition C*)

A formal proof is available in Section B.1.1 of the Appendix. In the main text, we provide a more informal discussion.

First, Statement 1 draws a very sharp distinction between the two types of bracketer. It implies that a broad bracketer spends all time on activity B (and none on activity A), whereas a narrow bracketer spends all unreserved time on activity A (and only spends time on activity B to the extent that they have to). Statement 2 combines into one condition (henceforth Condition C*) the conditions for each of the two clauses of Statement 1 to hold.

Second, C* can only hold when μ is strictly less than unity. This is intuitive since, without it, the distinction between broad and narrow bracketing would collapse as all time would be unreserved. The solution regarding how to allocate time would depend on parameter values, however, it would not vary across the BBP and NBP as they would be one and the same problem.

With $\mu < 1$, C* holds given certain parameter values. In that case, BBP and NBP both have corner solutions, but their corner solutions do not coincide. This is possible because the two problems optimise different maximands (all income versus income from unreserved time) over different feasible ranges of $\bar{b}$.

To see the implications of this, it is instructive to graph the maximands and their feasible ranges. One particular case is illustrated in Figure 2.1, with the two maximands for the NBP and BBP as quadratics with a shared global minimum at $\bar{b} = \frac{R}{2}$, where R is as defined in C*. The maximands are vertically separated by a constant, when $\mu < 1$. Hence, each maximisation problem must only have corner solutions. The symmetry of the quadratics also implies that each problem has a unique solution whenever one end of its feasible range is further from $\frac{R}{2}$ than the other (the unique solution being the further one). Figure 2.1 illustrates a case where the upper end of the BBS feasible range is further from $\frac{R}{2}$ than the lower end. The lower end of the NBS feasible range, by contrast, is further from $\frac{R}{2}$ than the upper end (represented by black dashed vertical lines on the maximands). These two points together imply that both inequalities in C* hold in the illustrated case.[20]

Note that even with $\mu < 1$, C* is not guaranteed to hold. One way of thinking about what is required for C* to hold is to fix all parameters except $\frac{\alpha}{\beta}$. C* cannot hold if this ratio is 'too high' because that would make activity A so remunerative relative to activity B that even a broad bracketer would devote all time to activity A. C* would also fail if $\frac{\alpha}{\beta}$ is 'too low' because that would make activity B so remunerative relative to activity A that even a narrow bracketer would devote all time to activity B. Thus, for Statement 1 to hold, $\frac{\alpha}{\beta}$ must be neither 'too high' nor 'too

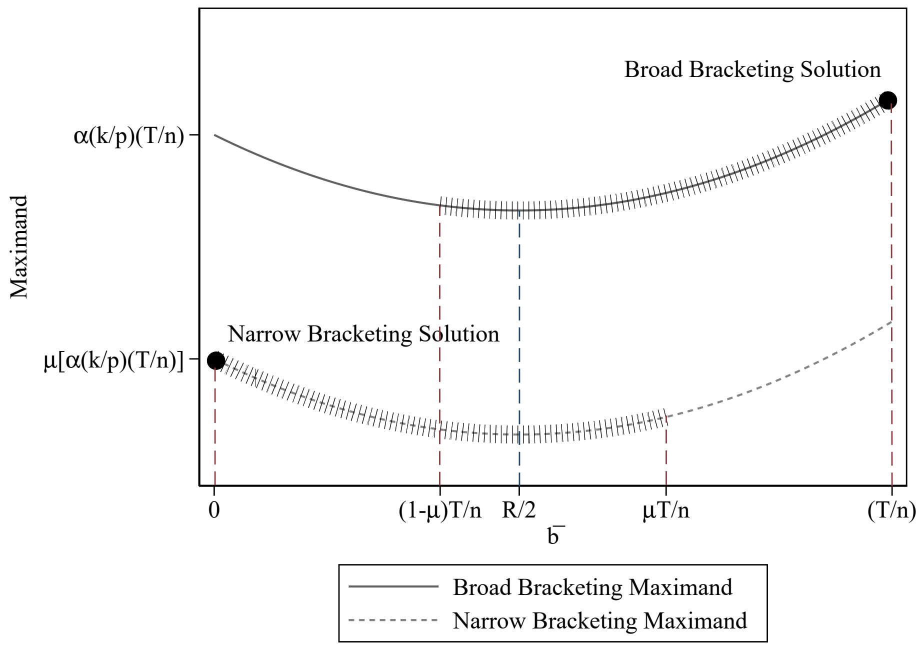


Figure 2.1: Broad and narrow bracket maximands

Notes. This figure is just one representation of the many possible maximands that can be shown. Varying the value of parameters will change the maximands and their feasible ranges.

[20] The equivalence of Statements 1 and 2 is established in Section B.1.1 of the Appendix.

low' in a sense that the inequalities of C* makes precise. Another way of thinking about C* is to fix all parameters except k. If this is 'too low', the agent is so unproductive they can never gain much from the increasing returns in activity B and even a broad bracketer would stick to activity A. If k is 'too high', the increasing returns from activity B kick-in so quickly that even a narrow bracketer would only do activity B. Thus, for Statement 1 to hold, k must be neither 'too high' nor 'too low' (again made precise by the inequalities of C*).

To sum up, our framework shows that it is possible that narrow and broad bracketing lead to different time allocations. When considering how to spend unreserved time, the narrow bracketer fails to account for something that the broad bracketers allows for, namely the impact of the use of unreserved time on the income generated from activity B in reserved time —they take insufficient account of the interdependency in the decision-problem. This factor gives the broad bracketer a reason to spend unreserved time on activity B that a narrow bracketer does not appreciate (namely, it will increase earnings generated in reserved time). When C* holds, the difference between considering and neglecting this factor drives the two types to 'opposite' corners of their feasible ranges for $\bar{b}$.

2.4 Experimental design and procedures

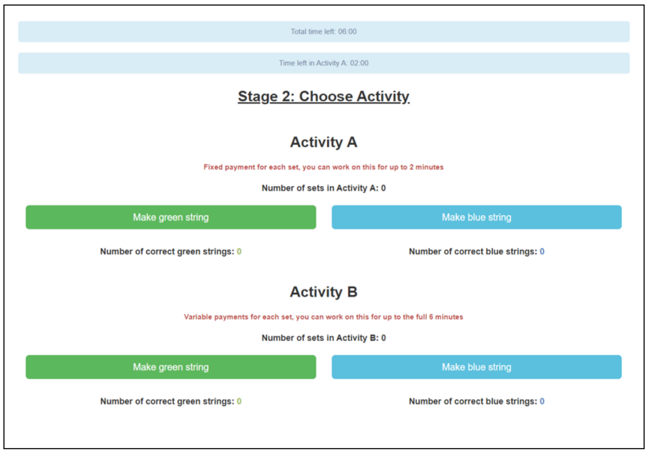
This section serves as a bridge between the theoretical framework and the experiments. In particular, we explain the basic environment that subjects face during the experiment, provide a rational for the chosen parameters, and the conditions under which C* holds. The general design and procedures remain identical across all studies, barring minor parameter tweaks that we discuss within the respective studies.

2.4.1 The experiment

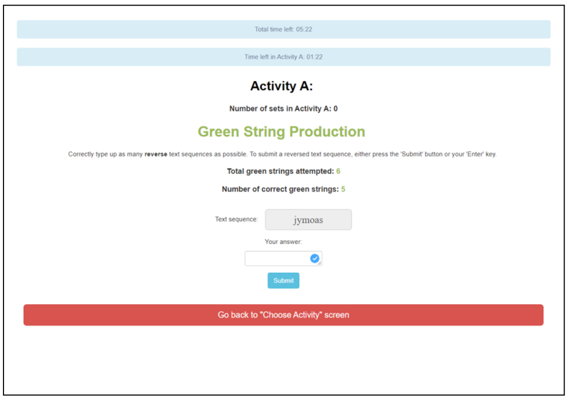
The experiment broadly consists of three stages. In *Stage 1*, subjects are introduced to the task and the experimental environment. They must answer nine comprehension questions correctly and engage in a practice round before moving on to the next stage. In *Stage 2*, they work to produce output and income. In *Stage 3*, subjects complete a battery of questions on various metrics (see Section B.2 in the Appendix for more detail).

Figure 2.2 showcases three example screenshots of the working environment that subjects face in Stage 2. Figure 2.2a represents the 'home-screen' where they must choose how to allocate their time across activity A and B. The activities each involve the *same underlying string-reversal task* but differ in their *payments* and *timing*. Specifically, the task presents subjects with sequences of randomly generated six-letter strings (e.g. 'oiggcj') and requires them to reverse-type these strings (e.g. 'jcggio'). We chose this task as it is easy to understand and unlikely to have high levels of intrinsic motivation. Moreover, in a sufficiently short time interval we expect the task to yield a predictable (ideally linear) relationship between time spent and output. For each activity, the analogue in the experiment of the theoretical framework's concept of a component (n) is the production of reversed-strings of a particular colour. In this chapter, we set $n = 2$ and operationalise it with 'green strings' and 'blue strings'.[21] We also set $p = 2$, meaning that overall

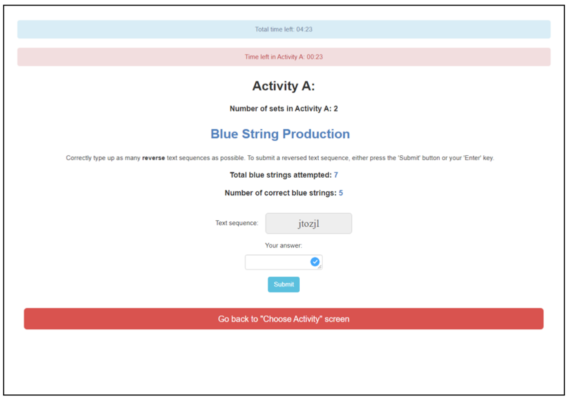
[21]It is important to note that using $n = 2$ is not a strict requirement for identifying narrow bracketing in time



(a) Choose-activity (home-screen) screen



(b) Activity A 'Make green string' screen



(c) Activity A 'Make blue string' screen

Figure 2.2: Stage 2: Task environment

Notes. This stage is where subjects work on tasks of their choice. They are free to switch between them as they please. These represent three amended screenshots of the environment. Real subjects would have seen 'Activity F' and 'Activity V', where activity 'F' and 'V' stand for 'Fixed' and 'Variable' payment to help subjects remember the differences between the two activities. In Example screen 2.2a (top), subjects are in the 'home-screen' and must choose between the component activities of A and B. Example screen 2.2b (bottom-left) provides a hypothetical instance of a subject who has clicked on 'Make green string' in activity A and has attempted several strings. Example screen 2.2c (bottom-right) is for several attempts in the 'Make blue string' component to generate sets. The experiment was programmed on LIONESS Lab (Giamattei et al., 2020). At this stage, subjects have already completed reading through instructions, answering comprehension questions and participated in a practice round where they learn how to generate string reversals.

activity-output is achieved by subjects constructing *sets* of strings. Two green strings and two blue strings together, all correctly reversed, comprise a single set in either activity A or B. For example, a subject who correctly solves five green and five blue strings in one of the activities has produced two complete sets in that activity.[22]

Subjects are paid according to the number of correctly solved sets they produce in each activity, with the following payoff functions: $y_A = \alpha(f_A) = \alpha(\# \, of \ sets \ in \ A)$ and $y_B = \beta(f_B)^2 = \beta(\# \, of \ sets \ in \ B)^2$. Payments are in the form of Experimental Currency Units (ECUs), which are

allocation. For instance, it is possible to have $n = 1$, such that the structure falls away. In this chapter, we choose $n = 2$ as this allows us to both identify bracketing behaviours, while also conveniently supporting the investigation of crucial research questions in Chapter 3.

[22]Due to the weakest-link structure, they produce the minimum of the two coloured strings rounded down. That is: $min\lfloor \frac{5}{2}, \frac{5}{2} \rfloor = 2$.

converted into a bonus at a pre-defined exchange rate (100 ECU = £1.00). We provide more detail about how these payoff functions are conveyed to the subjects below.

Subjects have a total of 360 seconds of working time. They can freely choose how to allocate their time between the two activities. However, the time that they can spend working on these activities differs. Activity A is available for *up to 120 seconds*, and activity B is available for *up to the full 360 seconds*. This implies that $\mu = \frac{1}{3}$. As illustrated in Figure 2.2a, subjects see two timers: the first governs the total time left; the other is the time left on activity A. Each timer remains fixed on the home-screen and only runs down when the subject is in one of the components. More specifically, the timer for activity A only runs down when the subject is working on activity A. Contrastingly, the total timer runs down when the subject is working on either activity A or B. This is in line with our mutually exclusive and exhaustive ways of allocating time (see Section 2.3). Figures 2.2b and 2.2c provide a hypothetical instance of a subject working.

2.4.2 Parameter choices

The time-limits

We considered 360 seconds a sufficiently short period to maintain a linear relationship between time spent and output and to avoid present-bias in evaluating strings.[23] Regarding the choice of time that can be spent on activity A, μ could not be too close to 1, as the NBP and BBP become very similar, and the adverse consequences (foregone income) from narrow bracketing are negligible in this case. On the other hand, if μ is very close to 0, almost all time is reserved, and the agent's decisions about unreserved time are inconsequential. The proportionate loss of income that a narrow-bracketer suffers relative to a broad bracketer is maximised at $\mu = 0.5$, but fairly insensitive to μ, for a typical subject, provided it remains in the range $0.25 < \mu < 0.75$ (see Section B.1.2 of the Appendix). As low values of μ increase the space of parameters for which C* holds, we selected a μ from the lower end of this range ($\mu = \frac{1}{3}$).[24]

The payoffs

A key issue when selecting payoffs is making the quadratic income returns to activity B easy to understand. We therefore set $\beta = 1$, which means that income in activity B is $y_B = (\# of\ sets\ in\ B)^2$. Thus, producing x sets in activity B earns x^2 units of experimental currency in total for that activity. This payoff structure is easy to explain to a subject: their first set in activity B earns them 1; their second 3; their third 5; their fourth 7 ECU; and so on, with each successive unit earning them 2 more than the previous one.[25] The result is that, for any number of activity B sets completed, the total income from activity B is its square, as required. In contrast, each set in activity A earns a constant α, an integer value we can set. How exactly we convey the payoffs to subjects is provided in Section B.2 of the Appendix.

[23]The potential of a present-bias is further reduced by paying subjects at the end of the experiment.

[24]It raises $2 - \mu$ and lowers μ, expanding the range of parameter values of k for which both its inequalities hold.

[25]For example, producing one set in activity B yields 1; producing two such sets yields $1 + 3 = 4$; producing three such sets yields $1 + 3 + 5 = 9$; producing four such sets yields $1 + 3 + 5 + 7 = 16$; producing five such sets yields $1 + 3 + 5 + 7 + 9 = 25$; and so on, with each increment of one set to production raising total income from activity by two more than the previous such increment did. As $x^2 - (x - 1)^2 = 2x - 1$, it follows that if, for each $x \geq 1$, the x^{th} set in activity B earns the subject $2x - 1$ units of experimental currency, achieving quadratic total returns to activity B. It does so via x separate prices for the successive sets with the property that they sum to x^2.

Given these payoffs, we can now provide an example of how narrow bracketing is seductive but suboptimal relative to broad bracketing. Assume a subject can generate a total of 6 sets of strings. Given $\mu = \frac{1}{3}$ a narrow bracketer who devotes all unreserved time to activity A, solves 2 sets in activity A, and 4 sets in activity B.[26] A broad bracketer, on the other hand, would devote all time to activity B, solving 6 sets. With the same payment structure as above and setting $\alpha = 9$ for purposes of the example, then a narrow bracketer receives the payoff vector $\{9, 9, 1, 3, 5, 7\}$ and a broad bracketer receives $\{1, 3, 5, 7, 9, 11\}$. A narrow bracketer's error occurs because they are in effect choosing between the following price vectors for the two sets producible in unreserved time: $\{9, 9\}$, $\{9, 1\}$, and $\{1, 3\}$. Viewed in this way, the first option (producing both sets in activity A) seems the more lucrative. A broad bracketer, on the other hand, is choosing the following price vectors considering the overall available time: $\{9, 9, 1, 3, 5, 7\}$, $\{9, 1, 3, 5, 7, 9\}$, and $\{1, 3, 5, 7, 9, 11\}$. Since $\{1, 3, 5, 7\}$ occur in all three of these sequences, the only difference is between $\{9, 7\}$, $\{9, 9\}$, and $\{9, 11\}$. Seen like this, it is clear that an all-activity B strategy is best. Put crudely, the narrow bracketer compares the value of two sets in activity A to the first two values achievable in activity B. In contrast, the broad bracketer compares them to the last two values achievable in activity B.

2.4.3 Conditions for C* to be met

We aimed to choose parameters such that the number of subjects for whom C* holds is maximised. One challenge is that we cannot control k, the individual string-reversal speed of a subject. Since k is unique to each subject, we cannot guarantee that C* holds for all subjects.[27] Nevertheless, we can choose parameter values following an expectation about the distribution of string-reversal performance based on the string-reversal task in Chapter 1 ($\frac{kT}{np}$). C* defines a range for $(\frac{\alpha}{\beta})(\frac{p}{k})$, governed by $\frac{T}{n}$ and μ. However, it can be rearranged to define a range for $\frac{kT}{np}$, ruled by $\frac{\alpha}{\beta}$ and μ. Multiplying C* by $\left(\frac{np}{kT}\right)\left(\frac{\beta}{\alpha}\right) > 0$, we obtain: $\left(\frac{1}{\mu}\right)\left(\frac{\alpha}{\beta}\right) > \frac{kT}{np} > \left(\frac{1}{2-\mu}\right)\left(\frac{\alpha}{\beta}\right)$. Where, $\frac{kT}{np}$ represents the total number of sets a subject would complete if they maximised their number of completed string reversals in the time-frame T. Table 2.1 summarises the parameters used for Study 1.

Parameters	**Values**
T	360 seconds
μ	$\frac{1}{3}$
Activity A work restriction	120 seconds
Activity B work restriction	360 seconds
$\frac{\alpha}{\beta}$	9
n	2
p	2

Table 2.1: Parameters in Study 1

We can plug these parameters into the reformulated version of C* to see the performance bounds in which the condition holds. We expect C* to hold for any subject whose value of k

[26] We assume consistency across output generation.

[27] Figure B1 in the Appendix shows three-dimensional graphics for how the ability to produce output in the activities interacts with earnings across different parameter levels that we employ later.

lies within $27\ sets > \frac{kT}{np} > 5\ sets$.[28] For performances that are larger than or equal to 27 sets or fewer than 6 sets, behaviour types cannot be determined, since narrow and broad bracketers are predicted to do the same thing. More specifically, the optimal time allocation for a subject who generates more than 26 sets in 360 seconds is always to spend all time on activity B. On the other hand, a subject whose performance is lower than 6 sets in the 360 seconds should always spend as much time on A as possible. Using the distribution of correct string reversals in Chapter 1, we anticipate that we can define types for at-least 70% of the sample (see Figure B3 in the Appendix). This is likely to be a lower bound, since performance in the previous chapter was unincentivised. In fact, as Section 2.5 shows, we can classify types for 95% of our sample (substantially more than previously anticipated).

2.5 Study 1: Does narrow bracketing in time exist?

Study 1 explores the extent of narrow bracketing in our experimental environment. Since broad bracketing is payoff maximising, any amount of narrow bracketing is suboptimal.[29] As our theoretical results are characterised by corner solutions, we expect to observe behaviour at either extremes of time spent on activity A, with little intermediate behaviour. While our framework provides us with tools to measure broad and narrow bracketing, it does not make explicit predictions as to the relative occurrence of types. Given the considerable amount of narrow bracketing described in previous literature, we nevertheless anticipate to find some of it in our environment.

Specific details

We conduct two experiments using the parameters specified in Table 2.1. They differ in the pricing schedule for activity B, which was extended by two sets in one study (see Section B.2 in the Appendix). However, since there is no difference in results, we pool the data and discuss them jointly. The data was collected in April 2021, with a total of $N = 83$ observations. The experiment was programmed on LIONESS Lab (Giamattei et al., 2020) and subjects were recruited via Prolific (prolific.co). The subject pool was restricted to subjects in the United Kingdom who had prior experience on the platform. Table B1 in the Appendix provides the summary statistics. Subjects took an average of 23.11 minutes to complete the study (with narrow and broad bracketers earning an average of £7.65/hr and £9.04/hr in total, respectively).

Results: Production of strings and sets

Before analysing time spent on activity A, we first look at performance in the string-reversal task, as this defines the number of subjects who fall within C^* and for which we can assign types. Figure 2.3 shows the distribution of the (i) total number of correct string reversals (top panel), (ii) total possible sets given the number of string reversals (middle panel), and (iii) the actual number of sets generated (bottom panel). $\bar{x}$ represents the mean for each distribution. In the middle panel, C^* LB and C^* UB refer to the lower- and upper-bounds constructed for C^* in Section 2.4.3. The reason why there are no bounds in the other panels is that the reformulation

[28] This is equivalent to: $\left(\frac{1}{\frac{1}{3}}\right)\left(\frac{9}{1}\right) > \frac{kT}{np} > \left(\frac{1}{2-\frac{1}{3}}\right)\left(\frac{9}{1}\right)$.

[29] As long as the string-reversal speed of a subject is sufficiently large.

of C* is based on the number of sets that would be completed ($\frac{kT}{np}$) if there were no wasted strings in the time frame (T).

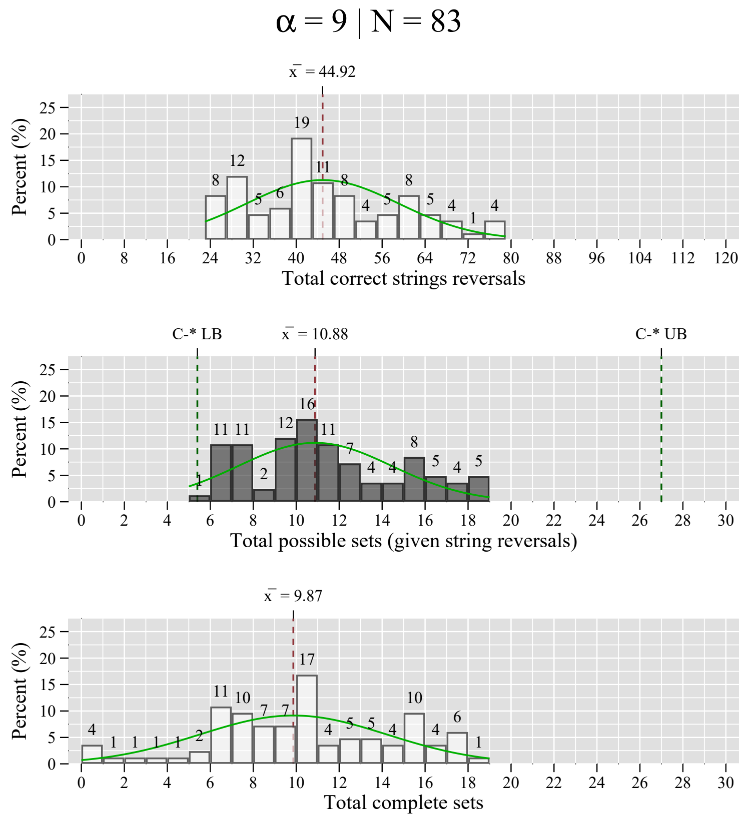


Figure 2.3: Performance for $\alpha = 9$

Notes. Only 4 subjects out of a total sample size of N=83 are outside C*(3 misunderstood how to generate a set, and 1 had a speed below C*).

Figure 2.3 shows that on average subjects complete 45 strings in the 6 minutes (top panel), translating into an average of 11 possible sets (middle panel). Only one subject has a speed that falls below the C* cut-off. Note that the number of complete sets produced depends on the ability to understand the weakest-link structure. That is, how good subjects are at splitting themselves evenly across the component activities they choose to work on. The bottom panel shows a significant downward shift in performance from number of possible sets to the sets actually produced (Wilcoxon rank-sum test, $p < 0.01$). Despite not handling the weakest-link structure perfectly, most subjects seem to understand how to generate sets and earn money from the task.[30] In particular, we can define types for 95% of our sample. Nonetheless, if we were to use sets produced, the vast majority of subjects would still be within C*, meaning that the overall results would not change drastically depending on which restriction criteria we choose to employ.[31]

[30] Only 3 subjects misunderstood the task and allocated all their time to one of the component activities, resulting in 0 sets. We exclude those observations from all future analysis.

[31] For completeness, in Section B.4.4 of the Appendix, we provide results restricted to those who fall within C* for set produced (bottom panel of Figure 2.3).

Results: Existence of narrow bracketing

Critical for defining bracketing types is the time allocated to activity A. Figure 2.4 provides a histogram of time spent on activity A (in seconds), with $\bar{x}$ representing the mean. The top panel is for the full sample, including those who misunderstood how to generate a set and the one subject whose speed was below C^*. The bottom panel is restricted to those whose speed categorisation is within C^* range (in terms of possible sets that could be produced). The results show that an overwhelming majority of subjects invested maximal amounts of time in activity A (77% for those within C^*). A Wilcoxon signed-rank test confirms that the median time spent on activity A is significantly higher than zero ($p < 0.01$), which is suboptimal under payoff-maximisation.[32]

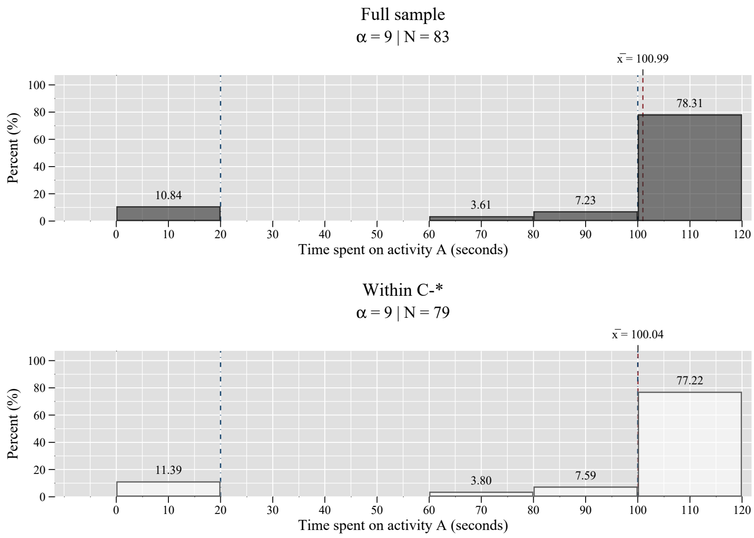


Figure 2.4: Time on activity A for $\alpha = 9$

Notes. $\bar{x}$ represents the mean. Each bin represents 20-second intervals. The left dash-dot line represents our broad bracketing cut-off (20 seconds on activity A) and the right dash-dot line our narrow bracketing cut-off (100 seconds on activity A).

Recall that for any subject whose speed puts them in the range at which C^* holds: if they are a broad bracketer, they should spend no time on activity A (0 seconds); and if they are a narrow bracketer, they should spend as much time as is allowed on activity A (120 seconds). However, in an experimental environment, we will naturally have some variability in the distribution of time spent on activities A and B.[33] Therefore, we implement a slightly more lenient classification when analysing the results. In all further analysis, we use these instead of the theoretical categories described previously. Adapting the terminology of prior literature to our environment (Barberis et al., 2006; Rabin & Weizsäcker, 2009; Ellis & Freeman, 2020), we categorise individuals into three distinct behaviour types:

i. Narrow bracketer: $Time_A \geq 100$ *seconds*

ii. Partial narrow behaviour: 20 *seconds* $< Time_A < 100$ *seconds*

[32] Given that the distribution of time spent on activity A is visibly bimodal and non-normal, we conduct non-parametric tests. Non-normality was confirmed by a Shapiro-wilk test ($p < 0.01$).

[33] Potential noise is increased through small lags between when individuals submit their choices and recording.

iii. Broad bracketer: $Time_A \leq 20$ *seconds*

The 'partial narrow bracketing' label describes subjects investing at least some time in activity A, but not all. We group it under 'narrow' bracketing because spending any time on activity A is suboptimal. Given these classifications and restricting the sample to subjects within C*, we see that 77% of behaviour is consistent with narrow bracketing, 11% with broad bracketing, and the remaining 12% with partial narrow bracketing. Study 1 thus leads to two key findings:

Result 1. *A large number of subjects choose activity A in our environment, revealing substantial levels of narrow bracketing behaviour. Suboptimal time allocation is highly prevalent within our set-up.*

Result 2. *In line with our theoretical framework, there is a large separation in types. Behaviour is consistent with either narrow or broad bracketers with little evidence of partial narrow bracketing.*[34]

Overall, the results from Study 1 show that our environment bottles a high volume of narrow-bracketing behaviour. While the prior literature discussed in Section 2.2 suggests some degree of narrow bracketing behaviour, we are surprised at the extent of it in our framework. Considering that broad bracketing in our set-up arguably does not require too much cognitive processing from the decision-maker, the high levels of suboptimal behaviour are quite notable. This is especially surprising given that our data is generated from a simple-laboratory environment, arguably devoid of the many complexities of daily life, where humans often do not have as clear an indication of output and returns to working on tasks as in our experiment. Given these striking results, a logical next step is to investigate their sensitivity.

2.6 Study 2: How robust is narrow bracketing?

We have seen that narrow-bracketing in time allocation is overwhelmingly more prevalent than broad bracketing. In Study 2, we aim to test how sensitive the result is to features of our environment and whether we can shift subjects to choosing the optimal time allocation. More specifically, we manipulate the relative price of activity A to B ($\frac{\alpha}{\beta}$). As we lower α (while keeping β fixed at 1), the loss in earnings from narrow bracketing increases, therefore making it less attractive for subjects.[35] In a subsequent data collection, we set the same parameters as outlined in Table 2.1, except we now have three piece-rate prices of activity A, namely $\alpha = \{9, 7, 5\}$. Referring back to Figure 2.1, a reduction in α lowers both intercepts of the maximands and shifts R, narrowing the gap between them. Therefore, reducing α makes it less likely the solution to the narrow and broad bracketing problem is to spend as much time on activity A as possible. Consequently, while we still predict a separation in types, we hypothesise that the level of narrow bracketing will attenuate as we lower the piece-rate price of activity A.

[34] The negligible evidence of partial narrow bracketing becomes much clearer in the subsequent studies.

[35] Keeping $\beta = 1$ means we can use an identical variable pricing schedule to the previous one to implement the quadratic returns to activity B.

Specific details

The data for this experiment was collected in May 2021. The total sample size was $N = 422$, equally split across three treatments and informed through *a-priori* power analysis (Power = 0.80). See Section B.3 in the Appendix for these calculations. Subjects were randomly allocated to one of three price conditions for α, making this a between-subject design. The experiment was programmed on LIONESS Lab (Giamattei et al., 2020). Table B2 in the Appendix provides the summary statistics across $\alpha = \{9, 7, 5\}$. Subjects were recruited via Prolific and took an average of 22.65 minutes to complete the study (with narrow and broad bracketers earning an average of £8.14/hr and £11.16/hr in total, respectively).[36]

Results: Production of strings and sets

When decreasing the fixed payments for activity A (α), the performance bounds that allow for the classification of bracketing under C* change. Across the differing levels of α, performance ($\frac{kT}{np}$) can be within the following bounds to allow for the classification of bracketing:[37]

$$\begin{aligned} \alpha = 9 &: 27\ sets > \tfrac{kT}{np} > 5\ sets \\ \alpha = 7 &: 21\ sets > \tfrac{kT}{np} > 4\ sets \\ \alpha = 5 &: 15\ sets > \tfrac{kT}{np} > 3\ sets \end{aligned}$$

The following analysis compares bracketing behaviours subject to whether C* holds for the specific treatment in question.[38] In line with C* outlined in the theoretical framework in Section 2.3, narrow bracketers should spend all time on activity A ($Time_A = 120$ seconds). Broad bracketers, by contrast should spend no time on activity A ($Time_A = 0$ seconds). For $Time_A = 120$ seconds to be the NBS, subjects need to be slow enough so that activity B does not yield a higher payoff. In other words, their string-reversal speed needs to be below an upper bound. As α falls, this upper bound must decrease, as there are now more narrow bracketers for whom activity A does not pay off. The condition thus becomes more restrictive. For $Time_A = 0\ seconds$ to be the BBS, subjects need to be fast enough (i.e., their string-reversal speed is above a lower bound). As α falls, the lower bound must decrease as well, as there are now more broad bracketers for whom activity B is optimal.

Figure 2.5 shows the performance distribution across the three levels of α. The results highlight that the distribution between the three treatments is relatively stable, with no significant differences in the means, barring weakly significant differences between $\alpha = 9$ and $\alpha = 5$ (Wilcoxon rank-sum test, $p < 0.10$).[39] As argued above, the theoretical framework implies that as the level of α decreases, the performance interval at which C* holds shrinks. This is shown in Table 2.2 where 92%, 88%, and 84% of the sample are within C* for $\alpha = \{9, 7, 5\}$ respectively. This is a result

[36] The higher average earnings in this experiment compared to Study 1 are a result of a higher fixed show-up fee.

[37] The general calculations are as follows: $\alpha = \{9, 7, 5\}$; $\left(\frac{1}{\frac{1}{3}}\right)\left(\frac{\alpha}{1}\right) > \frac{kT}{np} > \left(\frac{1}{2-\frac{1}{3}}\right)\left(\frac{\alpha}{1}\right)$

[38] Another approach would be an exclusion criterion where we analyse C* *between* treatments, therefore comparing behaviour across the lowest upper- and highest lower-bound of performance $15 > \frac{kT}{np} > 5$ for all three treatments. While this approach has the benefit of comparing evenly across treatments, such that a given subject's speed holds for all three levels of α, it restricts the sample used for analysis. Furthermore, the results between the two methods yield largely similar results.

[39] While this result may be explained by an income effect, it could also be random variation. A Kolmogorov–Smirnov test for equality of distributions shows no significant differences between the two.

of the upper bound of C* becoming more restrictive as we lower α. More explicitly, for people with a very high performance, their broad and narrow bracketing solutions are comparable (to spend no time on activity A).

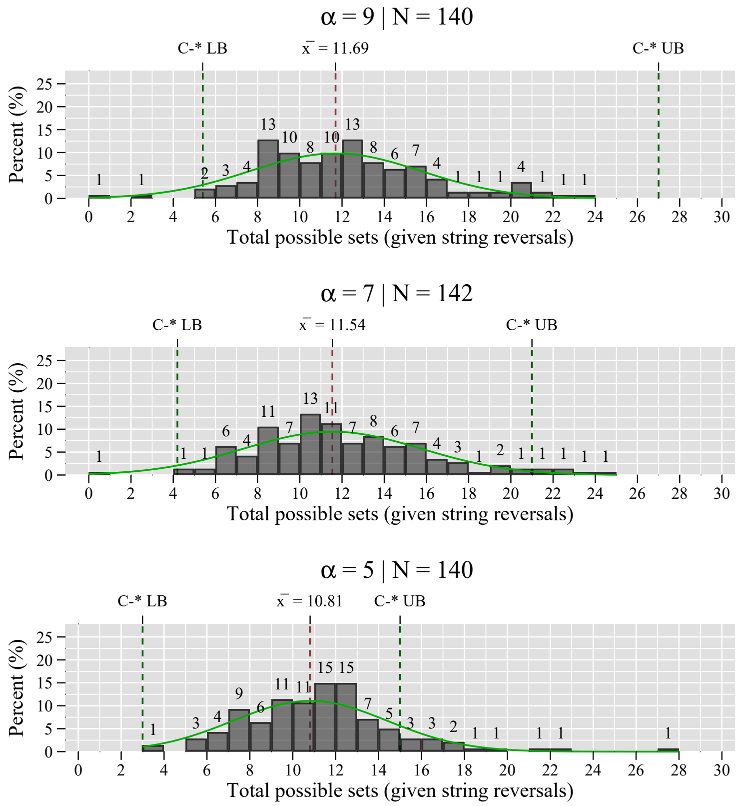


Figure 2.5: Performance for $\alpha = \{9, 7, 5\}$

Speed categorisation	$\alpha = 9$	$\alpha = 7$	$\alpha = 5$
Speed in C* range	129	125	118
	92.14%	*88.03%*	*84.29%*
Speed < C* range	3	1	2
	2.14%	*0.70%*	*1.43%*
Speed > C* range	0	6	15
	0.00%	*4.23%*	*10.71%*
Misunderstood	8	10	5
	5.71%	*7.04%*	*3.57%*
Total	140	142	140
	100.00%	*100.00%*	*100.00%*

Table 2.2: Speed categorisation for $\alpha = \{9, 7, 5\}$

Results: Robustness of narrow bracketing

Having established the bounds for which broad and narrow bracketing can be characterised, we now come to the main analysis of Study 2 and test the prevalence of narrow bracketing across treatments. Figure 2.6 shows histograms of time spent on activity A across treatments for two

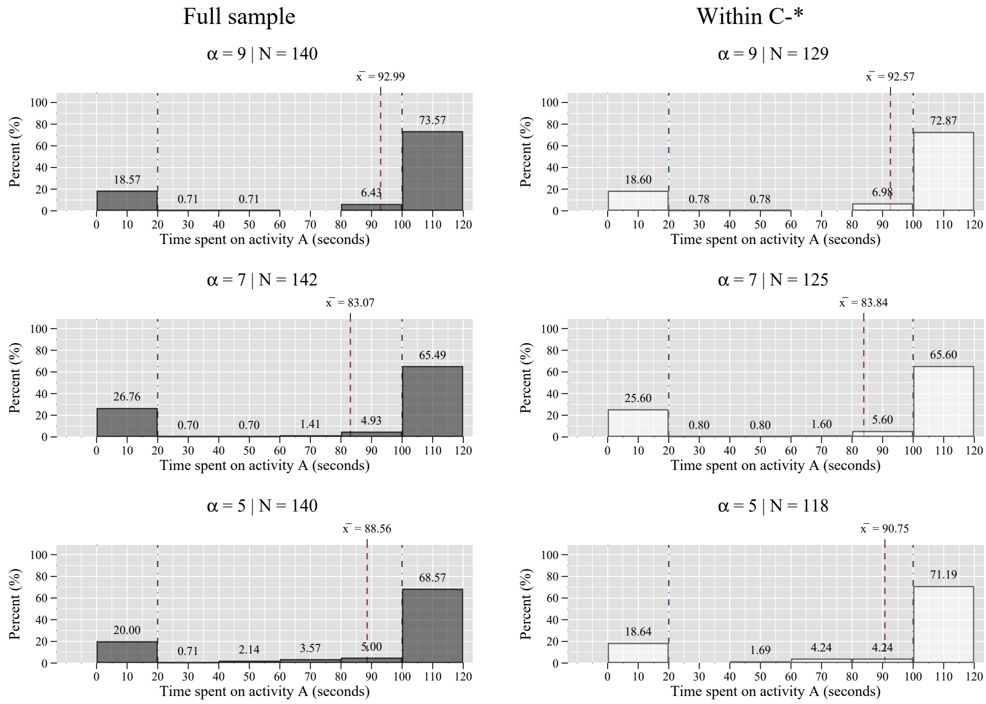


Figure 2.6: Time on activity A for $\alpha = \{9, 7, 5\}$

Notes. $\bar{x}$ represents the mean. The left vertical line represents our BB cut-off and the right vertical line our NB cut-off.

Comparisons	Full sample	Within C*
$\alpha = 9$ vs. $\alpha = 7$	0.328	0.400
$\alpha = 9$ vs. $\alpha = 5$	0.302	0.660
$\alpha = 7$ vs. $\alpha = 5$	0.928	0.640

Table 2.3: Overview of treatment comparisons for $\alpha = \{9, 7, 5\}$

Notes. * p < 0.1, ** p < 0.05, *** p < 0.01. All values reported are p-values two-sample Wilcoxon rank-sum tests. These results are also robust to other non-parametric tests, such as χ^2 tests and Kolmogorov–Smirnov tests for equality of distributions.

sample compositions: (i) full sample and (ii) within C*. Since spending no time on activity A is payoff maximising for those within C*, we test whether the median time spent on activity A is significantly greater than zero for all three conditions. Wilcoxon signed-rank tests show that the median time spent on activity A is significantly higher than zero (all $p < 0.01$). We then test differences in the distribution of time spent on activity A. As outlined above, we expect the time spent on activity A to decrease as we lower its piece-rate price. However, Wilcoxon rank-sum tests show that there is no statistical differences across treatments (see Table 2.3). This is surprising as making activity A less attractive should lower the prevalence of narrow bracketing.

Focusing on the sub-sample within C*, we find that the level of narrow bracketing is persistently high at 73%, 66%, and 71% across the three α levels. Similarly, the level of broad bracketing remains relatively low at 19%, 26%, and 19%, respectively. χ^2 tests confirm that the proportion of broad or narrow bracketers does not differ across treatments. However, it is interesting to note that there is a larger separation in behaviour types than in Study 1. Consequently, we see very little partial narrow bracketing, with just 9%, 9% and 10% across treatments, meaning that 90-91% of the data within C* are at either extreme of the time spent on activity A. This is in line with previous experimental evidence on partial narrow bracketing (Ellis & Freeman, 2020).

The results from Study 2 show that reducing α has negligible impact on the time spent on activity A and the subsequent level of narrow bracketing.[40] These results are quite unexpected, considering that losses from working on activity A are increasing significantly as we decrease the piece-rate price from working on it.[41] To give an idea as to the extent of losses, the subsequent analysis examines the efficiency losses from the persistent occurrence of spending time on activity A.

Consequentiality of losses

So far, we have established that spending time on activity A is a highly prevalent phenomenon and robust to increasing the losses from activity A. A crucial question is then: How consequential are losses from it? The simple answer is that subjects are losing a significant proportion of possible earnings by spending time in activity A. In our experiment, subjects can lose money by misallocating their (i) *time* (i.e., spending time on activity A when they should have spent it all on activity B) and (ii) *strings within activities* (i.e., not allocating their strings optimally across components). Specifically, misallocation of time is the difference between what the subject earned and what they would have earned had they optimally re-allocated the same number of sets to their optimal activity. On the other hand, misallocation of strings is the difference between what the subject earned and what they would have earned had they optimally re-allocated strings across colours within each activity. We can disentangle these two forms of losses as well as aggregate them across two types of subjects who spent: (i) less than or equal to 20 seconds on activity A ($Time_A \leq 20$ seconds), and (ii) greater than 20 seconds on activity A ($Time_A > 20$ seconds). This classification is because we observe little evidence of partial narrow bracketing.[42]

Figure 2.7 shows which percentage of earnings has been lost to misallocations of time, strings and on aggregate across treatments. The sample has been restricted to those within or above C*, since spending no time on activity A is optimal for these subjects. For those with $Time_A \leq 20$ *seconds* (top panel), there are no losses associated with the misallocation of time since these individuals are optimally generating sets in activity B. There is a small percentage of losses associated with misallocating strings, reflecting that some strings are being wasted across the two components within activity B.[43]

For those with $Time_A > 20$ *seconds* (bottom panel), losses are associated with misallocation of time are increasing across the three conditions (Wilcoxon rank-sum tests, all $p < 0.01$). In light of this, it is even more surprising that we observe such high levels of behaviour consistent with

[40] These results are also hold when running marginal effect probit regressions, controlling for demographics, performance and other correlates of choice behaviour (see Table B6 in the Appendix).

[41] Figure B6 in the Appendix provides a supplementary statistic on how close subjects are to optimal behaviour with respect to monetary maximisation. This measure is less restrictive as it includes those who fall outside C*. The results confirm that there are no treatment differences. Moreover, behaviour is far from optimal, highlighting the severity of suboptimal behaviour in our set-up as well as its apparent 'stickiness'.

[42] The results are not affected by this classification.

[43] There were two outliers for this sample that were removed, which completely skewed the results for the $\alpha = 9$ treatment. These subjects had almost all strings reversed in one activity, with a few in another activity.

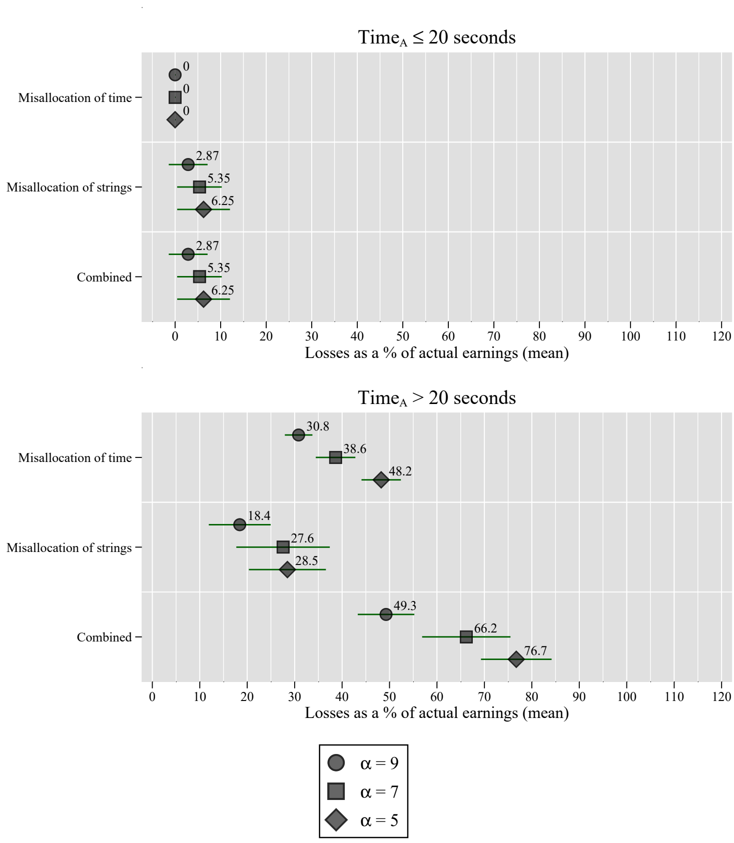


Figure 2.7: Efficiency losses for $\alpha = \{9, 7, 5\}$

Notes. Whiskers represent the 95% confidence intervals. To illustrate how these calculations are constructed, it is instructive to consider the case of a fictitious subject who spends time on activities A and B in the $\alpha = 9$ treatment when they should really only be spending time on activity B: activity A greens strings = 10 and blue strings = 9, activity B green strings = 20 and blue strings = 18, making a total of 57 string reversals. Sets produced in activity A = 4 and in activity B = 9, totalling 13 sets. The total number of possible sets is $\frac{57}{4} = 14$ (rounded down). Their earnings are $(4)(9) + 9^2 = 117$ Points (where points are exchanged at a given £ exchanged rate). Their potential earnings if they correctly allocated their 13 sets on activity B = $13^2 = 169$ Points. Their potential earnings if they correctly allocated their possible 14 sets on activity B = $14^2 = 196$ Points. Therefore, their total losses (combined) = $(\frac{196-117}{117})(100) = 67\%$. Their losses from misallocating their 13 sets = $(\frac{169-117}{117})(100) = 44\%$. Their losses from misallocating their 57 strings (14 possible sets) = $(\frac{196-169}{117})(100) = 23\%$.

narrow bracketing across the treatments. Regarding losses associated with the misallocation of strings, these are significantly higher than for broad bracketers (Wilcoxon rank-sum tests, all $p < 0.01$). While this result might indicate that these subjects do not understand the weakest-link structure as well, it may also reflect that it is more challenging to allocate strings optimally across four components instead of two (another perverse consequence of working on activity A).

The results show that narrow bracketers lose a substantial amount of their bonus by misallocating their time and strings (with average combined losses corresponding to 49%, 66% and 77% of possible earnings in $\alpha = \{9, 7, 5\}$, respectively). The persistently high levels of narrow bracketing, show that a large share of subjects are still willing to work on activity A, suggesting that there might also be non-monetary considerations when deciding how to allocate time in our set-up.

Result 3. *Reducing the returns from activity A does not significantly influence how much time subjects spend on it —even though subjects forego a substantial amount of income by doing so.*

2.7 Study 3: Pushing boundaries of the parameters

Study 2 decreased the relative price of activity A to dissuade subjects from spending time on it. The results show a persistently high level of activity A choice. While the strength and apparent robustness of this result highlights significant suboptimal behaviour, it does indicate that pecuniary factors are not the only consideration driving choice. In Study 3, we explore this by attempting to understand the boundary conditions at which suboptimal choice exists. Across Studies 3a and 3b, we sequentially implement four additional treatments that set (i) extremely low piece-rate prices of activity A, and (ii) completely remove the time restriction placed on it.

2.7.1 Study 3a: Extreme piece-rate manipulations

To investigate the role that pecuniary factors play in choice behaviour, we implement another two treatments, where the price of activity A is set to extremely low values: $\alpha = \{1.5, 0.5\}$. The rationale for $\alpha = 1.5$, is that no one who constructs more than one set should spend any time on activity A. However, since the piece-rate price of activity A (1.5) is still higher than the first set from activity B (1) this could be seen as an example of what we refer to as *ultra-narrow bracketing* and is consistent with melioration (Herrnstein & Prelec, 1991; Herrnstein et al., 1993). This would mean that subjects only compare each set of activity A with the first set generated in activity B with no foresight at all. A subject who is a meliorator, and only considers the best way of allocating their next set, will always spend as much time on activity A as possible. Their lack of foresight, however, needs to be immense as anyone who realises they will get onto a second set should see that they should invest all their time in activity B. On the other hand, if $\alpha = 0.5$, producing even one set in activity A is strictly dominated by producing one set in activity B. That is, the piece-rate returns to activity A (0.5) are always lower than the lowest piece-rate return from activity B (1). Therefore, if we observe a considerable number of subjects opting to spend time on activity A under this payoff scheme, then there must be other mechanisms in choice bracketing that dominate payoff considerations.

At this point, it is important to note that these treatments depart from our theory of broad and narrow bracketing behaviour. More specifically, with $\alpha = 1.5$, only exceptionally slow subjects generating one set could find it optimal to spend time on activity A; and with $\alpha = 0.5$ no subject should find it optimal to spend any time on activity A. Thus, only a minority of subjects could technically meet C* in the former case and none in the latter.

Procedures and results

The data for these treatments were collected in July 2021, with sample sizes of $N = 59$ and $N = 57$, respectively.[44] The experiment was programmed on LIONESS Lab (Giamattei et al., 2020). Table

[44] The sample size for these conditions and following ones is smaller because we are predominantly interested in understanding when subjects stop spending time on activity A. For testing whether the effect size is larger than zero, we need relatively small sample sizes.

B3 in the Appendix provides the summary statistics across $\alpha = \{1.5, 0.5\}$. Subjects were recruited via Prolific and took an average of 20.64 minutes to complete the study (with $Time_A > 20$ *seconds* and $Time_A \leq 20$ *seconds* earning an average of £8.43/hr and £12.57/hr in total, respectively).[45]

As before, our analysis focuses on time allocation. Figure 2.8 shows a histogram of the time spent on activity A. Since we drop the requirement of meeting C^*, the analysis is based on the full sample, excluding two subjects who failed to generate any sets.[46] The key question that we are interested in is whether extremely low levels of α are sufficient to stop subjects from choosing activity A. Although we can see that there are a number of subjects who chose to spend no time on activity A (35% in $\alpha = 1.5$ and 38% in $\alpha = 0.5$), we still see that the majority of subjects spend the maximal amount of time on it. Unsurprisingly, the median time spent on activity A is significantly higher than zero for both treatments (Wilcoxon rank tests, all $p < 0.01$). To investigate whether there is evidence of ultra-narrow bracketing, consistent with melioration, we test whether time spent on activity A in $\alpha = 1.5$ is significantly different to $\alpha = 0.5$. A Wilcoxon rank-sum test shows no evidence to support this hypothesis.[47]

We, therefore, pool the results from $\alpha = \{1.5, 0.5\}$ and compare them to the pooled results for $\alpha = \{9, 7, 5\}$. Compared to the higher piece rates, $\alpha = \{1.5, 0.5\}$ lead to a significant attenuation in time spent on activity A (Wilcoxon rank tests, all $p < 0.01$).[48] This result is important for two reasons. First, the significant attenuation in narrow bracketing highlights that our experimental

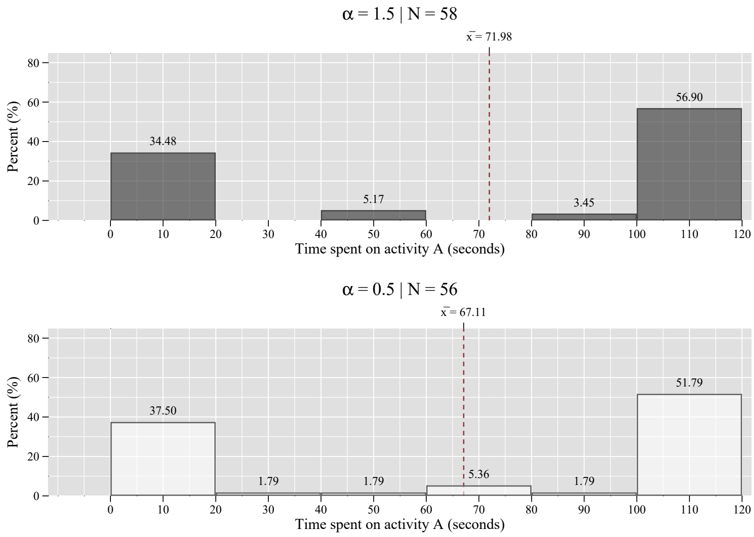


Figure 2.8: Time on activity A for $\alpha = \{1.5, 0.5\}$

Notes. $\bar{x}$ represents the mean. There is no range for C^*, because we are no longer considering bracketing behaviours. Instead, we are looking at the prevalence of suboptimal choice.

[45] Average earnings are higher in these sets of experiments because there is more optimal behaviour, barring subjects who fail to understand how to generate sets ($n = 2$).

[46] There were no subjects whose speed was so slow that they would not be able to generate more than one set.

[47] The negligible differences between $\alpha = 1.5$ and $\alpha = 0.5$ indicates minimal ultra-narrow bracketing (as we have defined above). However, this does not rule out 'ultra' narrow bracketing on more than one set (i.e., considering the earnings from the first two or more sets produced).

[48] We have pooled the data across $\alpha = \{1.5, 0.5\}$ to increase statistical power in the tests and because of the negligible statistical differences across the treatments. However, it should be noted that there is a significant attenuation in time spent on activity A when comparing $\alpha = \{9, 7, 5\}$ and $\alpha = 0.5$ (Wilcoxon rank test, all $p < 0.05$).

environment is capturing the bracketing behaviour of some subjects, as these subjects are clearly responsive to the changes in the price of activity A. Second, even though there is an increase in optimal behaviour, some subjects still spend their time sub-optimally. This behaviour cannot be explained by narrow bracketing, as there is no financial benefit to spending any time on activity A, particularly in the $\alpha = 0.5$ sample. This suggests that in addition to the bracketing behaviour observed, there must be other non-bracketing related factors influencing the choice of activity A, which we will explore in the following section.

Result 4. *Subjects are sensitive to extremely low prices of activity A. They spend less time on it when its piece-rate price becomes extremely low. However, there are still many subjects who choose to spend the maximal amount of time on activity A, suggesting that also non-monetary factors affect time allocation.*

2.7.2 Study 3b: Restriction effects

We have seen that while pecuniary factors play a significant role in determining bracketing behaviour, they cannot represent the full picture behind suboptimal choice. In Study 3b, we explore what happens if we lift the time restriction placed on activity A. Recall that for us to disentangle narrow bracketing behaviour from broad, C* needs to be met. While this involves a careful 'balancing act' of all parameters in our set-up, one key feature is that μ needs to be within a specific range ($0.25 < \mu < 0.75$) for bracketing types to be distinguishable from one another and for losses from narrow bracketing to be sufficiently consequential. Thus, while this parameter is part and parcel of the narrow bracketing problem that we have defined, we cannot currently disentangle the extent to which the restriction placed on activity A is driving choice behaviour. Given that there is a fairly vast array of research showing that people perceive things as more attractive the scarcer they are (Brock, 1968; Lynn, 1991; Verhallen & Robben, 1994; Gierl & Huettl, 2010; Worchel et al., 1975), it seems a reasonable assumption that one key mechanism underlying bracketing behaviour in our set-up is the pursuit of activity A merely because it is made artificially scarce. To investigate a *restriction effect*, we implement another two treatments that completely remove the time restriction placed on it. These treatments are named (i) $\mu = 1$ and (ii) $\mu = 1$ (Highlight), which we describe in detail below.[49]

In the previous studies, activity A could only be worked on for up to 120 seconds and activity B for up to 360 seconds. In Study 3b, by contrast, activities A and B can be worked on for up to 360 seconds. Moreover, to rule out monetary factors that explain activity A choice, we set $\alpha = 0.5$. This means that the previous $\alpha = 0.5$ results provide a clean benchmark through which we can investigate a restriction effect. Note that similar to the treatments in Study 3a, Study 3b departs from narrow and broad bracketing.[50] This is because with $\mu = 1$ and $\alpha = 0.5$, the optimal solution is to spend all time on activity B (even an ultra-narrow bracketer would do this).

[49]Note that when we state to a *restriction effect*, we are referring to a scarcity effect and not an *urgency effect*. This is because time on activity A can be spent at any point in the 6 minutes. That said, it is still possible to have an urgency effect if subjects perceive activity A to be more urgent (more on this below).

[50]With $\mu = 1$, C* cannot hold and there is no distinction between narrow and broad bracketing in our theoretical framework. Either corner solution could be optimal; and which one it is will depend on the subject's speed (k) and on $\frac{\alpha}{\beta}$.

The difference between $\mu = 1$ and $\mu = 1$ (Highlight) is that the latter highlights to subjects that they do not need to work on both activities. The exact additional wording is the following: *'Important note: You do not have to work on both activities. How much time you spend on each is your decision, however, this decision will affect your payment. Pay attention to the payment rules above.'* The rationale for the highlight is simply to emphasise that corner-solution behaviour should be considered rather than leaving it to subjects to work out. While testing an additional mechanism, since we find no significant differences across the two treatments, we use both to discuss removing the time restriction on activity A.

Procedures and results

These treatments were run separately in September 2021, with sample sizes of $N = 58$ and $N = 62$, respectively. The experiment was programmed on LIONESS Lab (Giamattei et al., 2020). Table B4 in the Appendix provides the summary statistics across $\mu = 1$ and $\mu = 1$ (Highlight). Subjects were recruited via Prolific and took an average of 20.92 minutes to complete the study (with $Time_A > 20$ *seconds* and $Time_A \leq 20$ *seconds* earning an average of £7.58/hr and £11.88/hr in total, respectively).

Figure 2.9 shows a histogram of time spent on activity A, with $\bar{x}$ representing the mean. The *x-axis* now runs from 0 to 360 seconds on activity A, as it was available for the entire time. Wilcoxon signed-rank tests show that the median time spent on activity A is still significantly higher than zero for both treatments (all $p < 0.01$). When comparing the share of optimal choices

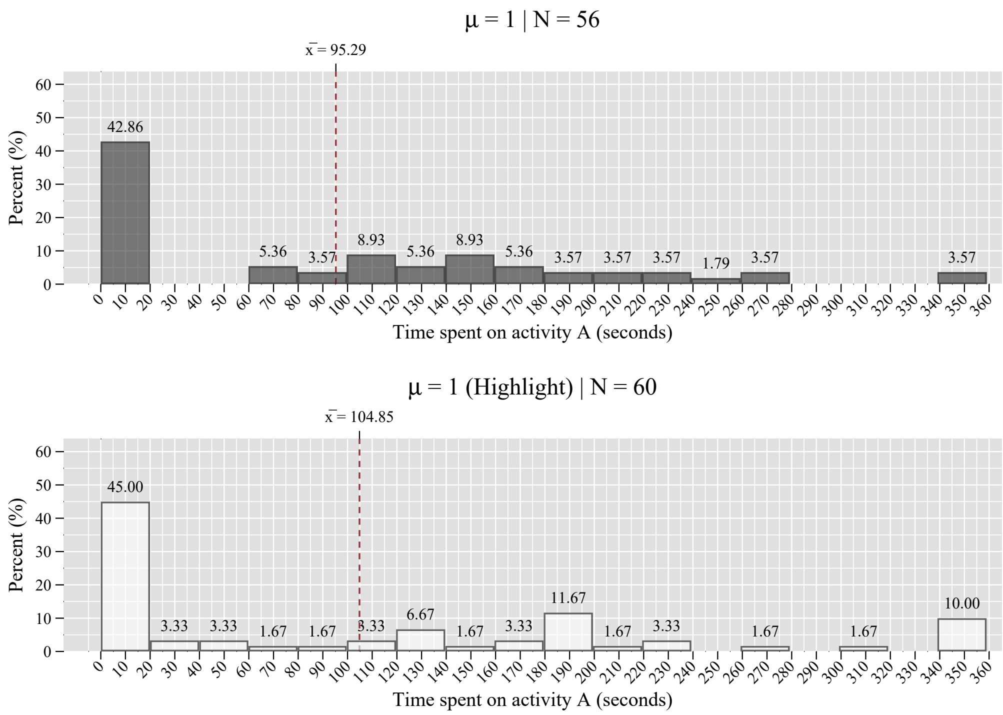


Figure 2.9: Time on activity A for $\mu = 1$

Notes. $\bar{x}$ represents the mean. There is no range for C* since we have moved away from the concept of bracketing as outlined in our theoretical framework.

in the new treatments to the ones in the $\alpha = \{1.5, 0.5\}$ conditions, there is a small, though statistically insignificant, increase in the proportion of subjects choosing to spend no time on activity A (43% and 45% for $\mu = 1$ and $\mu = 1$ (Highlight) relative to 34% and 38% in the $\alpha = \{1.5, 0.5\}$ conditions, respectively).

A part of this non-significance may be driven by a lack of statistical power. However, there is also some peculiar behaviour that we see for the subjects who spent 360 seconds on activity A. Recall that in the $\mu = 1$ treatments the lowest piece-rate price of activity B is higher than any piece-rate earnings one can obtain from activity A. It thus makes no sense to spend any time on activity A. If we exclude these individuals from the analysis on the grounds of misunderstanding the environment, then the proportion of subjects choosing to spend no time on activity A increases to 44% and 50% for $\mu = 1$ and $\mu = 1$ (Highlight) respectively (see Figure B7 in the Appendix). An independent χ^2 test now shows that these proportions are weakly statistically different from the $\alpha = \{1.5, 0.5\}$ conditions ($p < 0.10$).[51]

Result 5. *While there is still a significant number of subjects choosing to allocate their time suboptimally, removing the time restriction on activity A decreases the proportion of subjects choosing to spend time on it.*[52]

2.8 Pooled analysis across studies: Correlates of suboptimal choice

In this section, we pool the data across Studies 1, 2 and 3 to further explore different factors that might explain activity A choice and test whether our results are robust after controlling for potential correlates of suboptimal choice. In particular, we examine (i) comprehension of the instructions, (ii) relative attentional focus, as well as (iii) demographics and measures of cognitive ability. [53]

(i) Comprehension

Despite making activity A monetarily strictly dominated and removing the time restriction placed on it, we still observe that the overall level of suboptimal behaviour remains determinedly high under $\mu = 1$ treatments. A natural next step is to investigate the extent to which choice behaviour is correlated with comprehension. Subjects need to answer a total of nine comprehension questions to start the experiment (see Sections B.2 in the Appendix). If they do not get all questions correct in three attempts, they are excluded from the study.[54] Our measure of comprehension thus runs from $0 - 3$ mistakes. We distinguish between subjects with a 'good' comprehension (0 or 1 mistakes) and a 'bad' comprehension (2 or 3 mistakes). These categories are in line with a median split.[55]

[51]We compare the pooled data across the two treatments to increase statistical power.

[52]Although this is currently not a unanimous result, in the section that follows, when controlling for comprehension, we will see that it is.

[53]In Section B.4.6 of the Appendix, we also explore the role that two factors: (i) *activity ordering* and (ii) *expectations about performance* play with respect to choice behaviour. While these are two features that one might initially expect to correlate with choice behaviour, they ultimately have no impact on choice.

[54]This happened for 10% of the subjects who participated in our experiment.

[55]Across all sessions, 54% of the sample made 2-3 mistakes, while 46% made 0-1 mistakes.

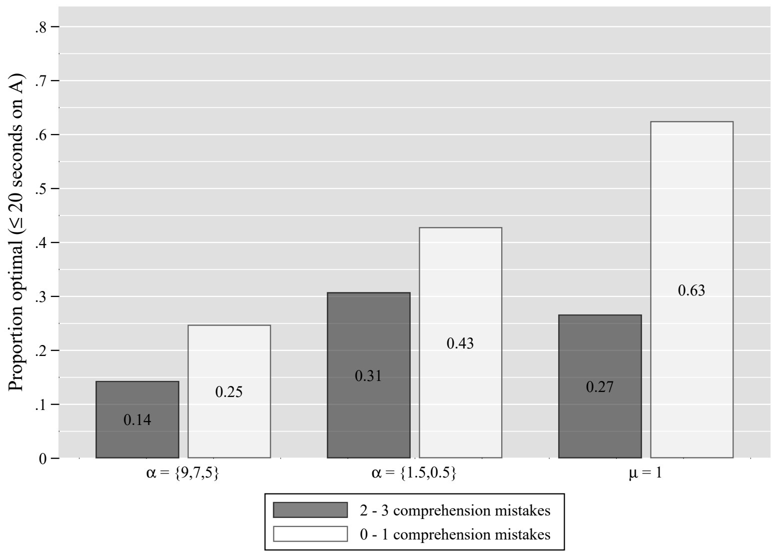


Figure 2.10: Optimal behaviour across all sessions and comprehension

Figure 2.10 shows the proportion of individuals allocating their time optimally across comprehension and studies. As before, 'optimal' behaviour refers to $Time_A \leq 20$ seconds and 'suboptimal' behaviour $Time_A > 20$ seconds. In all further analysis, we restrict the sample to subjects who (i) were within or above C* for $\alpha = \{9, 7, 5\}$ and (ii) understood how to generate at-least one set for all treatments.[56] As Figure 2.10 shows, comprehension matters for optimal choice in an intuitive way, with individuals who have good comprehension allocating time significantly more often to activity B (Wilcoxon rank-sum tests, all $p < 0.05$). However, this raises a potential concern, as differences in comprehension might be weakening treatment effects. For example, there is a significant increase in the proportion of optimal choice when moving from the treatments with $\mu = \frac{1}{3}$ to $\mu = 1$ for those with "high" comprehension (Wilcoxon rank-sum tests, all $p < 0.01$).[57] Nevertheless, even when focusing on subjects who clearly understand the environment very well, their behaviour is still highly consistent with narrow bracketing. In the $\alpha = \{9, 7, 5\}$, for instance, just 25% of the sub-sample with high comprehension allocates time optimally. The importance of comprehension is confirmed in a regression analysis (see Table B6 in the Appendix). While comprehension is thus an important factor, our main results are robust to including comprehension as a control.[58]

(ii) Attentional focus

In Chapter 1 we have seen that *attention* on timing relative to payment attributes matters for individual choices. Thus, it seems plausible that it might matter in the context of narrow bracketing, where there is a tension between timing and payment features of the environment. A subject

[56] Since the level of α in Study 1 is the same as in the $\alpha = 9$ treatment in Study 2, this data has been pooled. Therefore, $\alpha = \{9, 7, 5\}$ refers to data from Studies 1 and 2.

[57] This result holds when comparing $\alpha = 0.5$, $\mu = \frac{1}{3}$ with $\alpha = 0.5$, $\mu = 1$. Figure B12 in the Appendix disaggregates Figure 2.10 across all treatments.

[58] Note that unsurprisingly comprehension is significantly correlated with measures of cognitive ability (see Figure B8 in the Appendix).

more focused on the time dimension (neglecting payment attributes) might be more attracted to the restriction placed on activity A and thus be more likely to narrow bracket. To investigate relative attentional focus, we ask subjects two *ex-post* questions that vary on a scale from 1 (not at all) to 8 (very much):

1. *'While deciding which of the activities to spend time on, to what extent did you pay attention to their payment?'*

2. *'While deciding which of the activities to spend time on, to what extent did you pay attention to their time-limits?'*

Similar to Chapter 1, we can create a composite measure of relative attentional focus by subtracting attention to payments from attention to time-limits.[59] This measure ranges from −7 to +7, with higher values representing relatively more attention to timing than payments. Figure 2.11 provides the mean attentional focus on timing, payments and our composite measure. Averages are calculated across a binary variable of time spent on activity A ($Time_A > 20$ seconds) and

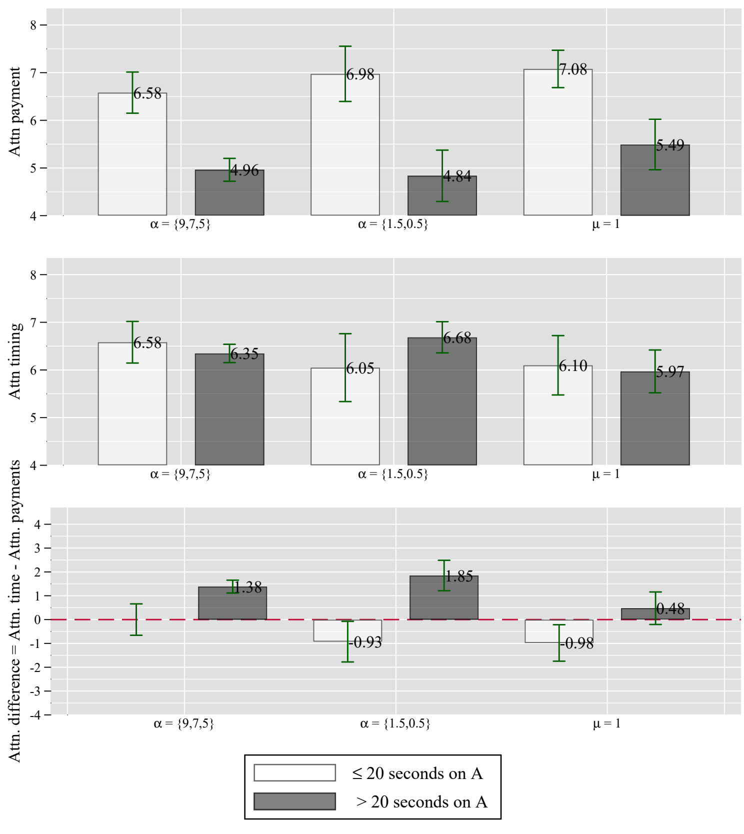


Figure 2.11: Attentional focus and perceived urgency

Notes. Whiskers represent the 95% confidence intervals.

[59]While it may matter whether subjects focus on payoffs/time for activity A versus activity B, we abstract from this difference. The idea is that if some subjects are *generally* paying more attention to elements regarding time, then they are likely paying more attention to the trade-off between the different time constraints and thus, neglecting the trade-off considerations for payoffs. The converse is true for attention to payoffs.

the treatment groups. The results show that subjects with optimal behaviour ($Time_A \leq 20$ seconds) are more focused on payments across all levels of α and μ (top panel)(Wilcoxon rank-sum tests, all $p < 0.01$). By contrast, there is no difference in attention to timing (middle panel). Combining these two attentional metrics, we can see an overall separation across the two behaviour types in terms of relative attentional focus. Optimal behaviour is associated with relatively more attention on payments (Wilcoxon rank-sum tests, all $p < 0.05$). Moreover, consistent with the attention-based narrative from Chapter 1, the treatments in $\mu = 1$ are associated with higher attention on payments, less attention on timings, and consequently relatively less attention on timings versus payments than in $\alpha = \{9, 7, 5\}$ and $\alpha = \{1.5, 0.5\}$ (Wilcoxon rank-sum tests, all $p < 0.05$).[60]

These results are again supported by a regression analysis, shown in Table B9 in the Appendix. Overall, the analysis indicates that attentional focus correlates significantly with time spent on activity A. Changing features of the environment that move relative attention away from payments and towards timings thus ought to lead to more suboptimal time allocation behaviour (more on this in Chapter 3).

(iii) Demographics and measures of cognitive ability

Much of the literature, as discussed in Section 2.2, has posited a strong linkage between cognition and choice-bracketing behaviour. However, the empirical evidence is mixed. Moreover, there is a lack of evidence surrounding which demographic characteristics correlate with bracketing, barring the evidence that women are more likely to exhibit certain choice bracketing phenomena than men (Rabin & Weizsäcker, 2009; Koch & Nafziger, 2019). Thus, in this section, we conduct an exploratory analysis regarding the role of specific demographic characteristics and measures of cognitive ability. The demographics we examine are gender, age and level of education. The measures of cognitive ability are: *perceived maths ability* (Koch & Nafziger, 2019); two questions concerning their *numeric ability* with respect to increasing quadratic functions (Lipkus et al., 2001; Bitterly et al., 2020); three questions from an alternate form of the *cognitive reflection test* (Thomson & Oppenheimer, 2016)[61]; and a measure regarding whether subjects integrate multiple lotteries or if they view them in isolation, known as *lottery isolation* (Tversky & Kahneman, 1985).[62] See Section B.2 in the Appendix for a detailed description of these metrics.

To test the effect of demographics and cognitive ability, we run probit regressions with suboptimal behaviour ($Time_A > 20$ *seconds*) as the dependent variable on demographics, measures of cognitive ability and controls. The results are presented in Table 2.4. Across all specifications (columns 1-7), we see a strong and significant effect for age and tertiary education on time allocation. Older subjects and those with tertiary levels of education, all else constant, are associated with lower probabilities of suboptimal behaviour. When it comes to gender, results are less clear. Women are more likely to engage in suboptimal time allocation in our environment,

[60]See Figure B13 in the Appendix for the results aggregated across time spent.

[61]This new version is a response to the original being highly overused and confounded with numeracy (Welsh et al., 2013; Campitelli & Gerrans, 2014)

[62]For budgetary reasons, these measures were all unincentivised. As a data quality check, we compared the data from these variables with other papers, showing that our data is comparable. For example, on average, subjects got 58.16% of the cognitive reflection test questions correct in our sample relative to 56.20% in Thomson & Oppenheimer (2016).

	Full sample						
	(1) ME	(2) ME	(3) ME	(4) ME	(5) ME	(6) ME	(7) ME
DEMOGRAPHICS							
Female	0.06* (0.04)	0.04 (0.04)	0.06 (0.04)	0.06* (0.04)	0.06* (0.04)	0.04 (0.03)	0.05 (0.03)
Age	-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)
Tertiary education	-0.07** (0.03)	-0.07** (0.03)	-0.07** (0.03)	-0.07** (0.03)	-0.07** (0.03)	-0.07** (0.03)	-0.07** (0.03)
COGNITIVE ABILITY							
Perceived maths ability		-0.02** (0.01)				-0.01* (0.01)	
Base. Numeric ability=0							
Numeric ability=1			0.04 (0.06)			0.06 (0.06)	
Numeric ability=2			-0.10* (0.06)			-0.06 (0.06)	
Base. Cognitive reflection=0							
Cognitive reflection=1				-0.10* (0.05)		-0.09 (0.06)	
Cognitive reflection=2				-0.13*** (0.05)		-0.10** (0.05)	
Cognitive reflection=3				-0.19*** (0.06)		-0.12** (0.06)	
Lottery isolation					0.05 (0.04)	0.05 (0.03)	
Cognition (standardised)							-0.07*** (0.02)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	694	694	694	694	693	693	693
Psuedo R^2	0.10	0.11	0.12	0.11	0.10	0.13	0.12
Log-lik	-359.40	-356.07	-351.03	-355.54	-358.12	-345.73	-351.01
χ^2	71.98	75.70	77.84	81.53	75.08	91.87	85.01

Table 2.4: Marginal effects: Demographics and cognitive ability

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: A binary measure taking on a value of 1 if $Time_A > 20$ seconds, and 0 otherwise. Explanatory variables include: demographics such as gender, age, and education; as well as measures of cognitive abiltiy such as perceived maths ability (1 to 10), numeric ability (0 to 2), cognitive reflection (0 to 3) and lottery isolation (binary). A higher score thereby always refers to higher ability levels. Additional controls include treatments, subject performance regarding total strings correctly reversed, activity order, and expectations about performance.

but the effect is only significant in 3 out of 7 specifications. The insignificance in columns 2,3, and 6, suggests that the gender difference may be driven by differences in maths ability and not cognitive reflection. Indeed, we find that women are associated with lower perceived maths and numeracy scores but not cognitive reflection or lottery isolation (see Table B11 in the Appendix).

Moving to our measures of cognitive ability, we see that under our measures of numeracy, it is perceived maths ability (column 2) that appears to be a more significant correlate of suboptimal time allocation than the actual numeracy score (column 3). On the other hand, while there is evidence to suggest that cognitive reflection (column 4) is a significant correlate of suboptimal behaviour, there is no such evidence for lottery isolation (column 5). When controlling for all measures in one regression (column 6), the measure of cognitive reflection appears the most robust. Finally, we construct a general measure of what we refer to as 'cognition', by summing and standardising the three cognitive measures (numeric ability, cognitive reflection and lottery isolation). Column 7 shows that a one standard deviation increase in this measure is associated with a seven percent decrease in the probability of spending $Time_A > 20$ seconds on activity A.

Overall, the results thus show that cognitive ability plays an essential part in time allocation behaviour in our environment. They also suggest that it is not necessarily a question of numeric ability that determines whether or not subjects broad bracket their time allocation decisions, but rather cognitive reflection. In particular, broad bracketing in time allocation appears to strongly correlate with cognitive reflection. In other words an individual's ability to bypass their intuitive response (i.e., to work on the initially higher paying alternative) and to engage their more thought-out and reflective response (i.e., the globally payoff maximising option) is crucial in our context. These results align with dual-process theory (Epstein, 1994; Sloman, 1996; Chaiken & Trope, 1999; Stanovich & West, 2000; Kahneman & Frederick, 2004), where 'System 1' thinking involves cognitive processes that require little conscious deliberation relative to 'System 2' processes that are slow and consume more cognitive bandwidth.

Taken together, these exploratory results provide important insights into correlates of narrow bracketing. In particular, we show that cognitive reflection, age and education play a crucial role for the prevalence of suboptimal behaviour.

2.9 Concluding remarks

In this research, we set out to answer two questions. Does narrow bracketing in time allocation exist? And if so, how robust is it? To answer these questions, we first develop a theoretical framework in which broad and narrow bracketing can yield different time allocations across two activities. We then operationalise the theory in an experimental design.

In Study 1, we observe large volumes of suboptimal time allocation, representing substantial narrow bracketing behaviour. In line with the theory, we find a significant dichotomy in behaviour types, with little behaviour consistent with partial narrow bracketing. While previous literature suggests some degree of narrow bracketing, the extent of suboptimal behaviour in our set-up is surprisingly high, particularly given that narrow bracketers forego a substantial amount of income. In Study 2, we thus examine how sensitive narrow bracketing is to decreases in the opportunity costs associated with working on activity B, the globally payoff maximising option. The results show no significant decrease in time spent on activity A, the suboptimal option, suggesting that narrow bracketing is very robust in our context.

Consequently, in Study 3, we explore the boundary conditions of suboptimal time allocation in our environment. We find that suboptimal behaviour can be reduced by reducing the piece-rate incentives from working on activity A. In addition, it seems that the time restriction placed on activity A makes the latter more attractive, in line with the literature on scarcity effects. We thus find that both pecuniary and non-pecuniary motives drive time allocation in our experiment. Finally, after pooling the data across all studies, we explore a range of correlates for time allocation and find that comprehension, attentional focus, demographics, and cognitive measures (cognitive reflection and numeric ability) play a significant role.

To our knowledge, these results are a first of their kind in the choice-bracketing literature, revealing high propensities of narrow bracketing in time allocation within a controlled experimental environment. They also showcase considerable efficiency losses incurred when interdependencies in time allocation decisions are not considered. While we explore correlates of suboptimal behaviour in our environment and have isolated factors that significantly attenuate narrow bracketing, there are still several open questions that we hope future research can tackle. For example, while we observe larger shares of optimal behaviour in treatments that remove the time restriction on activity A, there are likely still other factors that favour suboptimal choice. Some of these might be idiosyncratic reasons such as random choice, misunderstandings, or people following simple heuristics. It does seem plausible that there might be a $\frac{1}{N}$ rule where subjects divide their time and output by the number of components within the decision environment. This heuristic has been shown in portfolio allocation and is a particular case of a more general one coined 'naive diversification' (Simonson, 1990; Read & Loewenstein, 1995; Benartzi & Thaler, 2001, 2007). There might also be a general tendency for subjects to want to work on all tasks, leaving 'no stone unturned'. This might reflect a natural curiosity of wanting to explore what is in each component (Berlyne, 1954, 1978; Loewenstein, 1994; Reio Jr et al., 2006; Kidd & Hayden, 2015).

While we hope that this chapter gives rise to future research exploring additional factors that influence narrow bracketing in time allocation and how it can be avoided, we can clearly answer our two key questions. Narrow bracketing in time allocation is both a pervasive and remarkably persistent phenomenon

CHAPTER 3

Urgency & Busyness —factors that induce suboptimal time allocation?

Abstract

Utilising the novel experimental paradigm developed in Chapter 2, we conduct an experiment to test the effect that inducing two forms of *scarcity* have on narrow bracketing behaviour. More specifically, we experimentally manipulate the time at which particular tasks are available (*urgency*) and the extent to which subjects are short of time (*busyness*). In our framework, these manipulations are expected to increase the propensity of suboptimal time allocation behaviour as urgency and busyness result in an increased focus on narrow choices and disregard for the broader picture. Contrasting these hypotheses, our experimental results show negligible support for urgency and busyness increasing suboptimal time allocation. Analysis of *ex-post* questionnaire responses suggests that, while the experimental manipulations objectively affected urgency and busyness, the subjects, on average, did not perceive them as such. After dichotomising these measures, in lieu of treatment manipulations, we observe a significant correlation between high perceived urgency and busyness on narrow bracketing behaviour. These results demonstrate the importance of shifting perceptions and offer some interesting implications for future research.

3.1 Introduction

Scarcity is not just a physical constraint. It is also a mindset. When scarcity captures our attention, it changes how we think—whether it is at the level of milliseconds, hours, or days and weeks. By staying top of mind, it affects what we notice, how we weigh our choices, how we deliberate, and ultimately what we decide and how we behave. When we function under scarcity, we represent, manage, and deal with problems differently."

—Mullainathan and Shafir (2013)

Numerous important economic outcomes are a result of multiple smaller decisions, meaning that it is crucial to account for all the *interdependencies* that might arise across choices to make the highest quality decisions. The failure to account for such interdependencies is known as *narrow bracketing*. In Chapter 2, we saw that not only is narrow bracketing pervasive in time allocation, but it is also surprisingly robust —showing that suboptimal time allocation appears ubiquitous in time-use decisions. In this research, we utilise the same experimental and theoretical framework that disentangles narrow from broad bracketing behaviours and use it to investigate mechanisms that we hypothesise to induce *even* more narrow bracketing behaviour. Specifically, we are interested in the effect that *urgency* and *busyness* have on suboptimal time allocation behaviour.

We study this by experimentally manipulating the time at which particular tasks are available and the extent to which subjects are short of time. Building on our previous work on the *mere urgency effect* in meaningful ways (see Chapter 1), endowing specific tasks with an appearance of urgency (i.e., spurious urgency) is expected to induce narrow bracketing, because of an unwarranted lure of tasks that appear urgent. Similarly, giving subjects more tasks to do *per-unit* of time available, and in this sense, making them busier, is also predicted to exacerbate narrow bracketing. While urgency and busyness are distinct in the sense that the former is a restriction on the availability of tasks and the latter a restriction on how short an individual is in time, they fall under the same category of *scarcity* —that Mullainathan & Shafir (2013) argue forms its own distinct psychology.

In their seminal book titled *Scarcity: Why Having Too Little Means So Much*, Mullainathan & Shafir (2013, p.8) define scarcity as the mindset of *"... having less than you feel you need"* and conjecture that any form of scarcity (from time, money, and calories to companionship) all have remarkably similar effects on cognition and behaviour. Specifically, scarcity leads to attentional shifts described as 'tunnelling', that result in neglect of anything outside of the cognitive tunnel. As an example, adapted from Shah et al. (2012), consider a very busy academic who is currently consumed with the demands of an imminent work deadline (i.e., a manuscript submission). This deadline causes them to have a heightened focus (a 'focus dividend') on the pressing needs of the task. However, while this focus has some beneficial features, it results in tunneling and neglect of things outside of the tunnel (i.e., the academic might forget to prepare next week's lecture, or forget to pick up their kids from school).[1]

In this sense, there is a double-edged sword to scarcity. On the one hand, it helps concentrate cognitive resources toward whatever is made scarce. However, on the other, it leads to attentional neglect of other things. This rationale is neatly illustrated in the following passage from Mullainathan & Shafir (2013, p.19): *"... our argument in this book is quite simple. Scarcity captures our attention, and this provides a narrow benefit: we do a better job of managing pressing needs. But more broadly, it costs us: we neglect other concerns, and we become less effective in the rest of life"*. This psychology generates sharp predictions in our framework concerning bracketing behaviour and manipulating urgency and busyness. More specifically, these forms of scarcity are predicted to increase a subject's propensity to narrow bracket (or tunnel as Mullainathan & Shafir (2013) refer).[2] This occurs because such forms of scarcity are hypothesised to make subjects place undue attention on temporal features in the decision-environment (at the expense of other things). On the other hand, we also hypothesise that a heightened focus (a focus-dividend) leads to subjects who are narrow bracketing to become better at allocating their resources in the narrow-bracketed problem.

Contrasting these predictions, our experimental results show negligible support for urgency and busyness affecting bracketing behaviour of subjects in our framework. Specifically, there is minimal evidence to support that the spurious urgency of an option adversely affects the time allocation behaviour of subjects. For busyness, there is no evidence that the required number of tasks per-unit of time available adversely affects time allocation behaviour. Regarding a focus-dividend, we find that subject's narrow bracketing make fewer mistakes in the task itself, on

[1] The distinction between a 'focus-dividend' and tunneling is described in the following passage by Mullainathan & Shafir (2013, p.33): *"**Focus is a positive**: scarcity focuses us on what seems, at that moment, to matter most. **Tunnelling is not**: scarcity leads us to tunnel and neglect other, possibly more important, things"* [emphasis added].

[2] In this research, we hold the view that tunneling is conceptually analogous to 'narrow bracketing'.

average.[3] Since narrow bracketing in our set-up, by definition requires working on more subtasks (as outlined in Chapter 2), these results are consistent with the notion that having more to do (even if this is a choice) increases focus and concentrates cognitive resources to their most granular level, in line with the psychology of scarcity.

A crucial consideration concerning the results surrounding time allocation is that *perceptions* of scarcity are key to the psychology that Mullainathan & Shafir (2013) propose. Through *ex-post* measures, we investigate perceptions surrounding urgency and busyness, finding that there are surprisingly no differences across treatments. This shows that perceptions of urgency and busyness are insensitive to our manipulations, surprising given that our treatments were designed to manipulate *real* urgency and busyness.[4] As a thought-experiment, we dichotomise these variables, splitting them into high and low perceived urgency and busyness, and use them in lieu of treatment manipulations. These results show a significant correlation between such perceptions and suboptimal time allocation of subjects. Viewing the spuriously urgent option as more urgent or who perceiving themselves oneself as busier in the task are significantly associated with higher levels of narrow bracketing. Thus, these results in terms of perceptions are strongly aligned with those suggested by the psychology of scarcity. This work, to our knowledge, is the first to contribute to the literature on how such forms of scarcity affect choice bracketing in time allocation. Our framework provides an important stepping stone for future research to explore the role of urgency and busyness in time-use decisions.

The rest of this chapter is organised as follows. Section 3.2 briefly goes through the literature surrounding the psychology of scarcity. Section 3.3 provides the experimental design, hypotheses and experimental procedures. Section 3.4 showcases the results from an experiment. Section 3.5 discusses the implications of these results and areas for future research.

3.2 Literature surrounding psychology of scarcity

There is a growing body of experimental, quasi-experimental, survey and neuroscience research investigating how people respond to various forms of *scarcity*.[5] This literature posits some specific characteristics and mechanisms surrounding scarcity. The first is that various forms of scarcity have similar effects on cognition and decision-making behaviour (i.e., there are common mental processes that are triggered by scarcity of *any* resource) —for example, the psychology of poverty (i.e., financial-scarcity) shares common mechanisms with the psychology of being busy (i.e., time-scarcity). Second, scarcity imposes itself on us and captures our attention. As mentioned above, people experiencing scarcity have a heightened focus on pressing needs. This results in a focus-dividend, which in some respects improves decision making. However, this increased focus can cause tunneling, resulting in attentional neglect for things outside of the tunnel, and can worsen decision making. Third, one negative downstream consequence of this

[3] More specifically, we observe that subjects who narrow bracket are associated with fewer mistakes as a proportion of total output relative to those who broad bracket in their time allocation.

[4] In Section 3.3.1, we outline in detail what we mean by 'real' as this is an important consideration.

[5] For example, see Cannon et al. (2019); Carvalho et al. (2016); Huijsmans et al. (2019); Mani et al. (2013); Mullainathan & Shafir (2013); Kurtz (2008); Roux et al. (2015); Shah et al. (2012, 2015, 2018); Sharma & Alter (2012); Spiller (2011); Tomm & Zhao (2016); De Bruijn & Antonides (2021); Zhao & Tomm (2018); Hamilton et al. (2019a,b); Kalil et al. (2022); Binkley & Bejnarowicz (2003); Goldin & Homonoff (2013); and Maćkowiak et al. (2021).

heightened attentional focus is depleted cognitive functioning (referred to as a 'bandwidth tax').

The effects of how scarcity, across different resources, can result in attentional shifts that explain behaviour such as over-borrowing is experimentally studied by Shah et al. (2012). In one of their experiments, the authors create a game akin to the *Angry Birds* video-game, called 'Angry Blueberries'. In this computerised game, the objective is to slingshot blueberries into waffles, earning as many points as possible. In their experiment, subjects are assigned either to a poor or rich group, determining how many blueberries each subject have to throw (the poor having fewer than the rich), over ten rounds. In addition, some subjects can borrow with interest. The authors find that the blueberry poor spend more time focussing on each throw (i.e., have a heightened focus on what is restricted), but borrow more each round (i.e., have attentional neglect in the domain that is unrestricted, how much they can borrow), earning less overall than the rich.

In another experiment, Shah et al. (2012) manipulate how short subjects are of time while playing *Family Feud*, a trivia game where each question allows multiple answers. Time poor subjects have smaller budgets than do the rich in terms of time allowed each round, and subjects play until exhausting their budgets. In the same way as before, poor subjects overborrow each round relative to the rich, but are more focussed on the task at hand (they make more guesses and earn more points). While these are two distinct decision environments involving different forms of scarcity, the results show similar behaviours. These behaviours are analogous to bracketing. Scarcity induces narrow bracketing through a heightened focus on what is restricted. While this leads to a benefit in the narrow-bracketed problem (better focus on pressing needs), it costs the decisionmaker more broadly (they overborrow).

To explore how scarcity can impose itself on our cognitive functioning, Mani et al. (2013) study Indian sugar cane farmers, who earn the bulk of their income once a year after harvest, having to ensure their funds keep them going until the following harvest. As a result, these farmers are cash poor before harvest and cash richer afterwards. The authors conduct a series of cognitive tests on the farmers two-months before and two-months after harvest, finding a significant drop in cognitive performance. More recently, however, this result of reduced cognitive bandwidth after induced scarcity has been scrutinised. For example, in a prominent replication exercise, Camerer et al. (2018) fail to reproduce the results from the first experiment by Shah et al. (2012), showing that scarcity-induced focus leads to increased cognitive fatigue in future tasks. As a response, driven by *"sheer curiosity and some concern"*, Shah et al. (2019, p.2) conduct a replication exercise across all the experiments in their paper with very high-powered designs. While their findings are in line with Camerer et al. (2018), finding no evidence to support the cognitive depletion result, they still find strong evidence for their central hypothesis that scarcity leads to over-borrowing (true for multiple kinds of resources), and that it increases focus on how people use their resources (focus-dividend).

In this research, we contribute to the scarcity literature in several important respects. For the mere urgency effect, our work in Chapter 1 has shown that it is a robust phenomenon. However, this is within the confines of a very specific experimental paradigm. That is, the original design is in a one-shot dichotomous choice environment. In this environment, subjects make real time use decisions under spurious urgency in continuous time. Thus, this research further extrapolates the result to other domains. Moreover, to our knowledge, this is the first work that exogenously manipulates the level of busyness in a controlled experimental environment. While Zhu et al. (2018) find that those who perceive themselves as busier are more likely to make dominance violations, this is does not involve an exogenous variation in busyness. Moreover, while Shah et al. (2012) investigate time-limits, this is not in the context of someone choosing how to allocate their time. Thus, while being amongst the first to contribute to the literature on scarcity in the face of time allocation, this work also offers a key contribution with respect to investigating its affects in the contexts of choice bracketing.

3.3 The experiment

3.3.1 Design

In this experiment, we utilise the same theoretical and experimental paradigm that we constructed in Chapter 2 to investigate the impact of manipulating (i) *urgency* and (ii) *busyness* on the propensity to narrow bracket. Table 3.1 shows three-cells of a *2x2* experimental design that we implement.[6] Subjects are randomly allocated to one of three conditions, 'Baseline', 'Urgent' and 'Busy' —making this a between-subject design.[7]

	Low-busyness	High-busyness
Non-urgent	**Baseline**	**Busy**
Urgent	**Urgent**	-

Table 3.1: Experimental design

Analogous to Chapter 2, the experiment is broadly comprised of three stages. *Stage 1* includes instructions, comprehension questions and a practice round; *Stage 2* involves working on the activities of choice; and *Stage 3* consists of a series of *ex-post* comprehension questions. There were several adjustments to the instructions and decision-screen in this experiment relative to the studies in Chapter 2. These changes were largely to improve comprehension, which was observed to correlate significantly with choice behaviour. While we provide details for some of the key adjustments and their rationale below, more information is provided in Section C.4 of the Appendix. To explain differences across treatments, it is instructive to first recap the experimental environment used in Chapter 2, along with adjustments in the visual display. Where necessary, we also provide reminders of certain theoretical parameters and considerations. This environment is our Baseline condition, which we then manipulate in the Urgent and Busy treatments.

[6] We decided to implement the three-cells instead of the four to maximise statistical power, given our budget.

[7] We chose a between-subject design instead of a within-subject one for two reasons. First, balanced randomisation across the three cells should control for any differences in the sample across gender, age and education —factors that we have seen are significantly correlated with choice-bracketing behaviour in time allocation. Second, we were conscious of the experiment taking too long in a within-subject design. Had the experiment taken longer, we would also need to remunerate subjects more and this would reduce our statistical power.

Baseline decision-environment

Figure 3.1 provides a screenshot of the Baseline decision-environment that subjects face. As before, subjects must choose how to allocate their time across activities A and B —which involve the *same underlying task* but different *payoffs* and *time availabilities.*[8] Specifically, the task provides subjects with sequences of randomly generated six-letter strings ('sdfjiw') and requires them to reverse type these strings ('wijfds'). However, each activity has two different coloured strings, making a total of four colours (i.e., yellow, blue, green and red). These string-colours are randomised on screen for each subject to mitigate any ordering effects.[9] In the example, activity A has yellow and blue strings, while activity B has green and red strings.

Within each activity, subjects need to reverse strings and produce *sets.* Two strings of each colour form a set in an activity, and subjects are paid according to the number of sets they produce for that activity. For example, a subject who produces six yellow and eight blue strings in activity A, will be paid for three sets in that activity.[10] Recall that this is due to the *weakest-link structure*, where activity output is determined by the lowest number of correctly reversed coloured strings within an activity, as described in Section 2.3.[11] Subjects accrue earnings in the form of points, which are converted into pounds at a pre-defined exchange rate (100 points = £1.00). As in Chapter 2, the payment schedules for each activity differ. Specifically, $y_A = \alpha(\#\ of\ sets\ in\ A)$ and

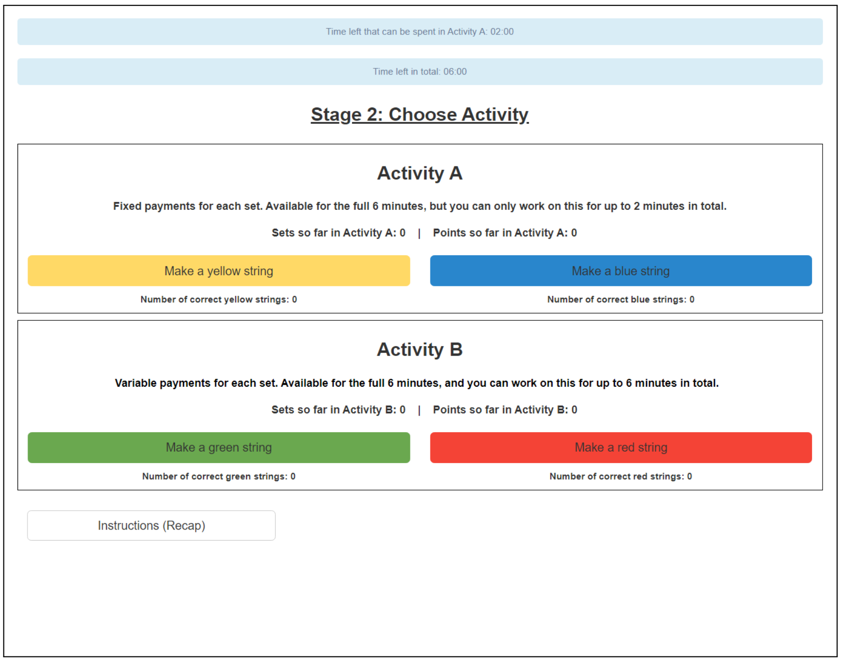


Figure 3.1: Choose-activity screen: Baseline condition

Notes. This is the Baseline condition decision-screen where subjects decide what colour strings to reverse. At this stage, subjects have already completed reading through instructions, answering comprehension questions and participated in a practice round where they learn how to generate string reversals.

[8] The actual labels for subjects were activity 'F' and 'V', standing for 'Fixed' and 'Variable' payments, as before.

[9] To elaborate, if more people have 'blue' as their favourite colour, and blue is shown first all the time, this would confound the result. Since they are randomised, there a $4! = 24$ colour combinations (set of yellow, red, green and blue colours).

[10] Note that different coloured strings from different activities do not form sets. That is, if a subject correctly reverses six yellow and eight red strings (with no correct string reversals in blue or green), they will produce no sets.

[11] That is: $min\lfloor \frac{6}{2}, \frac{8}{2} \rfloor = 3$.

$y_B = \beta(\#\ of\ sets\ in\ B)^2$. That is, income generated from activity A is constant, while income from activity B is quadratic.[12] Subjects have a total of 360 seconds of working time. They can freely choose how to allocate their time between the coloured strings within activities. However, the time that they spend working on these differs. Specifically, activity A is available for *up to 120 seconds*, and activity B is available for *up to the full 360 seconds*. Recall that this implies $\mu = \frac{1}{3}$. These time-limits are reflected at the top of the screen in Figure 3.1, where there are two timers. The first is the *'Time left that can be spent in Activity A'* and the second is *'Time left in total'*. These timers do not run down when a subject is in the choose-activity screen.[13] The total timer runs down when a subject is on any of the coloured strings of activities A or B. However, the timer on activity A *only* runs down when a subject is working in activity A. This implies that within the 360 seconds of working-time, *any* 120 seconds can be spent on activity A. That is, the time restriction placed on activity A is 'delocalised' in the sense that it can be spent at any point within the total 360 seconds of working time. Once the total timer is exhausted, subjects are moved to Stage 3 of the experiment.

Let us now note some of the key changes in the display in comparison to the one shown in Figure 2.2 of Chapter 2. The first is that *each* string-reversal colour is now unique to one component (n) of an activity.[14] This contrasts with previously in Chapter 2, where there were green and blue strings in each activity. The rationale for this change was to clarify which strings generate sets in either of the two activities and to reduce possible confusion. Additionally, on the decision screen, we provide subjects with the number of 'points' they generate (for payment at the end of the experiment). We think that this allows subjects to see what sets, and therefore, points they were accumulating for each activity.[15] This is to minimise any mistake of thinking they might be in one activity, but are in fact, in the other. We also added a link (at the bottom of the page) that allows subjects to go back to the instructions in case they felt like they needed to refresh their knowledge of the set-up.

Manipulating urgency

The core aim of the Urgent treatment is to vary the degree of *spurious* urgency, in line with the mere urgency effect (Zhu et al., 2018). The urgency is 'spurious' because under certain conditions in our theoretical framework in Section 2.3, no subjects should be spending any time on activity A at all, if they want to maximise their payoffs. We manipulate *urgency* in this treatment by changing whether the time restriction placed on activity A is *delocalised* (non-urgent) or *localised* (urgent). Figure 3.2 provides a screenshot of the working environment in the treatment where we increase urgency. Upon comparison with Figure 3.1, there are no apparent visual differences, barring the string colours being in a different ordering, reflecting their randomisation. The key difference between the Urgent treatment relative to Baseline condition is in the timer for activity A at the top of the screen. As before, the total timer runs down when a subject is in any of the

[12] We implement the same non-uniform pricing schedule method as in Chapter 2 (see Section 2.4.2).

[13] As explained previously in Chapter 2, the rationale for this is so there is exactly 360 seconds on time spent working and as a result, to be true to our mutually exclusive and exhaustive set of ways of spending time as set-out in our theoretical model.

[14] Recall from Section 2.3 in Chapter 2 that activities A and B each sub-divide into $n \geq 1$ component activities.

[15] Another minor change is that we moved from the concept of Experimental Currency Units (ECUs) into points as we felt this was more digestible terminology

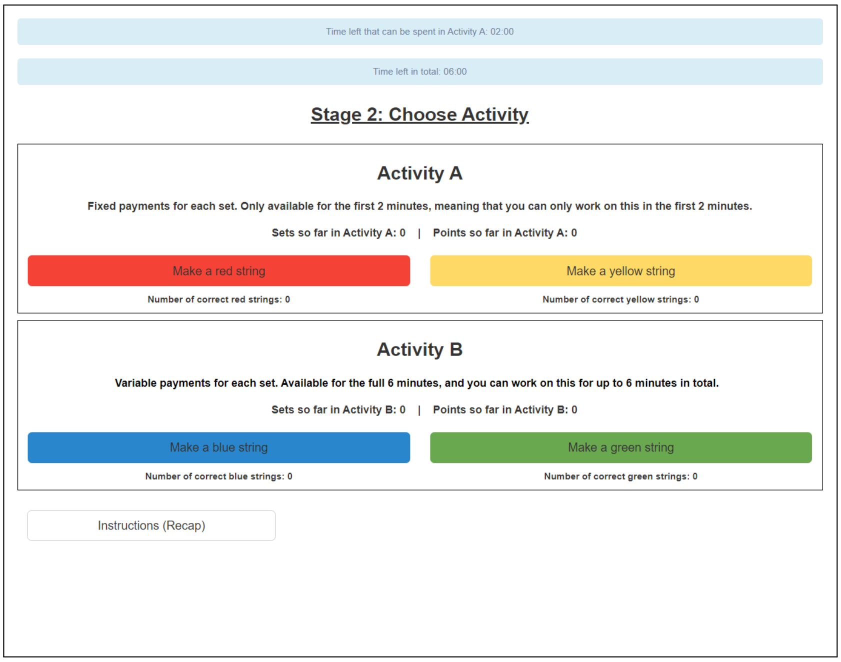


Figure 3.2: Choose-activity screen: Urgent treatment

Notes. This is the Urgent treatment decision-screen where subjects decide what colour strings to reverse. The key difference between this treatment and the Baseline condition is that the timer on activity A is fixed to the first 120 seconds of working time.

coloured strings of activity A or B. However, now different to the Baseline condition, the timer on activity A also runs down when a subject is working in activity A or B (much the same as the total timer). This implies that within the 360 seconds of working-time, *only the first* 120 seconds can be spent on activity A.[16] By tying the availability of activity A to the first 120 seconds, we believe this is likely to foster a sense of 'urgency', because if a subject would like to spend the maximum time allowed on activity A, they need to do so immediately. This is analogous to the mere urgency effect, since we are essentially adding urgency to a task that subjects should not be spending any time on, thus making it 'spurious' urgency.

Manipulating busyness

Before discussing how we manipulate busyness in our environment, we need to define what we mean by 'busyness'. An important idea premised in Mullainathan & Shafir's (2013) book in the context of time allocation is that scarcity of time should not be thought of as absolute (since we all have 24 hours in the day), but rather having a lot to do in the time that we have available. This refers to having more to do per-unit of time available. This is illustrated in the following excerpt from the book: *"Imagine a day at work where your calendar is sprinkled with a few meetings and your to-do list is manageable...Now, imagine another day at work where your calendar is chock-full of meetings. What little free time you have must be sunk into a project that is overdue. In both cases*

[16] We thank Daniel Read for pointing out at the NIBS September 2020 Workshop that our localised treatment combines localisation itself with achieving urgency by localising on the first available time, whereas an additional possibility would be to localise on the last available time. Budgetary constraints prevented us from exploring this additional possibility. However, we conjecture that almost all subjects would spend all unreserved time on activity B, if that time followed the reserved time (instead of preceding it), with the result that piece rates in activity B would already exceed those in Activity A before A became available.

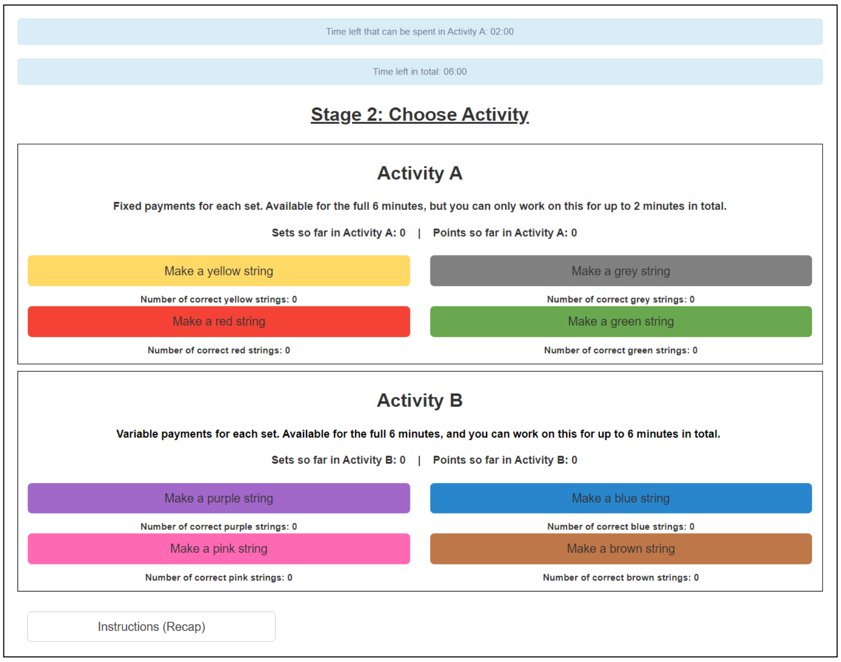


Figure 3.3: Choose-activity screen: Busy treatment

Notes. This is the Busy treatment decision-screen where subjects decide what colour strings to reverse. The key difference between this treatment and the Baseline condition is the number of component activities (n) and thus, the number of coloured strings per activity.

time was physically scarce. You had the same number of hours at work and you had more than enough activities to fill them. Yet in one case you were acutely aware of scarcity, of the finiteness of time; in the other it was a distant reality, if you felt it at all" (Mullainathan & Shafir, 2013, p.15). This is what we mean by being busier and is what we aim to manipulate in the Busy treatment.

To do this, we vary the number of components (i.e., string-reversal colours) that must be attended to produce a unit of activity output. Specifically, there are four string-reversal colours for each activity, making eight colours in total. This is illustrated in Figure 3.3, showing the working environment for the Busy treatment. The screenshot shows eight string-reversal colours across activities A and B, instead of the four in the Baseline condition and Urgent treatment. Notice that there are four new colours (i.e., purple, pink, brown and grey) adding to the previous set of four. As before, the order of colours is randomised on screen.[17] To produce a set in any one of the activities now requires one correctly reversed string from each of the four colours in that activity. For example, one correctly reversed yellow, grey, red and green string will produce one set in activity A, for which subjects get paid according to the same payoff schedules as in the Baseline condition and Urgent treatment.

One important feature of this busyness manipulation is that while we manipulate the number of components (i.e., string-reversal colours) to attend to at any given time to produce a set, we do not vary the total number of strings required for a set. Specifically, a total of *four* strings are always needed to produce a set, irrespective of treatment. This is to avoid a confound, since we do not want to drastically change the subjects income generating capacity and to consequently make different sections of the payoff schedules attainable. To illustrate the importance of this

[17] In the Busy treatment, there are $8! = 40,320$ colour combinations (set of yellow, red, green, blue, purple, pink, brown and grey colours).

feature, we need to briefly recall the main results from our theoretical framework outlined in Section 2.3 in Chapter 2.

Specifically, the main result is that under certain conditions, broad bracketers spend all time on activity B (and none on activity A), while narrow bracketers spend as much time on activity A as possible and the rest of the remaining time on activity B. This result is distilled in the following condition C*: $\frac{T}{n}(2-\mu) > R > \frac{\mu T}{n}$, where $R \equiv \left(\frac{\alpha}{\beta}\right)\left(\frac{p}{k}\right)$. From this, we can define $S = np$, where S refers to the minimum number of strings needed in total to produce one unit of activity output (in activities A or B). As before, n refers to the number of components (or string-reversal colours), and p is the number of such components required to produce one-unit of activity output. That is, each *set* requires p strings from each of n colours. Since S is constant across all treatments, it does not affect whether C* holds for any given subject. Specifically since $p = \frac{S}{n}$, we can rewrite C* in the result as: $\frac{kT}{S}[2-\mu] > \frac{\alpha}{\beta} > \frac{\mu k T}{S}$. The important implication of this reformulation is that n does not feature and all that remains is S, which is fixed across treatments. Moreover, since T, S and μ are all held constant across treatments, C* defines the same range of $\frac{\alpha}{\beta}$ in every treatment, for a subject of given k.

A crucial consideration is whether varying the number of string-reversal colours, while holding total output per treatment constant, affects busyness. Although there is no difference in the total time available or the total number of strings to reverse, the Busy treatment requires subjects to attend to four colours when generating sets for payment. Whereas, the Baseline condition and Urgent treatment requires subjects to balance their efforts across only two colours to generate sets. Thus, since the former adjusts the number that subjects need to attend to per-unit of time available, we argue that the Busy treatment makes subjects more busy. Whether subjects *perceive* this as such is also an important consideration —something that we also measure.

3.3.2 Hypotheses

The psychology of scarcity predicts that the experimental manipulations of *urgency* and *busyness* leads to subjects paying more undue attention to temporal features in the decision environment (at the neglect of pecuniary ones), resulting in more narrow bracketing behaviour as a result. In line with this prediction, for those who meet C*, we expect to observe a higher prevalence of subjects choosing to work on activity A in the treatments relative to the control, despite this being the suboptimal option in terms of income maximisation.

Hypothesis 1. *There will be a higher prevalence of narrow bracketing in the Urgent and Busy treatments relative to the Baseline condition.*

Another key prediction is that while we expect subjects to be more likely to narrow bracket in their time allocation, tunneling causes an increased focus on the task at hand. This results in subjects becoming better at solving the narrow bracket problem. In our context, this implies that subjects become (i) better at splitting their time, and therefore output, across components within an activity (i.e., better at allocating themselves between the coloured strings), as well as (ii) more focused in the tasks themselves when reversing strings (i.e., making fewer mistakes in each coloured string they work on). We test these two features by constructing two measures:

(i) *set-inefficiency rates*, and (ii) *string-inefficiency rates*, which we define in greater detail when examining the experimental results. Broadly, these metrics allow us to measure two distinct levels of focus-dividend behaviour. The former is a more general measure of efficiency between components, while the latter is a more granular measure of efficiency within components.

Hypothesis 2. *There will be lower set- and string-inefficiency rates in the Urgent and Busy treatments relative to the Baseline condition.*

In addition to the core predictions set out by the psychology of scarcity in terms of the observed behaviour in our environment, we can also investigate the mechanisms through which this behaviour occurs. Specifically, we can measure relative attentional focus on timing versus payoffs in the same manner as in Chapter 2. Specifically, we can utilise the two *ex-post* questions in Stage 3 on attention to (i) payments, and (ii) timings. We construct a composite measure of relative attentional focus by deducting attention to payments from attention to timings. Since these measures run on a scale from 1 (not at all) to 8 (very), the composite measure ranges from -7 to $+7$ (higher values represent higher relative attention to time-limits than payments).

Hypothesis 3. *Subjects will pay relatively more attention to timings than payments in the Urgent and Busy treatments relative to the Baseline condition.*

Lastly, another important question is whether the manipulations of *urgency* and *busyness* are *perceived* as such by subjects. This is interesting as while our treatments do manipulate *real* urgency and busyness, in the sense that (i) the availability of activity A is localised to the first 120 seconds; and (ii) subjects have more to do per-unit of time available —there is an open question about how experimental subjects will perceive them. As a result, as part of our *ex-post* questionnaire, we also ask subjects how urgent they perceive activity A to be and how busy they feel during the task:

1. *'How urgent did activity A feel?'* → scale from 1 (not urgent at all) to 8 (very urgent)
2. *'In general, how busy did you feel during the task?'* → scale from 1 (not busy at all) to 8 (very busy)

These questions can be used as manipulation checks of the treatments, where we predict higher perceived urgency and busyness in each respectively.

Hypothesis 4. *Subjects will perceive activity A to be more urgent in the Urgent treatment and themselves to be busier in the Busy treatment relative to the Baseline condition.*

3.3.3 Procedures

Recall from Chapter 2 that a key objective is to ensure that C^* holds for as many subjects as possible, as these are the subjects that we restrict our sample size to when conducting our empirical analysis. Another criterion is that we need a sufficiently low level of narrow bracketing in our Baseline condition since our treatments are designed to exogenously increase the propensity of it. In light of these considerations, we set the same parameters as in our '$\alpha = 7$' treatment in Chapter 2. Thus, $\frac{\alpha}{\beta} = 7$. Recall that we cannot set k as this is the latent ability of the subject to reverse strings. However, these parameters have all been set with a distribution of this measure in mind (from previous experiments reported in Chapter 2).[18] The reason for the decision to choose $\alpha = 7$ is that this value allowed for the least amount of narrow bracketing behaviour in our previous experiments, while still maintaining high levels of subjects who satisfy C*.

Having decided on $\alpha = 7$, we now need to specify need to specify performance bounds for which C* can hold, allowing us to identify bracketing types. As specified above, C* is written to define a range for $\frac{\alpha}{\beta}$, governed by $\frac{kT}{S}$ and μ. It can be rearranged to define a range for $\frac{kT}{S}$, ruled by $\frac{\alpha}{\beta}$ and μ: $\left(\frac{1}{\mu}\right)\left(\frac{\alpha}{\beta}\right) > \frac{kT}{S} > \left(\frac{1}{2-\mu}\right)\left(\frac{\alpha}{\beta}\right)$, where $\frac{kT}{S}$ represents the total number of sets a subject would complete if they maximised their number of completed sets in the time-frame T. With $\mu = \frac{1}{3}$, $\alpha = 7$, and $\beta = 1$, the number of sets generated needs to lie within the following bounds: $21\ sets > \frac{kT}{S} > 4\ sets$. Recall that we cannot distinguish the two types of bracketer outside these bounds, since narrow and broad bracketers are predicted to do the same thing in that case. At low levels of performance (# of sets < 5), spending maximal amount of time on activity A is payoff maximising, since there is insufficient scope for quadratic returns to activity B to 'kick-in'. For extremely high levels of sets produced (# of sets > 21), the returns to activity B have already surpassed the returns to activity A by μT, such that even a narrow bracketer would spend all time on activity B.

The experiment was programmed on LIONESS Lab (Giamattei et al., 2020). We recruited subjects through Prolific (prolific.co) and restricted the subject pool to those in the United Kingdom, and who had prior experience on the platform. Table C1 in the Appendix provides summary statistics across the treatments. The data was collected in one experimental session in November 2021, with a total sample size $N = 538$, equally split across the treatments and informed through *a-priori* power analysis (Power = 0.80).[19] Subjects took an average of 21.04 minutes to complete the study with narrow and broad bracketers earning an average of £8.55/hr and £11.65/hr in total, respectively.

3.4 Experimental results

3.4.1 Testing the effects of urgency and busyness

Performance and speed classification

Figure 3.4 shows the distribution of performance across the three experimental conditions, showing that the average possible sets generated is roughly constant. This is further illustrated in Table 3.2 with 87%, 90%, and 83% of the sample being within C* in the three conditions, respectively. Thus, performance concerning possible sets generated does not statistically differ across them (a χ^2 test for independent groups was conducted).[20] As a result, the sample for the remainder

[18]The average number of sets produced in previous experiments was 11, with a range from 0 to 28. See Chapter 2 for distribution of performance across the various experiments.

[19]See Figure C1 in the Appendix.

[20]Note that this does not refer to *actual* number of sets generated. Figure C4 in the Appendix, shows that average number of sets produced in the Busy treatment was slightly lower. This shows that subjects had a more challenging time allocating themselves across eight components in the Busy treatment relative to four components in the other

of the analysis is restricted those who fall within C*. However, for the sake of clarity, we also provide the full-sample results, despite negligible differences across the two.

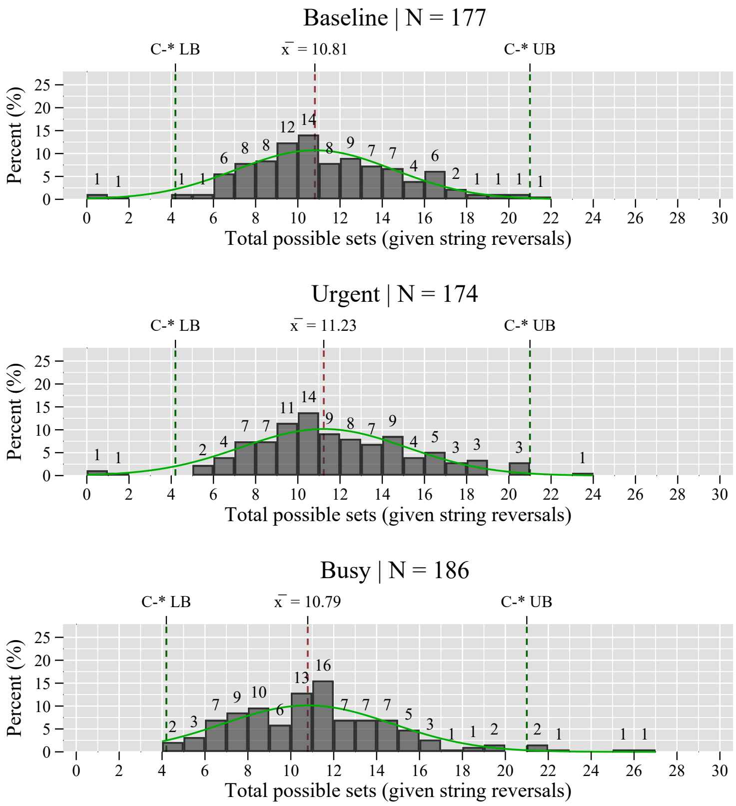


Figure 3.4: Performance

Speed categorisation	Baseline	Urgent	Busy
Speed in C* range	154 *87.01 %*	157 *90.23%*	155 *83.33%*
Speed < C* range	2 *1.13%*	1 *0.57%*	3 *1.61%*
Speed > C* range	1 *0.56%*	1 *0.57%*	5 *2.69%*
Misunderstood	20 *11.30%*	15 *8.62%*	23 *12.37%*
Total	177 *100.00%*	174 *100.00%*	186 *100.00%*

Table 3.2: Speed categorisation

Notes. Speed in C* range (21 sets > possible sets > 4 sets); Speed < C* range (possible sets < 5 sets); Speed > C* range (possible sets > 21 sets); Misunderstood (completely failed to generate a set).

Bracketing behaviour

Having identified subjects that lie within C*, we can now focus on bracketing behaviour. Figure 3.5 shows the distribution of time spent on activity A across the three-conditions, for the (i) full

two treatments, as predicted. It should be noted that the resultant losses in income from this were not substantial.

sample, and (ii) within C*. The *y-axis* represents the percentage (%) of subjects, and the *x-axis* is the time spent working on activity A, ranging from 0 to 120 seconds. As in Chapter 2, we empirically measure bracketing behaviour by categorising behaviour into three distinct groups depending on time spent on activity A: (i) narrow bracketer: $Time_A \geq 100$ *seconds*; (ii) partial narrow behaviour: 20 *seconds* $< Time_A <$ 100 *seconds*, and (iii) broad bracketer: $Time_A \leq 20$ *seconds*. Before analysing the result across treatments, we first highlight the re-occurrence of two key results that we found in Chapter 2. The first is a significant level of narrow bracketing behaviour in all three conditions (several Wilcoxon signed-rank tests, all $p < 0.01$). Second, there is a clear separation in type. More specifically, there is a minimal occurrence of partial narrow bracketing, with 92-94% of the subjects which satisfy C* locating near one extreme or the other. These results highlight that almost all subjects either spent very little or a lot of time on activity A, and the former is more common than the latter (accounting for roughly 50% of subjects, as compared with

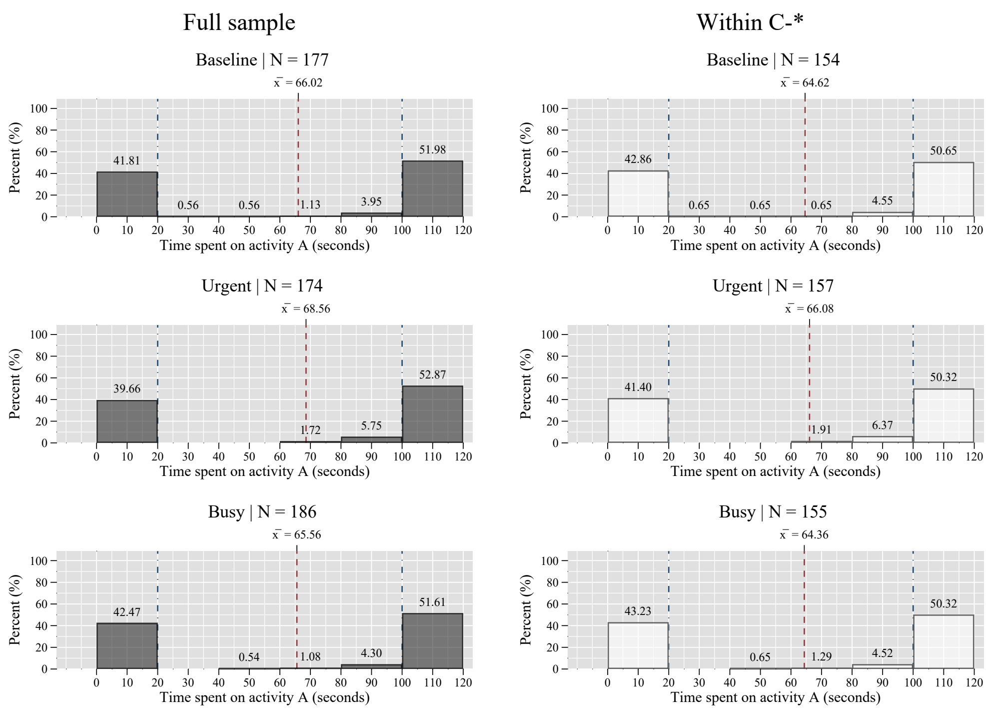


Figure 3.5: Time on activity A

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off.

Treatments	**Full sample**	**Within C***
Baseline vs. Urgent	0.937	0.900
Baseline vs. Busy	0.631	0.630
Urgent vs. Busy	0.550	0.686

Table 3.3: Overview of treatment comparisons

Notes. * p < 0.1, ** p < 0.05, *** p < 0.01. All values reported are p-values from two-sample Wilcoxon rank-sum tests. These results are also robust to other non-parametric tests, such as χ^2 tests and Kolmogorov–Smirnov tests for equality of distributions.

around 40%). These results are somewhat contrasted to the $\alpha = 7$ treatment in Chapter 2 (see Figure 2.6), which we suspect reflects better levels of comprehension achieved in this experiment relative to the previous studies. Given the hypotheses stated above, we expect a greater propensity for subjects to narrow bracket under induced urgency and busyness relative to the Baseline condition. We test the following null-hypothesis that time on activity A is equal across the three conditions —H_0: $Time_A$ (Baseline) = $Time_A$ (Urgent) = $Time_A$ (Busy). The results shown in Figure 3.5, highlight that there is no effect of either manipulation on the level of narrow bracketing.[21] This is reflected in Table 3.3, which shows the p-values and their corresponding significance level for two-sample Wilcoxon rank-sum tests across the three treatments.[22] We should note that these results are also robust to other non-parametric tests (χ^2 tests and Kolmogorov–Smirnov tests for equality of distributions).

These results indicate that inducing spurious urgency and busyness in this framework has no effect on the propensity to narrow bracket. While these results do not align with Hypothesis 1 that scarcity will induce greater degrees of suboptimal time allocation, we can dig at them further. In the decision screens, shown in Section 3.3.1, we record where a subject is at any point through the duration of the task. This allows us to see which activity any given subject is in at any second throughout the 360-seconds of time-spent working. This allows us to use a 'dynamic' measure of time spent on activity A in addition to the 'aggregate' one from before to get a more nuanced insight into behaviour across the treatments. Figure 3.6 shows the results across the three conditions for the dynamic measure of time spent. The *y-axis* now represents the proportion of subjects in activity A at any given second of time spent working for the total 360 seconds. Therefore, each second represents the proportion of subjects in the sample working on activity A. In the Baseline condition, 49% of the sample is working in activity A in the first second. Then at 120-seconds, this proportion drops dramatically to around 10% since many people have reached the maximum time of 120-seconds. The small number of people show that some are choosing to work on activity A later, though the majority are spending it in the initial stages of the task.

Moving to the Urgent treatment, the proportion of subjects working on activity A is around 58% in the first second (9% higher than in the Baseline condition). Since activity A is only available for the first 120-seconds in this treatment (by function of the experimental design), no-one is spending time on it past this point. This shows that while the experimental manipulation of urgency results in a higher proportion of subjects initially choosing to work on activity A, this effect is 'washed-out' by subjects choosing to work on activity A in the Baseline condition past 120 seconds. In the Busy treatment, there are very similar proportions of subjects at each-second in activity A relative to the Baseline condition. These results are exemplified in Table 3.4, showing ordinary least squares (OLS) estimates for several regressions at different 'blocks' of time spent

[21]Given that narrow bracketing behaviour in Chapter 2 was correlated with rates of comprehension, gender, education and age, we also disaggregated the measure of time spent on activity A across these metrics to investigate if there were treatment differences. The results from this endeavour show that the time spent on activity A does not differ along these metrics (see Figures C8, C9, C10 and C11 in the Appendix).

[22]In the same way as in Chapter 2, we can also investigate a supplementary statistic on *closeness to optimality* of subjects with respect to monetary maximisation. We can evaluate how close a subject's behaviour is to 'optimal' relative to their string-reversal speed. This measure is less restrictive as it includes those who fall outside the speed defined in C*. The results for this metric show no significant differences across the three treatments (see Figure C6 in the Appendix).

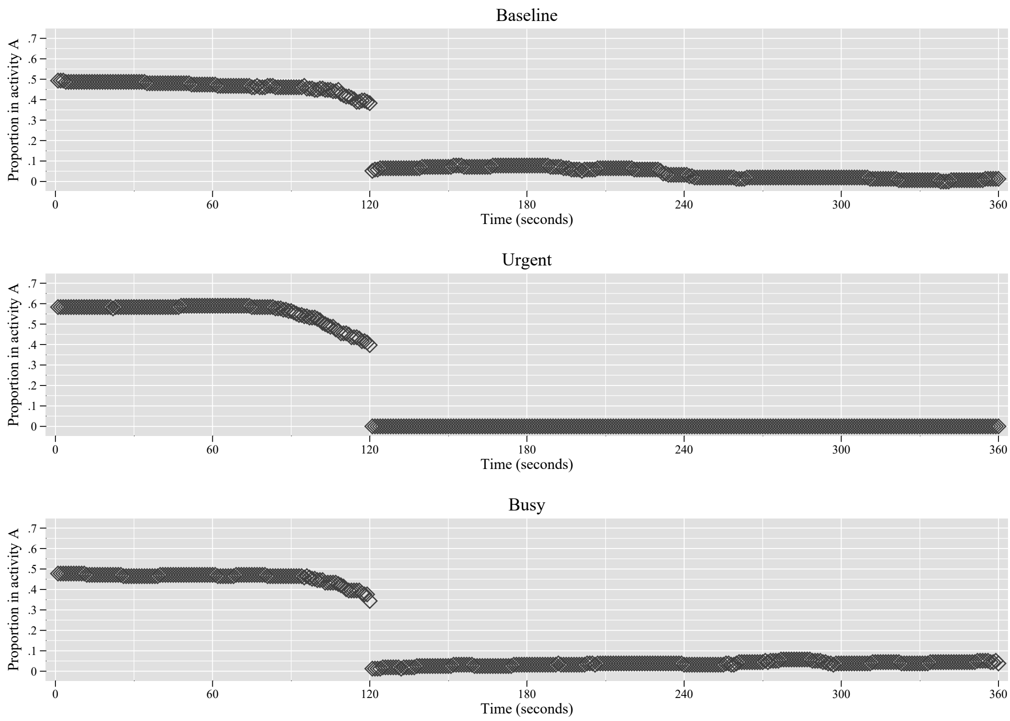


Figure 3.6: Activity ordering effects

	Time A in blocks						Time A overall
	(1) 1-60	(2) 61-120	(3) 121-180	(4) 181-240	(5) 241-300	(6) 301-360	(7) 1-360
Base. Baseline							
Urgent	0.11**	0.10*	-0.07***	-0.06***	-0.02*	-0.01**	0.01
	(0.05)	(0.05)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Busy	-0.07	-0.06	-0.05**	-0.03	0.02	0.03**	-0.02
	(0.06)	(0.05)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	1.06***	1.07***	0.13***	0.08*	-0.00	-0.03	0.38***
	(0.15)	(0.13)	(0.05)	(0.04)	(0.03)	(0.02)	(0.05)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	467	467	467	467	467	467	467
R^2	0.08	0.10	0.05	0.04	0.03	0.04	0.10

Table 3.4: Ordinary least squares: Activity ordering effects

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: Proportion of subjects in activity A across blocks of 60-second intervals. Independent measures include: treatments with Baseline as the base. Additional controls include age, gender, tertiary education, subject performance in terms of sets generated, as well as comprehension. The sample is restricted to those who generated more than 0 sets and had a speed classification within C* since it is optimal for these subjects to spend no time on activity A at all.

working on activity A. To elaborate, we regress the three treatments (with the Baseline condition as the base), on the proportion of subjects working in activity A in 60-second intervals up to 360 seconds. From 1-60 (column 1) and 61-120 (column 2), we see that there is a significantly higher proportion of subjects working on activity A in the Urgent treatment relative to the Baseline condition (albeit weak significance between 61-120 seconds). From 121-360 (columns 3 to 6), we see significantly fewer subjects working in activity A relative to the Baseline condition, a result which comes about given the experimental design. Note that in the last specification (column 7), there is no urgency effect relative to the Baseline condition, showing that the overall effect of the treatment is null. We observe null results for the Busy treatment relative to the Baseline condition (barring a strangely lower and higher level between 121-180 (column 3) and 301-360 seconds (column 6), respectively —which we cannot explain).[23] Taken together, while these results show no effect of the busyness manipulation in time allocation, there are some interesting insights with the urgency manipulation. Specifically, there appears to be an effect of urgency in terms of how time is allocated dynamically, however, this has no effect on the overall allocation of time to activity A. This result is consistent with there being a *certain* number of subjects who want to do 120 seconds of activity A, and the size of this set is independent of treatment. Thus, the observed urgency effect is not a case of urgency making activity A more attractive, but rather an effect of urgency making 'early' the only time that the 120 seconds of activity A can be done.

Result 1. *There is no evidence to support an urgency or busyness effect on time allocation (with our aggregate measure). However, there is some evidence to support an effect of the urgency manipulation on how time is spent (with our dynamic measure).*

Focus-dividend behaviour

We now turn to investigating the other key prediction that tunneling causes an increased focus on the task at hand (i.e., focus-dividend behaviour). To investigate this, we look at two measures, namely the (i) *set-inefficiency rate* and (ii) *string-inefficiency rate*. The set-inefficiency rate is the difference between the number of sets created and the number that could have been, given the subject's string-reversal speed, had they allocated those strings across colours optimally, expressed as a proportion of the latter. For each subject, it is calculated in the following way: $\textit{set-inefficiency} = \frac{\textit{Possible sets}-\textit{Actual sets}}{\textit{Possible sets}}$. On the other hand, the string-inefficiency looks at the proportion of unsuccessful string reversals relative to the total number of string-reversal attempts. It is calculated as: $\textit{string-inefficiency} = \frac{\textit{String mistakes}}{\textit{Total string attempts}}$. Thus, both measures represent a proportion, where the former is a measure of *between* component (string-colour) inefficiency and the latter *within* component (string-colour) inefficiency. Hypothesis 2 predicts lower set- and string-inefficiency rates in the treatment conditions relative to the Baseline condition. Therefore, we test several null-hypothesis for set- and string-inefficiency rates across conditions —H_0: inefficiency-rates (Baseline) = inefficiency-rates (Urgent) = inefficiency-rates (Busy) for broad and narrow bracketers (and aggregated across them).

Figure 3.7 shows the set- and string-inefficiency rates across the three treatments and a binary

[23] It should be noted that this dynamic measure of time is a slightly noisy measure since it cannot account for any computer lag that a subject might have. Further, since it is calculated by subtracting multiple timestamps every time a subject switches between components, there is also rounding-off differences that might not be taken into account. For example, this measure records over 120 seconds for some subjects in the experiment, even though this is impossible. As a result, for subjects who clearly spent their 120-seconds all at once (i.e., in the first 120 seconds), but their measure recorded more than 120-seconds, we simply cleaned the data to 120-seconds. In any cases where this was not possible, we dropped the observations. To check that we do not obtain any drastically different results across treatments, we run the exact OLS regression estimations with the uncleaned data.

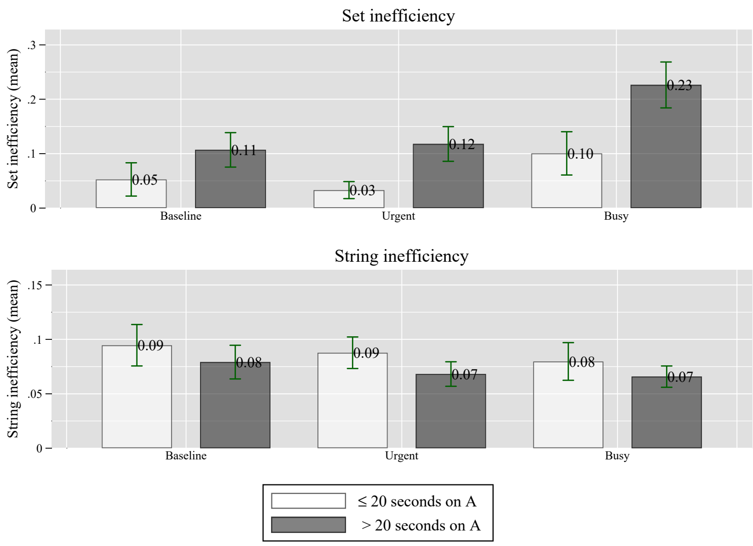


Figure 3.7: Focus-dividend behaviour

measure of time spent on activity A: referring to broad bracketing ($Time_A \leq 20$ seconds) and a more general conception of narrow bracketing ($Time_A > 20$ seconds) behaviour.[24] This more general conception is because we observe very little evidence of partial narrow bracketing, in line with the results from Chapter 2. Figure C13 in the Appendix provides the inefficiency rates across the three treatments (but aggregated across bracketing behaviour).[25] We first examine set-inefficiency (top panel), with three key results. The first is that narrow bracketers have higher levels of set-inefficiency relative to broad bracketers —true for all conditions (Wilcoxon rank-sum tests, all $p < 0.01$). The second is that set-inefficiency rates are significantly higher in the Busy treatment relative to the Urgent treatment and Baseline condition (Wilcoxon rank-sum tests, all $p < 0.01$). The third is that narrow bracketers have relatively higher set-inefficiency than broad bracketers in the Busy treatment compared to Baseline condition and Urgency treatment. These results are supported by an OLS estimation where we interact the treatments with the binary measures of time spent on activity A, yielding a positive interaction for the Busy treatment, controlling for demographics, performance and comprehension (see Table C2 in the Appendix).

These results sharply contradict the prediction that urgency and busyness will lead to a focus-dividend, resulting in better allocation across components. However, in some sense, they are quite intuitive in our setting because spending time on activity A means that subjects have more components to attend to and deliver equality than if they just worked on activity B. To illustrate this possible explanation further, we measure the number of entries into the components worked. That is, how often a subject switches from one string-colour to another. Across the three treatments, narrow bracketers are associated with significantly more entries into components than broad bracketers, with an average of 8.56 and 5.71, respectively (Wilcoxon rank-sum test, $p < 0.01$). Moreover, in the Busy treatment, narrow bracketers are associated with an average of 12.50 entries and broad bracketers 8.47 entries (Wilcoxon rank-sum test, $p < 0.01$). These results

[24]The key results are not affected by this classification.

[25]We also disaggregate the inefficiency rates across narrow and broad bracketers. These results are provided in Figures C14 and C15 in the Appendix

highlight the fact that more components worked means more opportunities to drop strings along the way and lose out on set generation.[26]

Let us now move to string-inefficiency (bottom panel), for which we observe two key results. The first is that on aggregate across the three treatments, narrow bracketers are associated with lower levels of string-inefficiency than broad bracketers (Wilcoxon rank-sum test, $p < 0.05$). However, within each treatment, these results are only significant for the Urgent treatment (Wilcoxon rank-sum test, $p < 0.05$). The second is that the Busy treatment compared to the Baseline condition, is weakly associated with lower levels of string-inefficiency (Wilcoxon rank-sum test, $p < 0.10$). The latter results are more significant in an OLS estimation, controlling for demographics, performance and comprehension (see C2 in the Appendix). These results, in contrast to the set-inefficiency rates, align with the hypothesis that urgency and busyness lead to increased focus-dividend behaviour. On average, subjects who are narrow bracketing make fewer mistakes as a proportion of total string attempts relative to those who are broad bracketing. These results are interesting as they highlight that having more to do, in the sense of more string-reversal colours to attend to, make subjects 'focus' more in each task, and make fewer mistakes as a result. To investigate this further, we construct a variable that is the number of components (i.e. string-reversal colours) a subject worked on. Recall that in the Baseline condition and Urgent treatment, the number of components worked can range from two to four, whereas, in the Busy treatment, it ranges from four to eight —due to the weakest-link structure.[27] Indeed, several regressions (see Table C3 in the Appendix) show a significantly negative relationship between the number of components worked and string-inefficiency. Moreover, these regressions show a significantly positive relationship between number of components worked and set-inefficiency.

The key takeaways from this analysis are that the number of components worked is significantly positively correlated with set-inefficiency and negatively correlated with string-inefficiency. We assign this result to the fact that the task environment requires cognitive processing at different 'levels'. The first and most broad is thinking about allocating time across activities A and B. It asks subjects to deliberate on how they want to allocate their time (to earn the most income). The second, slightly more focused level, is allocating oneself across components (related to set-inefficiency) to maximise the number of sets generated in their chosen activity (or activities). The third and most granular is how to apply oneself to the string-reversal task, where presumably a goal is to maximise the number of strings and the number of minimise mistakes (related to string-inefficiency). While quite nuanced, the results are consistent with the notion that tunneling concentrates cognitive resources to their most granular level.

Result 2. *While there is no evidence to support a between-component focus-dividend, there is evidence to suggest a within-component focus-dividend across conditions. Working on more components is associated with fewer string-reversal mistakes as a proportion of total strings attempted in the task.*

[26]This is also reflected in the fact that narrow bracketers are associated with significantly fewer completed sets on average than broad bracketers across the three treatments (Wilcoxon rank-sum test, $p < 0.01$). See Figure C5 in the Appendix.

[27]Recall that minimum number of components worked cannot be zero in the Baseline condition and Urgent treatment, nor can it be less than three in the Busy treatment, since we exclude individuals who misunderstand how to generate a single set.

3.4.2 Ex-post measures

Attentional focus

After having investigated treatment behaviour in the previous sections, we now turn to the analysis of our ex-post measures. Figure 3.8 shows the means for attentional focus across the three treatments and the binary measure of time spent on activity A. Figure C16 in the Appendix provides the attentional differences across the three treatments (but aggregated across bracketing behaviour). In line with Hypothesis 3, we test several null-hypothesis for attentional focus across conditions —H_0: attn. difference (Baseline) = attn. difference (Urgent) = attn. difference (Busy) —for broad and narrow bracketers (and aggregated across them). As observed in Chapter 2, there is clear and significant evidence to show that narrow bracketers focus relatively more on timings than payments when compared to broad bracketers (Wilcoxon rank-sum tests, all $p < 0.01$). Moreover, we see that broad bracketers are paying equivalently high attention to both payment and timing attributes, reflecting that their average relative attentional focus is not significantly different from zero (see bottom panel).

Upon examination of any treatment variations, a Wilcoxon rank-sum test across the three treatments shows that relative attention to timings versus payments is significantly higher for the Urgent treatment compared to the Baseline condition ($p < 0.05$). Moreover, we observe weakly greater relative attention to timings than payments for narrow bracketers in the Urgent

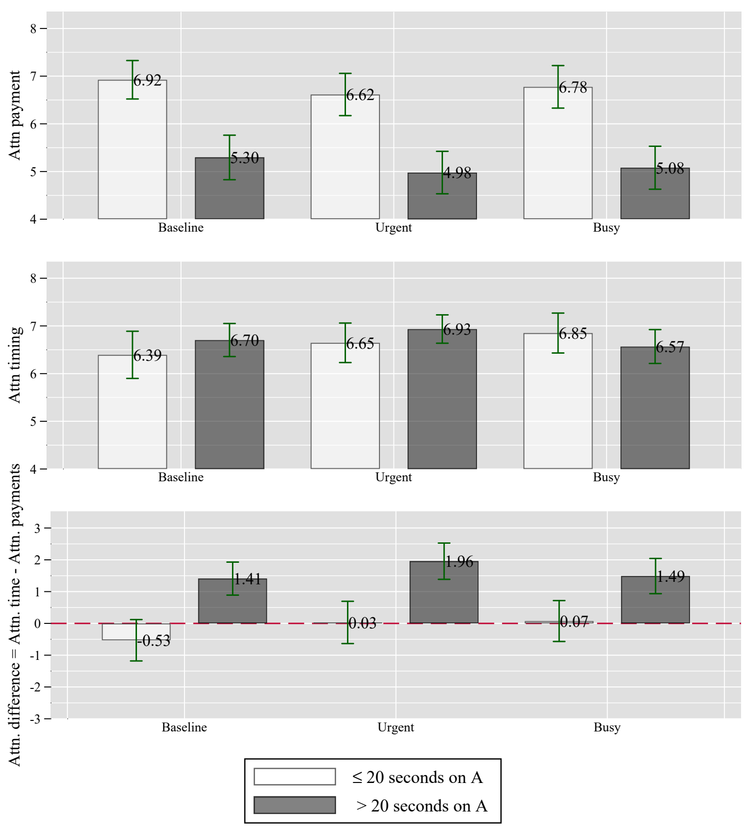


Figure 3.8: Relative attentional differences

treatment relative to narrow bracketers in the Baseline condition (Wilcoxon rank-sum test across the three treatments, $p < 0.10$). However, these results are not robust in a regression analysis, controlling for demographics, performance and comprehension (see Table C4 in the Appendix). In aggregate, the results from the relative attentional focus measure align with those observed above for time spent on activity A across the treatments. While we observe some treatment variation in the Urgent treatment, we cannot identify a clear result.

Result 3. *Although the relative attentional focus is significantly correlated with bracketing behaviour, there are negligible statistical differences across treatments.*

Manipulation checks

To take stock of the results so far, in line with the findings from Chapter 2, we have seen that narrow bracketing behaviour strongly correlates with a higher relative attentional focus on timing than payments. Moreover, while there are some differences across treatments, these results are not unanimous, which may explain why there is only weak evidence supporting an urgency effect and no evidence supporting a busyness effect. We now turn to whether the urgency and busyness manipulations are *perceived* as such by subjects. Figure 3.9 shows the means of these measures plotted against the treatment conditions, as well as the binary measure of time spent on activity A.[28] The top graph represents perceived urgency, and the bottom perceived busyness. The results show that narrow bracketing is associated with significantly higher levels of perceived urgency relative to broad bracketing —true for all treatments (Wilcoxon rank-sum tests across the three treatments, all $p < 0.01$). Further, contrary to what we would expect in the Urgent treatment, there are no higher levels of perceived urgency (Wilcoxon rank-sum test). While

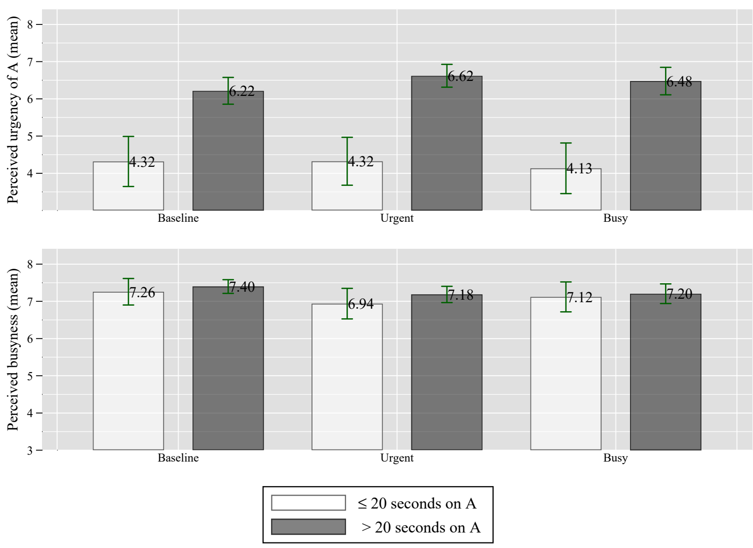


Figure 3.9: Perception of urgency and busyness

[28]Figure C18 in the Appendix, provides the results aggregated across bracketing behaviour. Moreover, Figure C17 in the Appendix provides the distributions of the measures. While both are leftward skewed, the busyness measure is severely left-skewed, with 50% or more of the sample across conditions stating they felt 'maximally' busy.

this is surprising concerning Hypothesis 4, it is in line with the weak results we observe for the relative attentional measure.

To explore this conjecture, we present an OLS regression analysis in Table 3.5. Columns 1 and 2 are specifications for attentional difference, columns 3 and 4 are perceived urgency and, columns 5 and 6 are perceived busyness. The dependent variable in columns 1, 3 and 5 is the proportion of time spent in activity A in the first 120 seconds. For someone who spent the first 120 seconds in activity A, this measure would take on a proportion of 1. Whereas, for someone who spent 120 seconds at any point after the first 120 seconds, this measure would take on a value of 0. These specifications are restricted to narrow bracketers (i.e., only includes those who spent $Time_A > 20$ seconds *in total*). This is because we are interested in attentional differences across the subjects who spent at-least some time on activity A. The results of these specifications show that there is no significant correlation between any of the perceived measures (relative attention on time, perceived urgency and busyness) and the ordering in which a subject worked on activity A. That is, there are no significant associations between any of the percevied measures and whether a subjet spent time on activity A in the first 120 seconds relative to those at any other point. Columns 2, 4, and 6, on the other hand, show these results for the full sample (who are within C* and generate more than one set). The dependent variable is the total proportion of time spent in activity A across the entire working time (and not just for the first 120 seconds). The results from these specifications show the same results as we see in Figure 3.9. Subjects who focus more on attention to timing than payments are significantly more likely to spend time on activity A. Moreover, those who perceive it as more urgent and themselves as busier (albeit with a weakly significant effect on busyness) are also more likely to spend time on activity A. Taken together, these results reaffirm that attention strongly correlates with allocating time in

	More attention on time		Perceived urgency of A		Perceived busyness	
	(1) 1-120	(2) 1-360	(3) 1-120	(4) 1-360	(5) 1-120	(6) 1-360
Attentional difference	0.01 (0.01)	0.02*** (0.00)				
Perceived urgency			0.01 (0.01)	0.03*** (0.00)		
Perceived busyness					-0.00 (0.02)	0.01** (0.00)
Constant	0.99*** (0.09)	0.33*** (0.04)	0.91*** (0.10)	0.16*** (0.04)	1.02*** (0.13)	0.31*** (0.05)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
N observations	268	467	268	467	268	467
R-squared	0.03	0.16	0.03	0.26	0.02	0.10

Table 3.5: Ordinary least squares: Urgency and attention

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variables: Columns 1, 3 and 5; the proportion of subjects in activity A from 0-120 seconds: Columns 2, 4 and 6; the proportion of subjects in activity A from 0-360 seconds. The independent measures in Columns 1 and 2; Attentional difference (-7 to +7). Columns 3 and 4; Perceived urgency of activity A (1 to 8). Columns 5 and 6; Perceived urgency of activity A (1 to 8). Additional controls include ordering of activities on screen, age, gender, tertiary education, treatments and subject performance in terms of sets generated. The sample has been restricted to only those that have spent > 20 second in activity A in columns 1, 3 and 5 and includes the full sample in columns 2, 4 and 6. All samples are for subjects within C*.

aggregate terms, but contrasting our previous result on dynamic timing here we do not observe these nuances.[29]

Moving to the perceived busyness measure, the first result that stands out is the high rate of average perceived busyness across all conditions. In fact, at-least 50% of the sample across conditions stated they felt 'maximally' busy. This highlights a possible *ceiling effect* issue and might explain why the level of perceived busyness does not differ across any of the conditions.[30] Further, contrasting urgency, narrow bracketing is not significantly associated with higher levels of perceived busyness. This is surprising since subjects are strictly busier by attending to more components when spending time on activity A. This also shows that increasing the number of components is not necessarily increasing perceptions of busyness *per se*. Indeed, several pairwise correlations show that perceived busyness is not correlated with the number of (i) components worked on, (ii) switching between components, or (iii) strings correctly reversed either (see Figure C19 in the Appendix). While these results might reflect the little heterogeneity in the perceived busyness measure itself, they may also highlight that we have not yet pinned down what makes subjects see themselves as more or less 'busy' in our set-up (more on this below). Nevertheless, one overwhelming result is that the current task environment is generating high levels of perceived busyness with all its moving components.

Result 4. *Both manipulation checks show that, on average, subjects do not perceive activity A to be more urgent in the treatment that localises it, nor do they perceive themselves to be busier in treatment that gives them more to do per-unit of time available.*

In the aforementioned analysis, we have seen some but arguably negligible differences between the Baseline condition and two treatments regarding relative attentional focus and perceptions of urgency and busyness. Recall that the aggregate measure of time on activity A showed no difference in the time allocated across treatments. However, there were differences in how that time was allocated dynamically, particularly for the Urgent treatment. This indicates that there might be perceived differences between subjects who spent their first 120 seconds of working time in activity A and those who did something else. If this is true, then, for example, a subject who focuses more on timings, perceives the activity as more urgent and themselves busier may be significantly correlated with spending time in activity A in the first 120 seconds than in the last 120 seconds (or another time period that is not within the first 120 seconds). This provides an insight into the relationship between different working patterns and perceptions of relative attentional focus, as well as perceived urgency and busyness.

3.4.3 Simulated experiment

In the previous section we established that there were no differences in perceived urgency and busyness across our treatment manipulations. However, we did find heterogeneity in individ-

[29]Since only a handful of subjects spend time on activity A outside of the first 120 seconds. We also see if this result holds with the data from the previous experiments in Chapter 2. See Table B10 in the Appendix. The key difference here is that the busyness measure is now highly significant (more on below).

[30]Note: when referring to a ceiling effect, this does not necessarily mean that subjects did not feel busier in the Busy treatment. Instead, it might be the case that the ceiling effect is on our measure of perceived busyness. We cannot disentangle what the underlying reasons are.

ual perceptions of these dimensions. Thus, we can now use this heterogeneity to construct a thought experiment where we dichotomise the measures of perceived urgency and busyness into low (from 1 to 5) and high (from 6 to 8) and use these in lieu of 'treatment' manipulations. Note that because the distribution of each measure is strongly leftward-skewed, there are more observations for the high end of both metrics than for the low end.[31] Moreover, since this is not an exogenous assignment into different treatment conditions, this thought experiment should only be considered as correlational evidence. Figure 3.10 shows the results for this exercise. For perceived urgency, there is a substantial increase in the time spent on activity A (Wilcoxon rank-sum test, $p < 0.01$). On the other hand, while there appears to be an attenuation in the expected direction for the perceived busyness measure, this result is not significant (Wilcoxon rank-sum test). That said, once controlling for demographics, performance and comprehension, it becomes statistically significant in a regression framework (see Table C5 in the Appendix).

Additionally, as a robustness check, we conduct the same analysis with all the data pooled from Chapter 2. This endeavour shows a strong and significant effect of both perceived measures on the time spent on activity A (Wilcoxon rank-sum test, $p < 0.01$). These results are provided in Figure C21 in the Appendix. Thus, in light of the results, particularly those after pooling

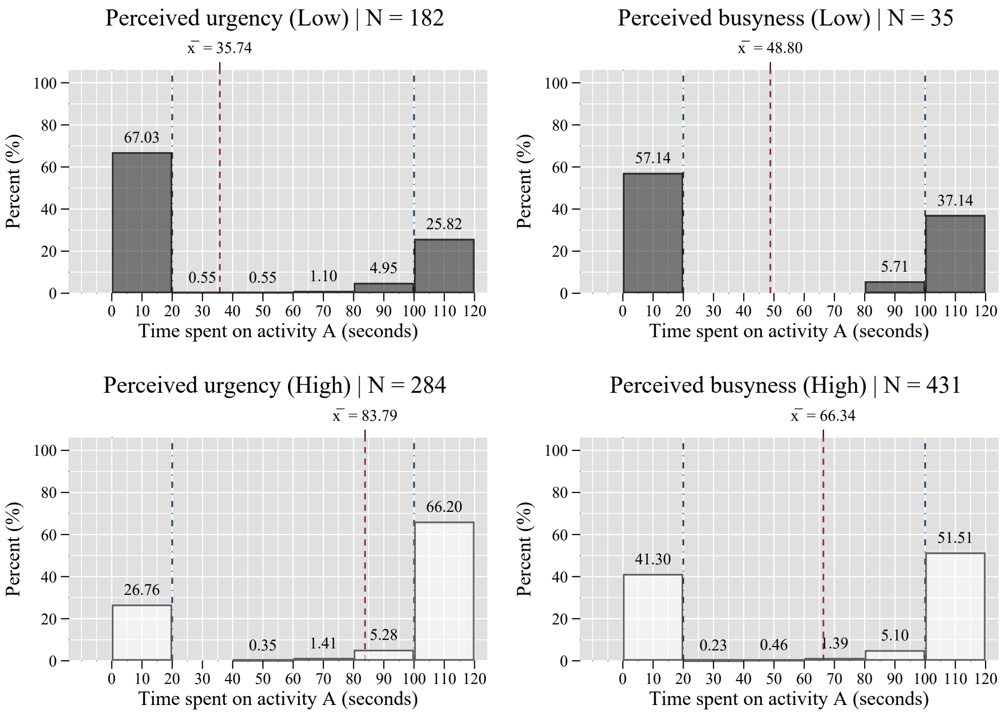


Figure 3.10: Time on activity A (dichotomised manipulation checks)

Notes. Top. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. Bottom. Whiskers represent the 95% confidence intervals. Note that there are no dash-dots as in the previous time on activity A figures, since we have moved away from the concept of bracketing as outlined in the theoretical framework.

[31] For example, the distribution of perceived busyness in the task is so skewed to the left that only 35 observations lie to the left of it. We ideally would have liked to split the measure into low (1 to 4) and high (from 5 to 8), however, there would only be 15 observations in this case. Moreover, it does not seem appropriate to split the measure along the median values, as we would just be comparing extreme perceptions with even more extreme perceptions.

the data from Chapter 2, there is a strong effect of perceived urgency and busyness on time spent on activity A. These results demonstrate that *perceptions* of urgency and busyness are an important correlate of time spent on activity A. They point to the possibility of a causal link between perceptions and time allocation behaviour that is worthy of further investigation. If such a causal link exists, then implementing treatments that exogenously increase urgency and busyness —but *most crucially*, subjects' perceptions —one should expect to observe more narrow bracketing behaviour.

Result 5. *After dichotomising the perceived urgency and busyness measures, in lieu of treatment manipulations, we observe the hypothesised effects in narrow bracketing behaviour.*

3.5 Discussion and conclusions

In this research, we designed an experiment that exogenously manipulated the (i) spurious urgency of a dominated activity, as well as (ii) the level of busyness that subjects experienced while working. Our main results show little support for the predictions that such elements lead to higher propensities of narrow bracketing behaviour. However, after investigating our findings more thoroughly, there is evidence to suggest that this might have been because subjects did not *perceive* there to be any differences in the treatment manipulations relative to the Baseline condition. That is, the manipulation checks showed negligible differences in perceived urgency and busyness across treatments. Consequently, as a thought experiment, we dichotomised these measures in lieu of treatment manipulations. This endeavour shows that high levels of perceived urgency and busyness are significantly associated with higher levels of narrow bracketing. Thus, while there appears to be a strong link between perceptions of urgency and busyness, and time allocation behaviour, such perceptions are relatively insensitive to our attempts to manipulate them. These results unearth some interesting points of discussion.

As illustrated in the quote opening this chapter, scarcity is not just a physical constraint but also a mindset. Thus, our positioning is that this work does not represent evidence against the psychology that Mullainathan & Shafir (2013) propose. As the theory predicts, perceptions of scarcity are evidently an important determinant of behaviour. This alludes to the fact that if these can be shifted, then behaviour might shift in turn. This point is also exemplified in Zhu et al. (2018), who find that in their treatment that exogenously varies spurious urgency of an option, subjects perceive these options to be more urgent than in the Baseline condition. Thus, a key implication for future research is that shifting perceptions of such features is not as trivial as it might initially seem. Indeed, Huijsmans et al. (2019, p.11699) discuss that *"...investigating how a scarcity mindset plays a role in decision-making in a laboratory setting is a challenging endevour."* Nonetheless, our findings offer some interesting possible avenues for future research.

One result that has puzzled us in this work and Chapter 2 is why people perceive activity A to be urgent in the delocalised condition. One possible interpretation is that the delocalised constraint in the Baseline condition relative to the localised constraint in the Urgent treatment is enough to make people perceive it as urgent —a misplaced 'urgency' effect. These speculations are supported by our investigation of the dynamic measure of time spent on activity A, showing that subjects who place relatively more attention on timings than payments as well as urgency of activity A are not more likely to work on activity A for the first 120 seconds than at some other point. Moreover, the results in the $\mu = 1$ treatments in Chapter 2, which makes activity A available for the full 360-seconds, also corroborate these findings —we observe significant attenuation in the propensity to work on activity A by removing the delocalised time restriction placed on it. Future research could investigate the perceived similarities and distinctions between varying forms of constraints placed on various time allocation tasks.

Regarding the busyness manipulation, one notable result is that subjects already perceive themselves as very busy in the Baseline condition. This might explain why we observe no significant correlation between perceived busyness and higher numbers of components worked, tasks switched between, or the number of strings produced in total. If there is a ceiling effect for subjects, then future research could attempt to lower levels of perceived busyness in the Baseline condition by changing the nature of the task (to reduce labour intensity). On the other hand, if there is a ceiling effect on the measure itself, then a more nuanced or granular measure of busyness could be implemented. Or it could involve a more nuanced measurement of features related to busyness. For example, some subjects might feel busy because they switch between tasks more often, and others might feel busier because they spend more time reversing strings. In other words, there might be distinctions between within- and between-task productivity, which affect perceptions regarding busyness that the current measure of perceived busyness cannot disentangle.

Relatedly, another area to explore is a treatment that jointly manipulates urgency busyness. Unfortunately, budget limitations meant that we could not explore the fourth-cell of our *2x2* experimental design in Table 3.1. However, the literature does link the two together. For example, Shah et al. (2012, p.683) discuss that *"... the busy often take extensions because they focus on urgent tasks, but neglect important tasks that seem less pressing."* Covey (2004) argues that busy people tend to neglect the important but not urgent tasks (in line with the mere urgency effect). Moreover, as discussed previously, Zhu et al. (2018), find that the busy (on a self-reported measure) are more likely to be susceptible to an urgency bias. Coupled with some of the refinements above, this would be an interesting area for future research to explore.

To conclude, to our knowledge, this work is the first that investigates how urgency and busyness affect time allocation behaviour in the context of choice bracketing. Ultimately, we find that individual perceptions of urgency and busyness are an important correlate of time allocation behaviour, while the 'real' manipulations of such features are not. However, what is less clear is what actually determines these perceptions. Nevertheless, our framework provides an important springboard for future research to explore the role of such features in time-use decisions. Borrowing words from Mullainathan & Shafir (2013, p.20), this research *"... is not meant to be the final word. [The psychology of scarcity] raises a new perspective on an age-old problem, one that ought to be seriously entertained. Anytime there is a new way of thinking, there are also new implications to be derived, new magnitudes to be deciphered, and new consequences to be understood."*

General conclusion

> *"We inhabit time as fish live in water. Our being is being in time."*
> —Carlo Rovelli (2017, p.1)

This thesis contributes to research surrounding *time allocation* and, in particular, circumstances under which suboptimal time-use behaviour might occur. We provide three self-contained, but also related, chapters (Chapters 1, 2 and 3), investigating various ways in which decision-makers deviate from 'optimal' time allocation behaviour. The work places itself at the intersection between literature surrounding *scarcity theory* and *choice bracketing*. Across three chapters, the key questions asked were: *'How robust is the mere urgency effect'* (Chapter 1), *'Does narrow bracketing in time allocation exist? And if so, how robust is it?'* (Chapter 2), and *'To what extent do urgency and busyness induce narrow bracketing in time allocation?'* (Chapter 3). The main takeaways and contributions relative to each chapter are summarised in the following paragraphs.

In Chapter 1, the key results from *four* replications show that while the *mere urgency effect* (Zhu et al., 2018) attenuated significantly after controlling for hypothesised confounds, it remained robust. Thus, our research confirms the striking result that people tend to pursue tasks characterised by a sense of urgency, even though these tasks are monetarily suboptimal.

We followed a three-stage sequential replication methodology: (i) re-analysed the data from the original study, (ii) replicated it within the original paradigm used to study it, and (iii) probed at the robustness of the effect. In stage (ii), our core aim was to provide what we considered the 'fairest' test of the mere urgency effect. Ultimately, while we made several important changes to the original design (i.e., randomised options, reduced scope for misunderstanding, removed redundant features) in pursuit of the fairest test, the most important was to provide 'informational balance' of attributes required for an *informed* choice at the moment of decision. The subsequent increase in optimal behaviour from this change highlights that the framing of the decision environment (i.e., how salient is the dominated nature of the suboptimal task) plays a crucial role in choice. Another interesting finding was that while we technically did not manipulate the level of spurious urgency in any of our replications, we observed changes in relative attentional focus on timings versus payoffs. This indicates that what people pay attention to and ultimately choose can be influenced by the relative prominence of specific attributes in the decision environment. Nevertheless, our results showed that spurious urgency shifts attention towards time over payoffs, underpinning the *attention-based psychological account* as a mechanism behind the effect (Zhu et al., 2018). Beyond the core contribution of adding to the evidential base of this novel phenomenon, this work also contributes to the growing literature surrounding the importance of conducting rigorous replication studies in the social sciences.

In Chapter 2, the crucial findings from *three* studies showed overwhelmingly strong evidence for the existence of narrow bracketing in time allocation as well as its robustness. These results highlight that numerous people evaluate the consequences of their time allocation choices separately (narrow bracket), rather than jointly (broad bracket). By implication, these individuals take insufficient account of the *interdependencies* in their time-use decisions, resulting in substantial efficiency losses.

To measure choice bracketing experimentally, we first developed a simple theoretical framework that predicts a separation between broad and narrow bracketers in time allocation. We then operationalised this theoretical framework in an experimental design. After finding surprisingly high levels of narrow bracketing behaviour in our first study, we shifted to investigating the robustness of the result in our second one. This involved increasing the foregone income from working on the suboptimal option. We found that narrow bracketing was fairly insensitive to such manipulations. Consequently, in our third study, we explored boundary conditions at which suboptimal behaviour occurred. An important insight from this endeavour is that while payoff considerations matter for bracketing, the restriction placed on tasks (a form of scarcity) also drives suboptimal choice. Finally, we pooled all the data together and found that attention, comprehension, cognitive ability and specific demographics correlate significantly with suboptimal time allocation. These results, to our understanding, are the first to empirically examine the extent of narrow bracketing in time-use in a controlled experimental environment. Moreover, we contribute to the literature by showing how bracketing relates to measures of cognition and demographics.

In Chapter 3, the key results from an experiment showed that experimentally manipulating *urgency* and *busyness* (two forms of scarcity) does not appear to induce narrow bracketing in time-use decisions. That said, we found that *perceptions* of urgency and busyness are significantly correlated with narrow bracketing. Viewing the spuriously urgent option as more urgent or perceiving oneself as busier in the task are significantly associated with higher levels of narrow bracketing However, they also show that shifting such perceptions is non-trivial.

To investigate how these two forms of scarcity affect narrow bracketing, we utilised the novel design we developed in Chapter 2. We manipulated urgency and busyness by (i) changing the time at which particular tasks were available and (ii) the extent to which subjects were short of time. A core strength of our experimental design is that we could manipulate such elements while holding income-generating capacity across manipulations constant, allowing us to measure bracketing behaviour consistently across conditions. Moreover, since Mullainathan & Shafir (2013) argue that it is *feelings* of scarcity which ultimately shift behaviour, another strength of our framework is that we built in manipulation checks surrounding *perceptions* of urgency and busyness. Overall, we learn that perceptions of scarcity are an important correlate of time allocation behaviour, and that such perceptions are difficult to shift in our experimental environment. These results suggest a possible causal link between perceptions and narrow bracketing behaviour that ultimately needs further research. If such a link exists, shifting perceptions of scarcity should also move time allocation behaviour accordingly. Overall, we stress that these results are not inconsistent with the psychology of scarcity that Mullainathan & Shafir (2013) propose, nor are they inconsistent with the mere urgency effect (Zhu et al., 2018). To our knowledge, this work is the first to investigate the impact of such forms of time-scarcity on suboptimal time-use behaviour in the context of bracketing. Our framework offers an important springboard for future research investigating such scarcity effects on suboptimal time-use behaviour.

While we provide in-depth discussions surrounding the results and areas for future research, a few points surround the chapters as a collective. One finding that has consistently surprised us is the sheer volume of narrow bracketing in time-use observed in Chapters 2 and 3. In Chapter

2, we provided substantial pecuniary incentives to decrease narrow bracketing, yet it remained robust. Moreover, in Chapter 3, we observed that subjects' narrow bracketing appeared to be focusing more granularly on *each* input in the tasks, often at the neglect of broader requirements. At the start of our work concerning bracketing behaviour, we began with a conception of 'narrow' and 'broad' bracketing (see our theoretical framework in Section 2.3). However, the results show that people are perhaps bracketing even more narrowly than we had envisaged.[1] An interesting possible direction for future research would be to explore *exactly* that which subjects are narrow bracketing on.[2] Nevertheless, these points underpin the fact that bracketing is a pervasive phenomenon in decision environments involving time allocation.

Another common theme across Chapters 1, 2 and 3 is the consistent role that attentional focus plays in choice behaviour. There is now a growing body of research surrounding how attention affects judgement and decision-making (Kahneman, 1973; Cohen et al., 1993; Styles, 2006; Bordalo et al., 2013; Caplin, 2016; Farmer & Matlin, 2019; Gabaix, 2019). While the findings across several studies presented in this thesis show that attention is a key correlate of behaviour, it was largely beyond the scope of this project to more deeply consider the role that attention plays in choice behaviour. That said, we suspect that there is much more to be learnt about attention and its relevance to choice in our set-up. Gabaix (2019, p.1) discusses that *"...a central, unifying theme for much of behavioural economics"* is "*inattention*", providing a literature review of several models of attention, methods and challenges in measuring it, behavioural and Bayesian theories surrounding it, and possible applications. We hope that future work might be able to more directly connect attention to the role of choice in the decision environments presented in this research.

Overall, numerous insights have been learned from the research presented in this thesis. However, perhaps the *key* takeaway we hope to depart with is that although 'time' plays an essential role in our lives, with evidence to suggest increasingly so, we surprisingly appear to be quite deficient at knowing how to use it optimally. In perhaps the first time-management self-help book, *On the Shortness of Life*, Seneca wrote in the first century, *"People are frugal in guarding their personal property; but as soon as it comes to squandering time they are most wasteful of the one thing in which it is right to be stingy."*[3] Once we appreciate this and understand where we might fall short in allocating our time, we can begin to recognise ways in which we can more effectively manage it.

[1] We began with the conception that narrow and broad bracketers think differently about *reserved* and *unreserved time*. That is, of the total time allowed while working (6 minutes), 2 minutes were *unreserved* and could be spent on either activity A or B. However, the remaining 4 minutes were *reserved* for activity B. We conjectured that broad bracketers consider the total available time and choose to allocate themselves to maximise income across the two activities. On the other hand, narrow bracketers consider unreserved and reserved time separately, choosing to maximise income in the unreserved time, taking as given income generated from the reserved time. However, it may be that narrow bracketers are rather thinking *'How do I make money from the next set (or the next few sets)?'* This would be consistent with narrow bracketing on the *piece-rate* returns to activity A rather than on income generated in unreserved time.

[2] This could be achieved through treatments that remove the time restriction placed on the suboptimal activity, but maintain higher piece-rate price levels associated with it. For example, in Chapter 2, we had $\mu = 1$, $\alpha = 0.5$. However, it would be interesting to explore dominated choice within $\mu = 1$, $\alpha = 7$. This would also allow one to separate the restriction effect from bracketing on individual piece-rate returns.

[3] This text is from Seneca et al. (2004, p.4), providing modern translations of the original work.

References

Aaron, E. & Spivey-Knowlton, M. (1998). *Frequency vs. probability formats: Framing the three doors problem.* Technical report, Cornell University.

Abeler, J. & Marklein, F. (2017). Fungibility, labels, and consumption. *Journal of the European Economic Association*, 15(1), 99–127.

Alempaki, D., Canic, E., Mullett, T. L., Skylark, W. J., Starmer, C., Stewart, N., & Tufano, F. (2019). Reexamining how utility and weighting functions get their shapes: A quasi-adversarial collaboration providing a new interpretation. *Management Science*, 65(10), 4841–4862.

Anderson, N. H. (1973). Serial position curves in impression formation. *Journal of Experimental Psychology*, 97(1), 8.

Andreoli-Versbach, P. & Mueller-Langer, F. (2014). Open access to data: An ideal professed but not practised. *Research Policy*, 43(9), 1621–1633.

Andreoni, J., Gravert, C., Kuhn, M. A., Saccardo, S., & Yang, Y. (2018). *Arbitrage or narrow bracketing? On using money to measure intertemporal preferences.* Technical report, National Bureau of Economic Research.

Arrow, K. J. (1966). Exposition of the theory of choice under uncertainty. *Synthese*, (pp. 253–269).

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3), 258.

Atkinson, J. W. (1953). The achievement motive and recall of interrupted and completed tasks. *Journal of Experimental Psychology*, 46(6), 381.

Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559.

Barberis, N., Huang, M., & Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American economic review*, 96(4), 1069–1090.

Barrero, J. M., Bloom, N., & Davis, S. J. (2021). *Why working from home will stick.* Technical report, National Bureau of Economic Research.

Becker, G. S. (1965). A theory of the allocation of time. *The economic journal*, 75(299), 493–517.

Benartzi, S. & Thaler, R. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic perspectives*, 21(3), 81–104.

Benartzi, S. & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics*, 110(1), 73–92.

Benartzi, S. & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American economic review*, 91(1), 79–98.

Berlyne, D. E. (1954). A theory of human curiosity. *British journal of psychology*, 45(3), 180.

Berlyne, D. E. (1969). The development of the concept of attention in psychology. In *Attention in neurophysiology* (pp. 1–26). Appleton-Century-Crofts New York.

Berlyne, D. E. (1978). Curiosity and learning. *Motivation and emotion*, 2(2), 97–175.

Bernheim, B. D., DellaVigna, S., & Laibson, D. (2019). *Handbook of Behavioral Economics-Foundations and Applications 2*. Elsevier.

Binkley, J. K. & Bejnarowicz, J. (2003). Consumer price awareness in food shopping: the case of quantity surcharges. *Journal of Retailing*, 79(1), 27–35.

Birnbaum, M. H. & Bahra, J. P. (2007). Gain-loss separability and coalescing in risky decision making. *Management Science*, 53(6), 1016–1028.

Bitterly, T. B., VanEpps, E., & Schweitzer, M. E. (2020). Exponential numeracy. *Available at SSRN 3746327*.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5), 803–843.

Botti, S., Broniarczyk, S., Häubl, G., Hill, R., Huang, Y., Kahn, B., Kopalle, P., Lehmann, D., Urbany, J., & Wansink, B. (2008). Choice under restrictions. *Marketing Letters*, 19(3-4), 183–199.

Brehm, J. W. (1966). A theory of psychological reactance. *Academic press*.

Brendl, C. M., Markman, A. B., & Messner, C. (2003). The devaluation effect: Activating a need devalues unrelated objects. *Journal of Consumer Research*, 29(4), 463–473.

Brock, T. C. (1968). Implications of commodity theory for value change. In *Psychological foundations of attitudes* (pp. 243–275). Elsevier.

Brock, T. C. & Brannon, L. A. (1992). Liberalization of commodity theory. *Basic and Applied Social Psychology*, 13(1), 135–144.

Brown, J. R., Kling, J. R., Mullainathan, S., & Wrobel, M. V. (2008). Why don't people insure late-life consumption? a framing explanation of the under-annuitization puzzle. *American Economic Review*, 98(2), 304–09.

Budd, J. W. (2018). Employment with a human face. In *Employment with a Human Face*. Cornell University Press.

Buonomano, D. (2017). *Your brain is a time machine: The neuroscience and physics of time*. WW Norton & Company.

Camerer, C. (1999). Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences*, 96(19), 10575–10577.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2), 407–441.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644.

Campitelli, G. & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? a mathematical modeling approach. *Memory & cognition*, 42(3), 434–447.

Cannon, C., Goldsmith, K., & Roux, C. (2019). A self-regulatory model of resource scarcity. *Journal of Consumer Psychology*, 29(1), 104–127.

Caplin, A. (2016). Measuring and modeling attention. *Annual Review of Economics*, 8, 379–403.

Carvalho, L. S., Meier, S., & Wang, S. W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American economic review*, 106(2), 260–84.

Chaiken, S. & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.

Chanel, O. & Chichilnisky, G. (2009). The influence of fear in decisions: Experimental evidence. *Journal of Risk and Uncertainty*, 39(3), 271–298.

Chanel, O. & Chichilnisky, G. (2013). Valuing life: Experimental evidence using sensitivity to rare events. *Ecological Economics*, 85, 198–205.

Choi, J. J., Laibson, D., & Madrian, B. C. (2009). Mental accounting in portfolio choice: Evidence from a flypaper effect. *American Economic Review*, 99(5), 2085–95.

Chong, A. (2020). In-depth: Time consciousness and discipline in the industrial revolution. *SJW*.

Cialdini, R. B. (1993). Influence: The psychology of persuasion (rev. ed.). *New York: Morrow*.

Cialdini, R. B. (2009). *Influence: Science and practice*, volume 4. Pearson education Boston.

Cohen, R. A., Sparling-Cohen, Y. A., & O'Donnell, B. F. (1993). *The neuropsychology of attention*. Springer.

Comer, J. M. (2019). *The ruthless elimination of hurry: How to stay emotionally healthy and spiritually alive in the chaos of the modern world*. WaterBrook.

Covey, S. R. (2004). *The 7 habits of highly effective people: Powerful lessons in personal change*. Simon and Schuster.

Czibor, E., Jimenez-Gomez, D., & List, J. A. (2019). The dozen things experimental economists should do (more of). *Southern Economic Journal*, 86(2), 371–432.

Darrah, C. (2007). *Busier than ever!: Why American families can't slow down*. Stanford University Press.

Davies, A. (2019). The industrial revolution and time. *The Open University*.

De Bruijn, E.-J. & Antonides, G. (2021). Poverty and economic decision making: a review of scarcity theory. *Theory and Decision*, (pp. 1–33).

Dingel, J. I. & Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189, 104235.

Dupas, P., Robinson, J., & Saavedra, S. (2014). The daily grind: Cash needs, labor supply and self-control. *Unpublished paper*, (pp. 1027–1047).

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What is meant by" replication" and why does it encounter resistance in economics? *American Economic Review*, 107(5), 46–51.

Eisendrath, B. (1991). What makes time tick, or has the industrial revolution really made clocks go faster? *Office of Elementary and Secondary Education, Smithsonian Institution, Washington, D.C.*

Ellis, A. & Freeman, D. J. (2020). Revealing choice bracketing. *arXiv preprint arXiv:2006.14869*.

Epstein, C. F. & Kalleberg, A. L. (2004). *Fighting for time: Shifting boundaries of work and social life*. Russell Sage Foundation.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American psychologist*, 49(8), 709.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4), 1645–1692.

Fallucchi, F. & Kaufmann, M. (2021). Narrow bracketing in work choices. *arXiv preprint arXiv:2101.04529*.

Farmer, T. A. & Matlin, M. W. (2019). *Cognition*. John Wiley & Sons.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, 13(1), 1–17.

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

Friedman, D. (1998). Monty hall's three doors: Construction and deconstruction of a choice anomaly. *The American Economic Review*, 88(4), 933–946.

Fudenberg, D. & Levine, D. K. (2006). A dual-self model of impulse control. *American economic review*, 96(5), 1449–1476.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2 (pp. 261–343). Elsevier.

Gagné, M. & Bhave, D. (2011). Autonomy in the workplace: An essential ingredient to employee engagement and well-being in every culture. In *Human autonomy in cross-cultural context* (pp. 163–187). Springer.

Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). Lioness lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, 6(1), 95–111.

Gierl, H. & Huettl, V. (2010). Are scarce products always more attractive? the interaction of different types of scarcity signals with products' suitability for conspicuous consumption. *International Journal of Research in Marketing*, 27(3), 225–235.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological review*, 102(4), 684.

Gneezy, U. & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The quarterly journal of economics*, 112(2), 631–645.

Goldin, J. & Homonoff, T. (2013). Smoke gets in your eyes: cigarette tax salience and regressivity. *American Economic Journal: Economic Policy*, 5(1), 302–36.

Goodie, A. S. & Fantino, E. (1999). What does and does not alleviate base-rate neglect under direct experience. *Journal of Behavioral Decision Making*, 12(4), 307–335.

Goswami, I. & Urminsky, O. (2015). More time, more work: How time limits bias estimates of project duration and scope. *More Work: How Time Limits Bias Estimates of Project Duration and Scope*.

Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. *Psychological foundations of attitudes*, (pp. 147–170).

Hamermesh, D. S. & Lee, J. (2007). Stressed out on four continents: Time crunch or yuppie kvetch? *The review of Economics and Statistics*, 89(2), 374–383.

Hamilton, R., Thompson, D., Bone, S., Chaplin, L. N., Griskevicius, V., Goldsmith, K., Hill, R., John, D. R., Mittal, C., O'Guinn, T., et al. (2019a). The effects of scarcity on consumer decision journeys. *Journal of the Academy of Marketing Science*, (pp. 1–19).

Hamilton, R. W., Mittal, C., Shah, A., Thompson, D. V., & Griskevicius, V. (2019b). How financial constraints influence consumer behavior: An integrative framework. *Journal of Consumer Psychology*, 29(2), 285–305.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–14).

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford publications.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan Jr, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of behavioral decision making*, 6(3), 149–185.

Herrnstein, R. J. & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives*, 5(3), 137–156.

Herrnstein, R. J. & Prelec, D. (1992a). *Melioration.* New York: Russell Sage Foundation.

Herrnstein, R. J. & Prelec, D. (1992b). A theory of addiction. *Choice over time. New York: Sage.*

Heyman, G. M. (1996). Resolving the contradictions of addiction. *Behavioral and Brain sciences*, 19(4), 561–574.

Hilgers, D. & Wibral, M. (2014). How malleable are choice brackets? the case of myopic loss aversion. *University of Bonn, mimeo.*

Hochschild, A. R. & Arlie, H. R. (1997). *The time bind: When work becomes home and home becomes work.* Macmillan.

Hogarth, R. M. & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1), 1–55.

Huijsmans, I., Ma, I., Micheli, L., Civai, C., Stallen, M., & Sanfey, A. G. (2019). A scarcity mindset alters neural processing underlying consumer decision making. *Proceedings of the National Academy of Sciences*, 116(24), 11699–11704.

Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review*, 39(1), 25.

Hunnicutt, B. (1988). *Work without end: Abandoning shorter hours for the right to work.* Temple University Press.

Johar, G., Meng, R., & Wilcox, K. (2015). Thinking about financial deprivation: Rumination and decision making among the poor. *ACR North American Advances.*

Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of risk and uncertainty*, 7(1), 35–51.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on brandstätter, gigerenzer, and hertwig (2006).

Juster, F. T. & Stafford, F. P. (1991). The allocation of time: Empirical findings, behavioral models, and problems of measurement. *Journal of Economic literature*, 29(2), 471–522.

Kahneman, D. (1973). *Attention and effort*, volume 1063. Citeseer.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49, 81.

Kahneman, D. & Frederick, S. (2004). Attribute substitution in intuitive judgment. *Models of a man: Essays in memory of Herbert A. Simon*, (pp. 411–432).

Kahneman, D. & Knetsch, J. L. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of environmental economics and management*, 22(1), 57–70.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, 98(6), 1325–1348.

Kahneman, D. & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management science*, 39(1), 17–31.

Kahneman, D. & Tversky, A. (1981). *The simulation heuristic.* Technical report, Stanford Univ CA Dept of Psychology.

Kalil, A., Mayer, S., & Shah, R. (2022). Scarcity and inattention. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2022-76).

Keren, G. & Wagenaar, W. A. (1987). Violation of utility theory in unique and repeated gambles. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 387.

Keynes, J. M. (1930). Economic possibilities for our grandchildren. In *Essays in persuasion* (pp. 321–332). Springer.

Kidd, C. & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, 88(3), 449–460.

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, 43(1), 39–58.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., et al. (2014). Investigating variation in replicability: A “many labs” replication project. *Social psychology*, 45(3), 142.

Koch, A. K. & Nafziger, J. (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, 162, 305–351.

Koch, A. K. & Nafziger, J. (2019). Correlates of narrow bracketing. *The Scandinavian Journal of Economics*, 121(4), 1441–1472.

Koch, A. K. & Nafziger, J. (2020). Motivational goal bracketing: An experiment. *Journal of Economic Theory*, 185, 104949.

König, C. J. & Kleinmann, M. (2004). Business before pleasure: no strategy for procrastinators? *Personality and Individual Differences*, 37(5), 1045–1057.

Kourouxous, T. & Bauer, T. (2019). Violations of dominance in decision-making. *Business Research*, 12(1), 209–239.

Krauss, S. & Wang, X.-T. (2003). The psychology of the monty hall problem: discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, 132(1), 3.

Kristofferson, K., McFerran, B., Morales, A. C., & Dahl, D. W. (2016). The dark side of scarcity promotions: How exposure to limited-quantity promotions can induce aggression. *Journal of Consumer Research*, 43(5), 683–706.

Kumar, A. & Lim, S. S. (2008). How do decision frames influence the stock investment choices of individual investors? *Management science*, 54(6), 1052–1064.

Kurtz, J. L. (2008). Looking to the future to appreciate the present: The benefits of perceived temporal scarcity. *Psychological Science*, 19(12), 1238–1241.

Levine, R. N. (2008). *A geography of time: On tempo, culture, and the pace of life.* Basic Books.

Levitt, S. D. & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1), 1–18.

Li, S. (2001). Extended research on dominance violations in similarity judgments: The equate-to-differentiate interpretation. *The International Journal of Creativity & Problem Solving*, 11(1), 13–38.

Linder, S. B. (1970). *The harried leisure class.* New York: Columbia University Press.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical decision making*, 21(1), 37–44.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1), 75.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological bulletin*, 127(2), 267.

Lynn, M. (1991). Scarcity effects on value: A quantitative review of the commodity theory literature. *Psychology & Marketing*, 8(1), 43–57.

Maćkowiak, B., Matějka, F., & Wiederholt, M. (2021). Rational inattention: A review. *ECB Working Paper.*

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *science*, 341(6149), 976–980.

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306(5695), 503–507.

Mehta, R. & Zhu, M. (2015). Creating when you have less: The impact of resource scarcity on product use creativity. *Journal of Consumer Research*, 42(5), 767–782.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2), 81.

Mueller-Langer, F., Fecher, B., Harhoff, D., & Wagner, G. G. (2019). Replication studies in economics—how many and which papers are chosen for replication, and why? *Research Policy*, 48(1), 62–83.

Mullainathan, S. & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

O'Donoghue, T. & Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, 116(1), 121–160.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology*, 45(4), 867–872.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5), 411–419.

Pearson, H. (2015). The time lab: why does modern life seem so busy? an oxford centre is trying to find answers with the world's biggest collection of time-use diaries. *Nature*, 526(7574), 492–497.

Perlow, L. A. (1999). The time famine: Toward a sociology of work time. *Administrative science quarterly*, 44(1), 57–81.

Pribram, K. H. & McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological review*, 82(2), 116.

Rabin, M. & Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *American Economic Review*, 99(4), 1508–43.

Rachlin, H. (1995). Self-control: Beyond commitment. *Behavioral and brain sciences*, 18(1), 109–121.

Read, D. & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, 1(1), 34.

Read, D., Loewenstein, G., & Kalyanaraman, S. (1999a). Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic. *Journal of Behavioral Decision Making*, 12(4), 257–273.

Read, D., Loewenstein, G., Rabin, M., Keren, G., & Laibson, D. (1999b). Choice bracketing. In *Elicitation of preferences* (pp. 171–202). Springer.

Redelmeier, D. A. & Tversky, A. (1992). On the framing of multiple prospects. *Psychological Science*, 3(3), 191–193.

Reio Jr, T. G., Petrosko, J. M., Wiswell, A. K., & Thongsukmag, J. (2006). The measurement and conceptualization of curiosity. *The Journal of Genetic Psychology*, 167(2), 117–135.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2), 900–926.

Roux, C., Goldsmith, K., & Bonezzi, A. (2015). On the psychology of scarcity: When reminders of resource scarcity promote selfish (and generous) behavior. *Journal of consumer research*, 42(4), 615–631.

Sabini, John amd Silver, M. (1982). Moralities of everyday life. *Oxford: Oxford University Press.*

Samuelson, P. (1963). Risk and uncertainty: A fallacy of large numbers. *scientia*, 57(98).

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, 26(5), 442–450.

Seneca, L. A. et al. (2004). *On the shortness of life*, volume 1. Penguin UK.

Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107), 682–685.

Shah, A. K., Mullainathan, S., & Shafir, E. (2019). An exercise in self-replication: Replicating shah, mullainathan, and shafir (2012). *Journal of Economic Psychology*, 75, 102127.

Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological science*, 26(4), 402–412.

Shah, A. K., Zhao, J., Mullainathan, S., & Shafir, E. (2018). Money in the mental lives of the poor. *Social Cognition*, 36(1), 4–19.

Sharma, E. & Alter, A. L. (2012). Financial deprivation prompts consumers to seek scarce goods. *Journal of Consumer Research*, 39(3), 545–560.

Sheehan, K. B. (2018). Crowdsourcing research: data collection with amazon's mechanical turk. *Communication Monographs*, 85(1), 140–156.

Shefrin, H. M. & Thaler, R. H. (1988). The behavioral life-cycle hypothesis. *Economic inquiry*, 26(4), 609–643.

Shin, J. & Ariely, D. (2004). Keeping doors open: The effect of unavailability on incentives to keep options viable. *Management science*, 50(5), 575–586.

Simon, H. A. (1957). Models of man; social and rational. *Wiley.*

Simon, H. A. (1979). Rational decision making in business organizations. *The American economic review*, 69(4), 493–513.

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing research*, 27(2), 150–162.

Simonson, I. & Winer, R. S. (1992). The influence of purchase quantity and display format on consumer preference for variety. *Journal of Consumer Research*, 19(1), 133–138.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1), 3.

Spiller, S. A. (2011). Opportunity cost consideration. *Journal of Consumer Research*, 38(4), 595–610.

Stanovich, K. E. & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and brain sciences*, 23(5), 645–665.

Starmer, C. (1999). Cycling with rules of thumb: An experimental test for a new form of non-transitive behaviour. *Theory and Decision*, 46(2), 139–157.

Stillwell, D. J. & Tunney, R. J. (2009). Melioration behaviour in the harvard game is reduced by simplifying decision outcomes. *Quarterly Journal of Experimental Psychology*, 62(11), 2252–2261.

Stracke, R., Kerschbamer, R., & Sunde, U. (2015). Prevalence and determinants of choice bracketing-experimental evidence.

Stracke, R., Kerschbamer, R., & Sunde, U. (2017). Coping with complexity–experimental evidence for narrow bracketing in multi-stage contests. *European Economic Review*, 98, 264–281.

Styles, E. (2006). *The psychology of attention.* Psychology Press.

Sunstein, C. R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty*, 26(2), 121–136.

Sunstein, C. R. & Zeckhauser, R. (2011). Overreaction to fearsome risks. *Environmental and Resource Economics*, 48(3), 435–449.

Svenson, O. & Edland, A. (1987). Change of preferences under time pressure: Choices and judgements. *Scandinavian Journal of Psychology*, 28(4), 322–330.

Svenson, O., Edland, A., & Slovic, P. (1990). Choices and judgments of incompletely described decision alternatives under time pressure. *Acta Psychologica*, 75(2), 153–169.

Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The quarterly journal of economics*, 112(2), 647–661.

Thompson, D. (2016). A brief economic history of time. *The Atlantic.*

Thompson, E. P. (1967). Time, work-discipline, and industrial capitalism. *Class: The Anthology*, (pp. 27–40).

Thomson, K. S. & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, 11(1), 99.

Tomm, B. M. & Zhao, J. (2016). Scarcity captures attention and induces neglect: Eyetracking and behavioral evidence. In *CogSci* (pp. 1199–1204).

Tunney, R. J. (2006). Short article: Preference reversals are diminished when gambles are presented as relative frequencies. *Quarterly Journal of Experimental Psychology*, 59(9), 1516–1523.

Tversky, A. & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In *Behavioral decision making* (pp. 25–41). Springer.

Tversky, A. & Kahneman, D. (1986). Judgment under uncertainty: Heuristics and biases.

Tversky, A. & Kahneman, D. (1989). Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81–126). Springer.

Verhallen, T. M. & Robben, H. S. (1994). Scarcity and preference: An experiment on unavailability and product evaluation. *Journal of economic psychology*, 15(2), 315–331.

Vorjohann, P. (2020). Reference-dependent choice bracketing. *Working Paper.*

Wajcman, J. (2014). Pressed for time. In *Pressed for Time.* University of Chicago Press.

Wang, X. T. (1996). Domain-specific rationality in human choices: Violations of utility axioms and social contexts. *Cognition*, 60(1), 31–63.

Warry, C. J., Remington, B., & Sonuga-Barke, E. J. (1999). When more means less: Factors affecting human self-control in a local versus global choice paradigm. *Learning and motivation*, 30(1), 53–73.

Welsh, M., Burns, N., & Delfabbro, P. (2013). The cognitive reflection test: How much more than numerical ability? In *Proceedings of the Annual Meeting of the Cognitive Science society*, volume 35.

Wickham, P. A. (2003). The representativeness heuristic in judgements involving entrepreneurial success and failure. *Management Decision*.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Worchel, S., Lee, J., & Adewole, A. (1975). Effects of supply and demand on ratings of object value. *Journal of personality and social psychology*, 32(5), 906.

Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 11(6), 495–506.

Zeigarnik, B. (1938). On finished and unfinished tasks.

Zhao, J. & Tomm, B. M. (2018). Psychological responses to scarcity. In *Oxford research encyclopedia of psychology*.

Zhu, M., Bagchi, R., & Hock, S. J. (2019). The mere deadline effect: Why more time might sabotage goal pursuit. *Journal of consumer research*, 45(5), 1068–1084.

Zhu, M. & Ratner, R. K. (2015). Scarcity polarizes preferences: The impact on choice among multiple items in a product class. *Journal of Marketing Research*, 52(1), 13–26.

Zhu, M., Yang, Y., & Hsee, C. K. (2018). The mere urgency effect. *Journal of Consumer Research*, 45(3), 673–690.

Zur, H. B. & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica*, 47(2), 89–104.

APPENDIX A

Chapter I

A.1 Instructions

[The experiments were programmed on Qualtrics (all material available on upon request). The Two dotted lines represent a new screen in the programme. Comments in footnotes and square brackets were not seen by readers. Red indicate treatment differences or differences across experiments. For example, in Section A.1.1, red indicates additional questions that we asked, which were not in the original Experiment IIB by (Zhu et al., 2018).]

A.1.1 Experiment IIB & Replication I

Preliminary instructions before treatment assignment

Thank you for participating in our study. By completing this survey or questionnaire, you are consenting to be in this research study. Your participation is voluntary and you can stop at any time.

It is very important that you read and follow the instructions carefully.

As survey-data is useless when participants do not read and follow instructions correctly, you will only be able to finish the survey and get paid when you follow the instructions.

Thank you!

. . . .

. . . .

Thank you for participating in this survey. Please read the text below and answer the question.

Most modern theories of decision-making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables, can greatly impact the decision process. In order to facilitate our research on decision-making we are interested in knowing certain factors about you, the decision maker. Specifically, we are interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the sports items below, type the word inside the quotation marks: 'typing' in the blank box, and proceed to the next screen by clicking on the 'NEXT' button. Thank you very much.

Which of these activities do you engage most regularly?

- skiing
- soccer
- snowboarding

- running
- hockey
- football
- swimming
- tennis
- cycling
- Other

....
....

Instructions

We are interested in studying how people choose between different tasks.

- In this study, you will choose one of two tasks to take, Task A or Task B.
- Each participant is only allowed to take ONE task (i.e., either Task A or Task B, but not both) in this study.
- Additionally, as clearly indicated in the sign-up page, each participant is only allowed to participate in the study ONCE.

Please click on 'CONTINUE' to read more about the two tasks.

Control (Pure Replication)

Instructions

In this study, you will choose one of two tasks to take, Task A or Task B.

Task A and Task B require you to perform the SAME activity and DO NOT differ in terms of difficulty or length. In both tasks, you will be asked to type some randomly generated 6-letter string (e.g., 'rlgows') in the reversed order (e.g., 'swoglr'). More specifically, you will be presented with 100 strings and you will be given a total of 3 minutes and asked to type as many strings as possible within the 3-minute window.

However, these two tasks differs in their bonus level. Specifically, one of the tasks will give you a bonus of 12 cents and the other task will give you a bonus of 16 cents. The bonus level of each task will be randomly determined by a randomizer on the next screen. And the bonus will be added to your Mturk payment after you complete the study.

Both tasks will expire in 50 minutes.

Before you proceed to the next screen, please answer the following questions to indicate that you understand the instructions.

1. How many cents will you get for completing these two tasks?
    - Both tasks will give me a bonus of 12 cents.
    - Both tasks will give me a bonus of 16 cents.
    - One of the task will give me a bonus of 12 cents and the other will give me a bonus of 16 cents. The final bonus level of each task will be determined by a randomizer.
    - One of the task will give me a bonus of 12 cents and the other will give me a bonus of 16 cents. The final bonus level of each task will be determined based on my performance.
    - It was not specified.
2. Task A and Task B require you to perform the SAME activity and DO NOT differ in terms of difficulty or length. Could you specify what is the activity you need to perform in these two tasks?
    - In both tasks, I will be given six 100-letter strings and a total of 5 minutes. My job is to type as many strings as possible within the 3-minute time window.
    - In both tasks, I will be given six 100-letter strings and a total of 3 minutes. My job is to type as many strings as possible within the 3-minute time window.
    - In both tasks, I will be given 100 six-letter strings and a total of 5 minutes. My job is to type as many strings as possible within the 5-minute time window.
    - In both tasks, I will be given 100 six-letter strings and a total of 3 minutes. My job is to type as many strings as possible within the 3-minute time window.
    - It was not specified.

Please click CONTINUE to let the system randomly decide the bonus level of each task.

System processing, please wait.

The system has randomly assigned the bonus level for Task A and Task B.

You can earn a bonus of 12 cents by completing Task A.

You can earn a bonus of 16 cents by completing Task B.

Both tasks will take exactly 3 minutes to complete and expire in 50 minutes, and they do not differ in terms of difficulty. Please choose below which task you would like to take.

- Task A
- Task B

Urgency (Pure Replication)

Instructions

In this study, you will choose one of two tasks to take, Task A or Task B.

Task A and Task B require you to perform the SAME activity and DO NOT differ in terms of difficulty or length. In both tasks, you will be asked to type some randomly generated 6-letter string (e.g., 'rlgows') in the reversed order (e.g., 'swoglr'). More specifically, you will be presented with 100 strings and you will be given a total of 3 minutes and asked to type as many strings as possible within the 3-minute window.

However, these two tasks differ in their bonus level and availability. Specifically, one of the tasks will give you a bonus of 12 cents and the other task will give you a bonus of 16 cents. The bonus level of each task will be randomly determined by a randomizer on the next screen. And the bonus will be added to your MTurk payment after you complete the study.

Tasks A and B also differ in their availability. One of the tasks will expire in 5 minutes, whereas the other task will expire in 50 minutes. The availability of each task will be determined by the system using a second randomizer.

Before you proceed to the next screen, please answer the following questions to indicate that you understand the instructions.

1. How many cents will you get for completing these two tasks?
   - Both tasks will give me a bonus of 12 cents.
   - Both tasks will give me a bonus of 16 cents.
   - One of the task will give me a bonus of 12 cents and the other will give me a bonus of 16 cents. The final bonus level of each task will be determined by a randomizer.
   - One of the task will give me a bonus of 12 cents and the other will give me a bonus of 16 cents. The final bonus level of each task will be determined based on my performance.
   - It was not specified.
2. How long will the two tasks be available?
   - Both tasks will expire after 5 minutes.
   - Both tasks will expire after 50 minutes.
   - One of the task will expire after 5 minutes and the other will expire after 50 minutes. The final availability of each task will be determined based on the bonus level.
   - One of the task will expire after 5 minutes and the other will expire after 50 minutes. The final availability of each task will be determined by a randomizer.
   - It was not specified.
3. Task A and Task B require you to perform the SAME activity and DO NOT differ in terms of difficulty or length. Could you specify what is the activity you need to perform in these two tasks?
   - In both tasks, I will be given six 100-letter strings and a total of 5 minutes. My job is to type as many strings as possible within the 3-minute time window.

- In both tasks, I will be given six 100-letter strings and a total of 3 minutes. My job is to type as many strings as possible within the 3-minute time window.
- In both tasks, I will be given 100 six-letter strings and a total of 5 minutes. My job is to type as many strings as possible within the 5-minute time window.
- In both tasks, I will be given 100 six-letter strings and a total of 3 minutes. My job is to type as many strings as possible within the 3-minute time window.
- It was not specified.

Please click CONTINUE to let the system randomly decide the bonus level of each task.

System processing, please wait.

The system has randomly assigned the bonus level for Task A and Task B.

You can earn a bonus of 12 cents by completing Task A.

You can earn a bonus of 16 cents by completing Task B.

Please click on CONTINUE to let the system randomly assign the availability of the two tasks.

System processing, please wait.

The system has completed the random assignment for task availability.

Task A will expire in 5 minutes.

Task B will expire in 50 minutes.

Both tasks will take exactly 3 minutes to complete and they do not differ in terms of difficulty. Please choose below which task you would like to take.

- Task A
- Task B

Task and ex-post questions

You have chosen Task [A/B]. Please click CONTINUE to start working on this task.

. . . .
. . . .

Please type the following randomly generated letter sequences in the reversed order. For example, for the first randomly generated letter sequence 'rlgows', you should enter 'swoglr' in the textbox below. When 3 minutes run out, you will be automatically directed to the next page.

rlgows oiggcj ylqtyc ebmvqi mdusmj
xrqefx mrbuzv cckdjb fhigmu staqym
jzcemu seflna cfjrow pkbpij fxawzz
heqppk bqgpvy losnak gowtxm nkjncw
snqkea luhcyu cmklzf vzvoof dllruf
flesbt nqffcl zrfyfa fpiiwg yoxzmk
fyrhjn yqmsuo bpdqpi rewsab rvwzhw
bdlnvl kvtvrq rqoyyq geigbc gcmsxn
kapjqa nvfxdh cxakgp nasyxl hmyvam
pcqqhy etoghj ewscke qarmly ztrqvv
jtylyx efhzoc fcktqe fgfaul uuvszq
wijdsv elatzp xqdzcv wdgokh slhflz
hlzdyi icbncj aqqoel yxqyva xisbbs
uzdoul dkxlyf ljlxkm kmogbe mvhuqb
eepyvq tojnzr zxqhfo edwpbk qassjz
xohebz ltkuxw dixlrj qspdxs aafdpf
gzwgws imverp ywhxqy quqzxi eszhdc
yddijo dcfkwm cgsnxc uthwal sywonr
fsqzlr qfyllo nfbvzm fqdojb xjfnvb
ngjeqq wogzld fehkjh hscgdc mhntog

. . . .
. . . .

Please briefly explain how you decided which task to work on in this study.

. . . .
. . . .

- While deciding which task to take, to what extent did you pay attention to the bonus amounts?

  (not at all ← 1 to 8 → very much)

- While deciding which task to take, to what extent did you pay attention to the task expiration time?

  (not at all ← 1 to 8 → very much)

1. What was the bonus of Task A in this study?
    - 12 cents
    - 16 cents
    - I don't remember
2. According to the instructions, which task was going to expire in 5 minutes?
    - Task A
    - Task B
    - None of tasks had an expiration time specified
3. Did the two tasks differ in terms of difficulty?
    - Yes
    - No
    - I don't remember

In general, how busy do you think you are?

(not at all ← 1 to 8 → very much)

- To what extent did you feel tired before doing your task?

  (not at all ← 1 to 8 → very much)

- To what extent did you feel tired after doing your task?

  (not at all ← 1 to 8 → very much)

When making your decision, were you aware that you could do the task immediately and end your session without having to wait for your task to expire?

- Yes
- No

. . . .
. . . .

Going back to your decision, which reason is most closely related to your decision? Please choose one.

- I was concerned with bonus amounts
- I was concerned about the task expiration time
- I thought the one task might be more difficult than the other
- I thought choosing one option would motivate me more
- I feared that I would be stuck in the survey for 50-minutes
- The one option would be less stressful
- I just picked the first option
- I picked randomly
- Other

. . . .
. . . .

[An unincentivised measure of patience from the *Global Evidence on Economic Preferences* (Falk et al., 2018). The authors measure these preferences through a combination of two survey measures, one quantitative and another qualitative. The quantitative consists of a series of five interdependent hypothetical binary choices employing the staircase method. The qualitative, on the other hand, is a participant's self-assessed level of a subject's willingness to wait on an 11-point Likert scale.]

1. How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?

   (Not willing at all ← 0 to 10 → Very willing to take risks)

2. Please consider the following hypothetical scenarios. Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present you with 5 different situations. The payment today is the same in each of these situations.The payment in 12 months is different in every situation. For each of the situations, we would like to know which you would choose. Please assume there is no inflation, i.e. future prices are the same as today's prices.

   Situation 1:
   Please consider the following: would you rather receive $100 today or $154 in 12 months?

   - Today
   - 12 months

   .
   .
   .

. . . .
. . . .

1. Gender
    - Male
    - Female
    - Prefer not to say
2. Age
    (numeric input)
3. How would you rate your English language ability?

    (not at all ← 1 to 8 → very much)

4. Going back to your decision, which reason is most closely related to your decision? Please choose one.
    - Less than high school diploma
    - High school diploma or GED
    - Some college, but no degree
    - Associates Degree (for example: AA, AS)
    - Bachelor's Degree (for example: BA, BBA, and BS)
    - Master's Degree (for example: MA, MS, and MEng)
    - Professional Degree (for example: MD, DDS, JD)
    - Doctorate (for example: PhD, EdD)
5. Which of the following income categories best describes your annual household income?
    - Less than $15,000
    - $15,000 to $29,999
    - $30,000 to $49,999
    - $50,000 to $74,999
    - $75,000 to $99,999
    - $100,000 and over
    - Prefer not to say
6. Now, you have finished the survey. Please let us know if you have any comments.
    (text input)

A.1.2 Replication II

[Since the preliminary instructions, task and ex-post questions are identical to EIIB and RI, we only provide instructions around the moment of decision.]

Control (Replication)

Please click CONTINUE to let the system randomly decide the bonus level of each task.

System processing, please wait.

The system has randomly assigned the bonus level for Task A and Task B.

You can earn a bonus of 12 (16) cents by completing Task A.

You can earn a bonus of 16 (12) cents by completing Task B.

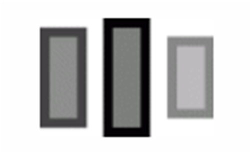
Both tasks will take exactly 3 minutes to complete and expire in 50 minutes, and they do not differ in terms of difficulty. Please choose below which task you would like to take.

- Task A
- Task B

Urgency (Replication)

Please click CONTINUE to let the system randomly decide the bonus level of each task.

System processing, please wait.

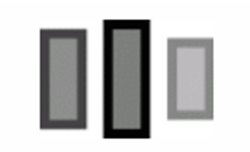


The system has randomly assigned the bonus level for Task A and Task B.

> You can earn a bonus of 12 (16) cents by completing Task A.
>
> You can earn a bonus of 16 (16) cents by completing Task B.

Please click on CONTINUE to let the system randomly assign the availability of the two tasks.

System processing, please wait.



The system has completed the random assignment for task availability.

> Task A will expire in 5 (50) minutes.
>
> Task B will expire in 50 (5) minutes.

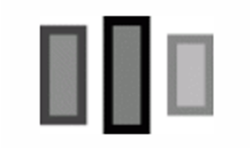
Both tasks will take exactly 3 minutes to complete and they do not differ in terms of difficulty. Please choose below which task you would like to take.

- Task A
- Task B

Urgency (Leave Highlight)

Please click CONTINUE to let the system randomly decide the bonus level of each task.

System processing, please wait.

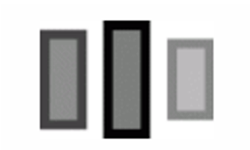


The system has randomly assigned the bonus level for Task A and Task B.

> You can earn a bonus of 12 (16) cents by completing Task A.
>
> You can earn a bonus of 16 (16) cents by completing Task B.

Please click on CONTINUE to let the system randomly assign the availability of the two tasks.

System processing, please wait.



The system has completed the random assignment for task availability.

> Task A will expire in 5 (50) minutes.
>
> Task B will expire in 50 (5) minutes.

Both tasks will take exactly 3 minutes to complete and they do not differ in terms of difficulty. Please choose below which task you would like to take.

(Note: the task expiry times only affect when the tasks are available for you to complete (not how long they will take). Whichever task you choose, it will only take 3-minutes to complete and then you can move on.)[1]

- Task A
- Task B

[1] Note that participants saw this text as red in the experiment. We made this text different as we wanted it to stand out in this treatment manipulation.

Urgency (Saliency)

Please click CONTINUE to let the system randomly decide the bonus level and availability of each task.

. . . .
. . . .

System processing, please wait.

. . . .
. . . .

The system has completed the random assignment for task availability and has randomly assigned the bonus level for Task A and Task B.

> You can earn a bonus of 12 (16) cents by completing Task A.
>
> You can earn a bonus of 16 (16) cents by completing Task B.
>
> Task A will expire in 5 (50) minutes.
>
> Task B will expire in 50 (5) minutes.

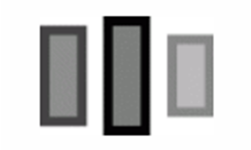
Both tasks will take exactly 3 minutes to complete, and they do not differ in terms of difficulty. Please choose below which task you would like to take.

- Task A
- Task B

Urgency (Reversal)

Please click on CONTINUE to let the system randomly assign the availability of the two tasks.

System processing, please wait.



The system has completed the random assignment for task availability.

> Task A will expire in 5 (50) minutes.
>
> Task B will expire in 50 (5) minutes.

Please click CONTINUE to let the system randomly decide the bonus level of each task.

System processing, please wait.

The system has randomly assigned the bonus level for Task A and Task B.

> You can earn a bonus of 12 (16) cents by completing Task A.
>
> You can earn a bonus of 16 (16) cents by completing Task B.

Both tasks will take exactly 3 minutes to complete, and they do not differ in terms of difficulty. Please choose below which task you would like to take.

- Task A
- Task B

A.1.3 Replication III

[Since the preliminary instructions, task and ex-post questions are identical to EIIB and RI, we only provide instructions around the moment of decision. Moreover, this experiment included an additional replication with randomization of task dominance, identical to Replication II (please see instructions above).]

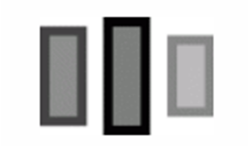
Urgency (Combined, Bonus amounts)

Please click CONTINUE to let the system randomly decide the bonus level and availability of each task.

....

....

System processing, please wait.



....

....

The system has completed the random assignment for task availability and has randomly assigned the bonus level for Task A and Task B.

> You can earn a bonus of 12 (16) cents by completing Task A.
>
> You can earn a bonus of 16 (16) cents by completing Task B.
>
> Task A will expire in 5 (50) minutes.
>
> Task B will expire in 50 (5) minutes.

Both tasks will take exactly 3 minutes to complete, and they do not differ in terms of difficulty. Please choose below which task you would like to take.

(Note: the task expiry times only affect when the tasks are available for you to complete (not how long they will take). Whichever task you choose, it will only take 3-minutes to complete and then you can move on.)

- Task A
- Task B

Urgency (Combined, Expiration)

Please click CONTINUE to let the system randomly decide the bonus level and availability of each task.

System processing, please wait.

The system has completed the random assignment for task availability and has randomly assigned the bonus level for Task A and Task B.

> Task A will expire in 5 (50) minutes.
>
> Task B will expire in 50 (5) minutes.
>
> You can earn a bonus of 12 (16) cents by completing Task A.
>
> You can earn a bonus of 16 (16) cents by completing Task B.

Both tasks will take exactly 3 minutes to complete, and they do not differ in terms of difficulty. Please choose below which task you would like to take.

(Note: the task expiry times only affect when the tasks are available for you to complete (not how long they will take). Whichever task you choose, it will only take 3-minutes to complete and then you can move on.)

- Task A
- Task B

A.1.4 Replication IV

[Since this replication involved rewriting the instructions entirely (up to the point of decision), we provide them in full detail up to this point.]

Preliminary instructions before treatment assignment

Thank you for participating in our study. By completing this survey or questionnaire, you are consenting to be in this research study. Your participation is voluntary and you can stop at any time.

It is very important that you read and follow the instructions carefully.

. . . .
. . . .

Please read the text below and answer the question.

Most modern theories of decision-making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables, can greatly impact the decision process. In order to facilitate our research on decision-making we are interested in knowing certain factors about you, the decision maker. Specifically, we are interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the sports items below, type the word inside the quotation marks: 'typing' in the blank box, and proceed to the next screen by clicking on the 'NEXT' button. Thank you very much.

Which of these activities do you engage most regularly?

- skiing
- soccer
- snowboarding
- running
- hockey
- football
- swimming
- tennis
- cycling
- Other

. . . .
. . . .

Instructions

We are studying how people choose between different options.

- In this study, you will choose one of two options, Option A or Option B.
- Each participant is only allowed to take ONE option (i.e., either Option A or Option B, but not both).
- Each participant is only allowed to participate in the study ONCE.

Please click on 'CONTINUE' to read more about the two tasks.

Control (Fresh-Eye)

Instructions

In this study, you will choose one of two options, Option A or Option B.

Each option requires you to perform the SAME 3-minute task, which is as follows:

The task

You will be shown 100 randomly generated 6-letter strings and you must type as many of them as you can in reverse order in 3 minutes.

As an example: If you see 'rlgows' for one string, you will need to type 'swoglr' for a correct reversal of that string.

Note. A 3-minute countdown timer will be displayed when you start the task. It is IMPOSSIBLE for you to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and you will move on.

The two options DIFFER in their bonus levels.

Bonus levels

In one of the options, you will be given a bonus of 12 cents. In the other option, you will be given a bonus of 16 cents. (Either way, the task is the same and is as described above.)

The bonus will be added to your MTurk payment after you complete the study.

When you must start

You must start the task at any time within 47 minutes of selecting your option.

Before you proceed, please answer the following questions to indicate that you understand the instructions.

1. How many cents will you get for completing these two options?
    - Both options will give me a bonus of 16 cents.
    - One of the options will give me a bonus of 12 cents and the other will give me a bonus of 16 cents.
    - It was not specified.

2. Options A and B require you to perform the SAME task. Could you specify what is the task you need to perform in these two options?

    - In both options, I will be shown 100 randomly generated 6-letter strings. My job is to type as many strings as I can in reverse order in 3 minutes.
    - In both options, I will be shown 100 randomly generated 6-letter strings. My job is to type as many strings as I can in reverse order in 5 minutes.
    - It was not specified.

3. When you start your 3-minute task, what will happen?

    - A 3-minute countdown timer will be displayed when I start the task. It is POSSIBLE for me to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and I will have to wait.
    - A 3-minute countdown timer will be displayed when I start the task. It is IMPOSSIBLE for me to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and I will move on.
    - It was not specified.

Please click CONTINUE.

. . . .
. . . .

	Task duration	**Bonus level**	**When you must start**
Option A	3 minutes	12 (16) cents	At any time within 47 minutes of selecting this option
Option B	3 minutes	16 (12) cents	At any time within 47 minutes of selecting this option

Both options require you to perform the SAME task.

Note. A 3-minute countdown timer will be displayed when you start the task. It is IMPOSSIBLE for you to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and you will move on.

Please choose below which option you would like to take.

- Task A
- Task B

Urgency (Fresh-Eye)

Instructions

In this study, you will choose one of two options, Option A or Option B.

Each option requires you to perform the SAME 3-minute task, which is as follows:

The task

You will be shown 100 randomly generated 6-letter strings and you must type as many of them as you can in reverse order in 3 minutes.

As an example: If you see 'rlgows' for one string, you will need to type 'swoglr' for a correct reversal of that string.

Note. A 3-minute countdown timer will be displayed when you start the task. It is IMPOSSIBLE for you to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and you will move on.

The two options DIFFER in (i) their bonus levels: and (ii) when you may start the task.

(i) Bonus levels

In one of the options, you will be given a bonus of 12 cents. In the other option, you will be given a bonus of 16 cents. (Either way, the task is the same and is as described above.)

The bonus will be added to your MTurk payment after you complete the study.

(ii) When you must start

In one of the options, you must start the task at any time within 2 minutes of selecting that option. In the other option, you must start the task at any time within 47 minutes of selecting that option. (Either way, the task is the same and is as described above.)

Before you proceed to the next screen, please answer the following questions to indicate that you understand the instructions.

1. How many cents will you get for completing these two options?
   - Both options will give me a bonus of 16 cents.
   - One of the options will give me a bonus of 12 cents and the other will give me a bonus of 16 cents.
   - It was not specified.
2. When must you start?
   - In both options, I must start the task at any time within 47 minutes of selecting my option.
   - In one of the options, I must start the task at any time within 2 minutes of selecting that option. In the other option, I must start the task at any time within 47 minutes of selecting that option.

- It was not specified.

3. Options A and B require you to perform the SAME task. Could you specify what is the task you need to perform in these two options?

   - In both options, I will be shown 100 randomly generated 6-letter strings. My job is to type as many strings as I can in reverse order in 3 minutes.
   - In both options, I will be shown 100 randomly generated 6-letter strings. My job is to type as many strings as I can in reverse order in 5 minutes.
   - It was not specified.

4. When you start your 3-minute task, what will happen?

   - A 3-minute countdown timer will be displayed when I start the task. It is POSSIBLE for me to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and I will have to wait.
   - A 3-minute countdown timer will be displayed when I start the task. It is IMPOSSIBLE for me to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and I will move on.
   - It was not specified.

Please click CONTINUE.

....
....

	Task duration	**Bonus level**	**When you must start**
Option A	3 minutes	12 (16) cents	At any time within 2 (47) minutes of selecting this option
Option B	3 minutes	16 (12) cents	At any time within 47 (2) minutes of selecting this option

Both options require you to perform the SAME task.

Note. A 3-minute countdown timer will be displayed when you start the task. It is IMPOSSIBLE for you to spend longer than 3 minutes on it. At the end of the 3 minutes, the task is over, and you will move on.

Please choose below which option you would like to take.

- Task A
- Task B

A.2 Power calculations

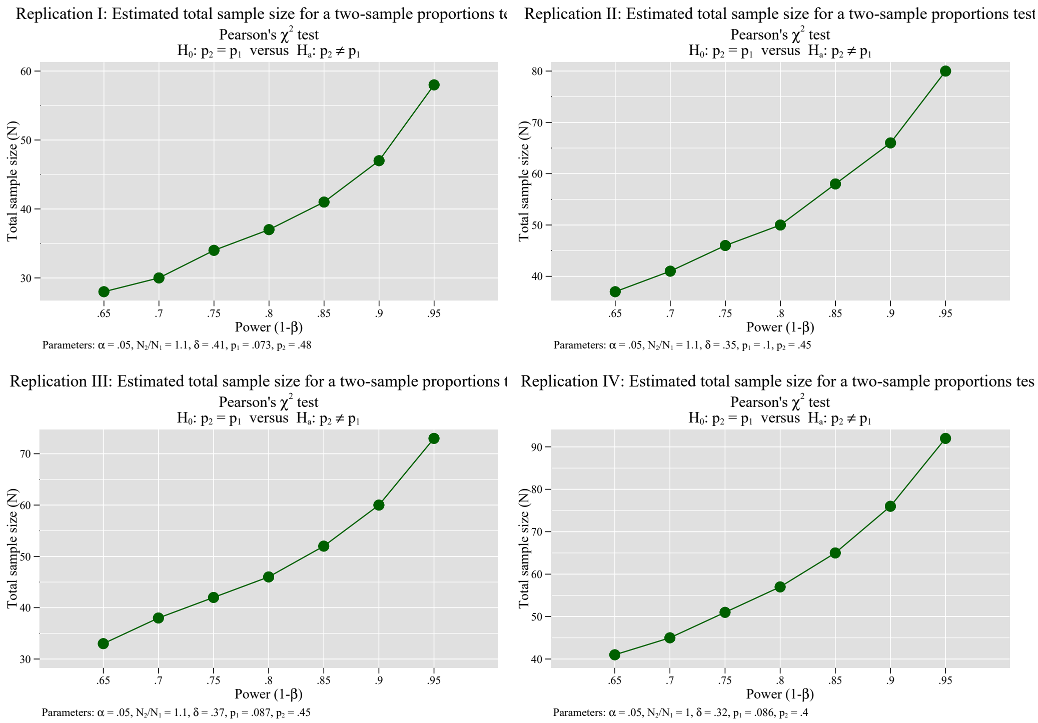


Figure A1: Power calculations

Notes. These power calculations are for Replications I, II, II and IV, and show what power we would obtain under various total sample sizes (N). The test for each is a two-sample χ^2 test.

A.3 Summary statistics

Experiment IIB and Replication I

	C(O)	U(O)	C(PR)	U(PR)
Decision metrics				
Dominated choice	0.07	0.48	0.12	0.44
	(0.26)	(0.50)	(0.32)	(0.50)
Bonus recall (wrong)	0.08	0.18	0.16	0.15
	(0.28)	(0.38)	(0.36)	(0.36)
Time recall (wrong)	0.44	0.12	0.45	0.09
	(0.50)	(0.33)	(0.50)	(0.29)
Attention bonus	7.26	5.70	7.24	6.14
	(1.71)	(2.48)	(1.61)	(2.36)
Attention availibility	4.02	5.56	3.85	5.14
	(2.70)	(2.56)	(2.74)	(2.76)
Attn. diff.	-3.24	-0.14	-3.39	-1.00
	(3.10)	(4.05)	(3.28)	(4.10)
Leave awareness	.	.	0.75	0.61
	(.)	(.)	(0.44)	(0.49)
Busy	.	.	5.40	5.45
	(.)	(.)	(1.80)	(1.77)
(Socio-)Demographics				
Age	36.50	35.94	35.74	37.46
	(11.38)	(10.67)	(12.04)	(12.46)
Female	0.51	0.50	0.56	0.50
	(0.50)	(0.50)	(0.99)	(0.98)
Less than high school diploma	.	.	0.01	0.00
	(.)	(.)	(0.07)	(0.07)
High school diploma or GED	.	.	0.09	0.11
	(.)	(.)	(0.29)	(0.31)
Some college, but no degree	.	.	0.24	0.20
	(.)	(.)	(0.43)	(0.40)
Associates Degree	.	.	0.08	0.12
	(.)	(.)	(0.27)	(0.32)
Bachelor's Degree	.	.	0.43	0.42
	(.)	(.)	(0.50)	(0.49)
Master's Degree	.	.	0.13	0.12
	(.)	(.)	(0.33)	(0.32)
Professional Degree	.	.	0.01	0.02
	(.)	(.)	(0.07)	(0.14)
Doctorate	.	.	0.03	0.01
	(.)	(.)	(0.16)	(0.10)
<$15,000	0.08	0.09	0.10	0.09
	(0.28)	(0.29)	(0.30)	(0.29)
$15,000-$29,999	0.20	0.16	0.14	0.14
	(0.40)	(0.37)	(0.35)	(0.35)
$30,000-$49,999	0.20	0.25	0.19	0.25
	(0.40)	(0.44)	(0.39)	(0.43)
$50,000-$74,999	0.26	0.21	0.24	0.24
	(0.44)	(0.41)	(0.43)	(0.43)
$75,000-$99,999	0.17	0.14	0.20	0.12
	(0.37)	(0.35)	(0.40)	(0.32)
>$100,000	0.09	0.15	0.12	0.14
	(0.29)	(0.36)	(0.33)	(0.35)
Prefer not to say	0.00	0.00	0.02	0.03
	(0.00)	(0.00)	(0.12)	(0.17)
English ability (1 to 8)	.	.	7.78	7.80
	(.)	(.)	(0.79)	(0.58)
Time Preference (std)	.	.	0.11	0.06
	(.)	(.)	(1.07)	(1.05)

Table A1: Summary statistics

Notes. Columns 1 and 2 (Experiment IIB from Zhu et al. (2018)): C(O) = C(Original), U(O) = U(Original). Columns 3 and 4 (Replication I): C(PR) = C(Pure Replication), U(PR) = U(Pure Replication)

Replication II

	C(Rep.)	U(Rep.)	U(LH)	U(S)	U(R)
Decision metrics					
Dominated choice	0.09 (0.28)	0.45 (0.50)	0.40 (0.49)	0.26 (0.44)	0.09 (0.29)
Bonus recall (wrong)	0.15 (0.36)	0.21 (0.41)	0.27 (0.44)	0.14 (0.35)	0.13 (0.33)
Time recall (wrong)	0.47 (0.50)	0.15 (0.36)	0.14 (0.34)	0.16 (0.37)	0.36 (0.48)
Attention bonus	7.33 (1.61)	5.94 (2.52)	6.18 (2.38)	6.66 (2.12)	7.31 (1.43)
Attention availibility	3.89 (2.72)	5.21 (2.93)	5.14 (2.79)	4.87 (2.72)	4.41 (2.75)
Attn. diff.	-3.43 (3.08)	-0.73 (4.27)	-1.05 (4.19)	-1.79 (4.14)	-2.90 (3.32)
Leave awareness	0.71 (0.45)	0.65 (0.48)	0.49 (0.50)	0.61 (0.49)	0.59 (0.49)
Busy	5.47 (1.92)	5.21 (1.95)	5.46 (1.89)	5.54 (1.74)	5.65 (1.83)
(Socio-)Demographics					
Age	36.17 (10.71)	37.40 (11.83)	36.99 (11.47)	37.54 (11.67)	36.80 (11.45)
Female	0.63 (0.85)	0.58 (0.84)	0.52 (0.50)	0.57 (1.06)	0.73 (1.22)
Less than high school diploma	0.01 (0.08)	0.00 (0.00)	0.00 (0.00)	0.01 (0.11)	0.01 (0.08)
High school diploma or GED	0.06 (0.24)	0.16 (0.37)	0.08 (0.28)	0.10 (0.30)	0.07 (0.26)
Some college, but no degree	0.24 (0.43)	0.25 (0.43)	0.27 (0.44)	0.23 (0.42)	0.16 (0.36)
Associates Degree	0.11 (0.32)	0.09 (0.28)	0.11 (0.32)	0.10 (0.31)	0.16 (0.36)
Bachelor's Degree	0.40 (0.49)	0.33 (0.47)	0.39 (0.49)	0.40 (0.49)	0.44 (0.50)
Master's Degree	0.15 (0.35)	0.12 (0.33)	0.13 (0.34)	0.10 (0.30)	0.14 (0.35)
Professional Degree	0.03 (0.16)	0.05 (0.22)	0.01 (0.11)	0.03 (0.17)	0.03 (0.17)
Doctorate	0.01 (0.08)	0.01 (0.08)	0.01 (0.08)	0.02 (0.15)	0.00 (0.00)
<$15,000	0.05 (0.21)	0.11 (0.32)	0.07 (0.25)	0.06 (0.24)	0.07 (0.25)
$15,000-$29,999	0.15 (0.35)	0.16 (0.37)	0.16 (0.37)	0.07 (0.26)	0.11 (0.31)
$30,000-49,999$	0.23 (0.42)	0.16 (0.37)	0.24 (0.43)	0.27 (0.44)	0.25 (0.43)
$50,000-$74,999	0.22 (0.42)	0.28 (0.45)	0.23 (0.42)	0.26 (0.44)	0.25 (0.43)
$75,000-$99,999	0.19 (0.39)	0.14 (0.35)	0.14 (0.34)	0.20 (0.40)	0.16 (0.37)
>$100,000	0.15 (0.36)	0.14 (0.35)	0.15 (0.36)	0.12 (0.33)	0.14 (0.35)
Prefer not to say	0.01 (0.12)	0.01 (0.11)	0.02 (0.13)	0.02 (0.13)	0.04 (0.19)
English ability (1 to 8)	7.87 (0.47)	7.83 (0.45)	7.83 (0.51)	7.88 (0.36)	7.84 (0.47)

Table A2: Summary statistics

Notes. C(Rep.) = C(Replication); U(Rep.) = U(Replication); U(LH) = U(Leave Highlight); U(S) = U(Saliency); U(R) = U(Reversal)

Replication III

	C(Rep.)	U(Rep.)	U(CP)	U(CA)
Decision metrics				
Dominated choice	0.08	0.36	0.21	0.25
	(0.28)	(0.48)	(0.41)	(0.44)
Bonus recall (wrong)	0.12	0.30	0.19	0.26
	(0.32)	(0.46)	(0.39)	(0.44)
Time recall (wrong)	0.42	0.14	0.23	0.30
	(0.49)	(0.35)	(0.42)	(0.46)
Attention bonus	7.22	5.89	6.80	6.82
	(1.50)	(2.38)	(2.03)	(1.84)
Attention availibility	3.88	5.29	4.91	4.77
	(2.64)	(2.69)	(2.75)	(2.56)
Attn. diff.	-3.34	-0.60	-1.89	-2.05
	(3.15)	(3.90)	(3.78)	(3.56)
Leave awareness	0.74	0.64	0.46	0.53
	(0.44)	(0.48)	(0.50)	(0.50)
Busy	5.52	5.48	5.54	5.54
	(1.54)	(1.75)	(1.50)	(1.72)
(Socio-)Demographics				
Age	37.26	37.04	37.86	37.17
	(11.75)	(10.74)	(12.41)	(11.74)
Female	0.50	0.51	0.64	0.58
	(0.50)	(0.50)	(0.85)	(1.05)
Less than high school diploma	0.00	0.01	0.00	0.01
	(0.00)	(0.08)	(0.00)	(0.08)
High school diploma or GED	0.07	0.08	0.05	0.05
	(0.26)	(0.27)	(0.23)	(0.23)
Some college, but no degree	0.24	0.19	0.21	0.16
	(0.43)	(0.39)	(0.41)	(0.37)
Associates Degree	0.10	0.10	0.12	0.13
	(0.30)	(0.30)	(0.33)	(0.34)
Bachelor's Degree	0.42	0.48	0.43	0.41
	(0.50)	(0.50)	(0.50)	(0.49)
Master's Degree	0.13	0.13	0.16	0.21
	(0.34)	(0.34)	(0.37)	(0.41)
Professional Degree	0.03	0.02	0.01	0.01
	(0.17)	(0.13)	(0.12)	(0.08)
Doctorate	0.01	0.01	0.01	0.02
	(0.08)	(0.08)	(0.08)	(0.13)
<$15,000	0.08	0.08	0.03	0.05
	(0.27)	(0.28)	(0.16)	(0.21)
$15,000-$29,999	0.12	0.10	0.17	0.15
	(0.33)	(0.30)	(0.37)	(0.36)
$30,000-$49,999	0.24	0.23	0.20	0.27
	(0.43)	(0.43)	(0.40)	(0.45)
$50,000-$74,999	0.26	0.23	0.26	0.17
	(0.44)	(0.43)	(0.44)	(0.38)
$75,000-$99,999	0.11	0.15	0.17	0.14
	(0.31)	(0.36)	(0.38)	(0.35)
>$100,000	0.17	0.17	0.15	0.19
	(0.38)	(0.38)	(0.36)	(0.39)
Prefer not to say	0.02	0.02	0.02	0.02
	(0.13)	(0.15)	(0.14)	(0.15)
English ability (1 to 8)	7.85	7.81	7.89	7.80
	(0.42)	(0.65)	(0.33)	(0.54)

Table A3: Summary statistics

Notes. C(Rep.) = C(Replication); U(Rep.) = U(Replication); U(CP) = U(Combined-Payoffs); U(CA) = U(Combined-Availability)

Replication IV

	C(FE)	U(FE)
Decision metrics		
Dominated choice	0.11 (0.32)	0.36 (0.48)
Bonus recall (wrong)	0.24 (0.43)	0.21 (0.41)
Time recall (wrong)	0.25 (0.44)	0.24 (0.43)
Attention bonus	7.45 (1.01)	6.81 (1.70)
Attention availibility	5.81 (2.28)	6.07 (2.35)
Attn. diff.	-1.64 (2.57)	-0.74 (3.18)
Leave awareness	0.28 (0.45)	0.18 (0.38)
Busy	5.58 (1.93)	5.31 (1.95)
(Socio-)Demographics		
Age	34.76 (11.10)	36.53 (12.05)
Female	0.52 (0.84)	0.41 (0.84)
Less than high school diploma	0.01 (0.08)	0.01 (0.11)
High school diploma or GED	0.09 (0.29)	0.12 (0.32)
Some college, but no degree	0.15 (0.36)	0.13 (0.34)
Associates Degree	0.13 (0.34)	0.08 (0.27)
Bachelor's Degree	0.47 (0.50)	0.47 (0.50)
Master's Degree	0.14 (0.35)	0.16 (0.37)
Professional Degree	0.00 (0.00)	0.02 (0.14)
Doctorate	0.01 (0.08)	0.01 (0.11)
<$15,000	0.09 (0.28)	0.06 (0.25)
$15,000-$29,999	0.18 (0.38)	0.21 (0.41)
$30,000-$49,999	0.23 (0.42)	0.27 (0.44)
$50,000-$74,999	0.23 (0.42)	0.25 (0.43)
$75,000-$99,999	0.14 (0.35)	0.12 (0.32)
>$100,000	0.13 (0.34)	0.08 (0.28)
Prefer not to say	0.01 (0.08)	0.01 (0.08)
English ability (1 to 8)	7.70 (0.74)	7.58 (0.83)

Table A4: Summary statistics

Notes. C(FE) = C(Fresh-Eye); U(FE) = U(Fresh-Eye)

A.4 Supplementary results

Task ordering effects

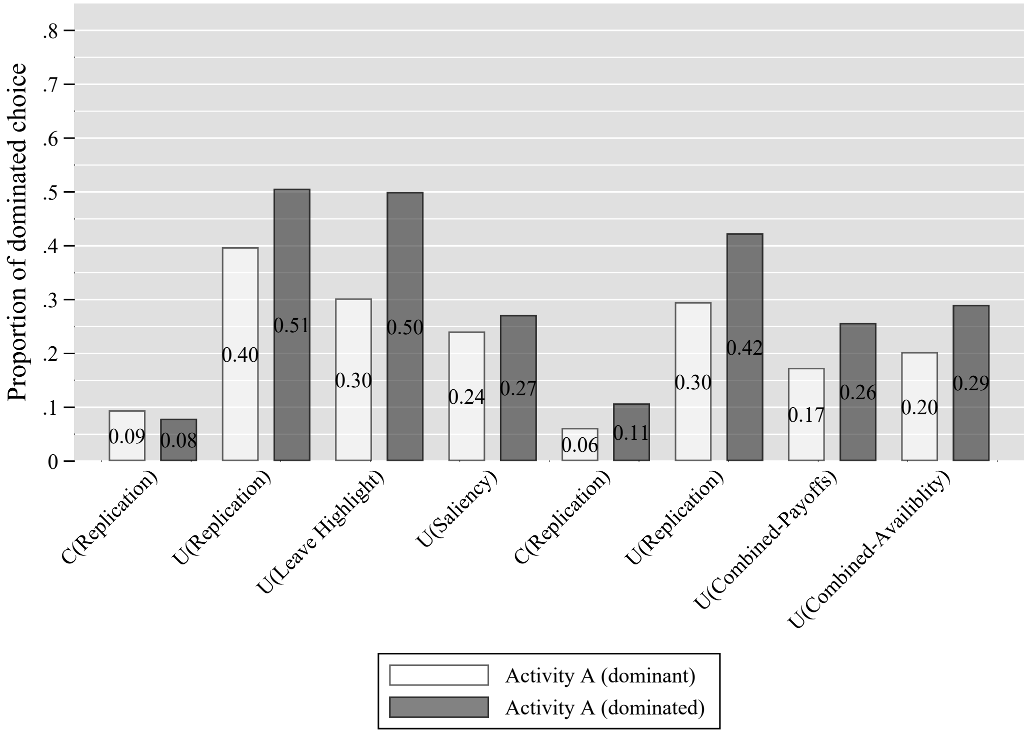


Figure A2: Task ordering effects (Replications II and III)

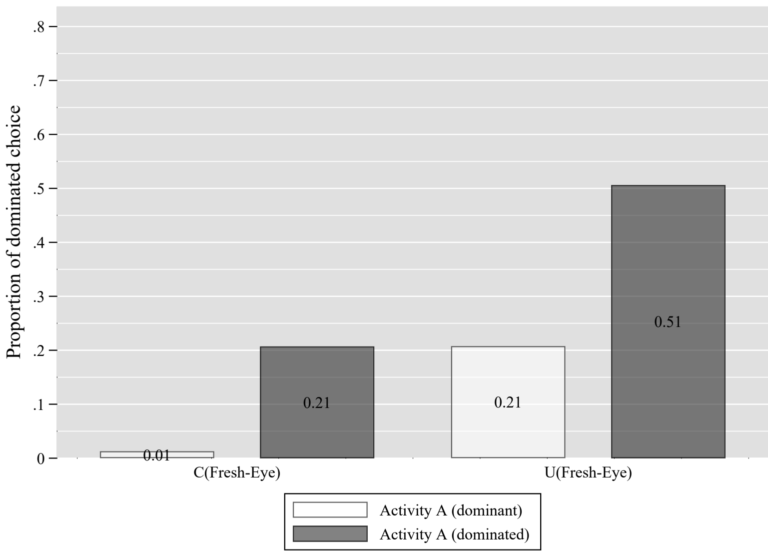


Figure A3: Task ordering effects (Replication IV)

Reasons for choice

Treatments	C(Pooled)	U(Pooled)
Dominating vs. Dominated	0.000***	0.000***

Table A5: Reasons for choice tests (Experiment IIB and Replication I, pooled)

Notes. * p < 0.1, ** p < 0.05, *** p < 0.01. All values reported are p-values χ^2 tests for multiple independent groups.

Treatments	Dominating	Dominated
C(Pooled) vs. U(Pooled)	0.001***	0.000***

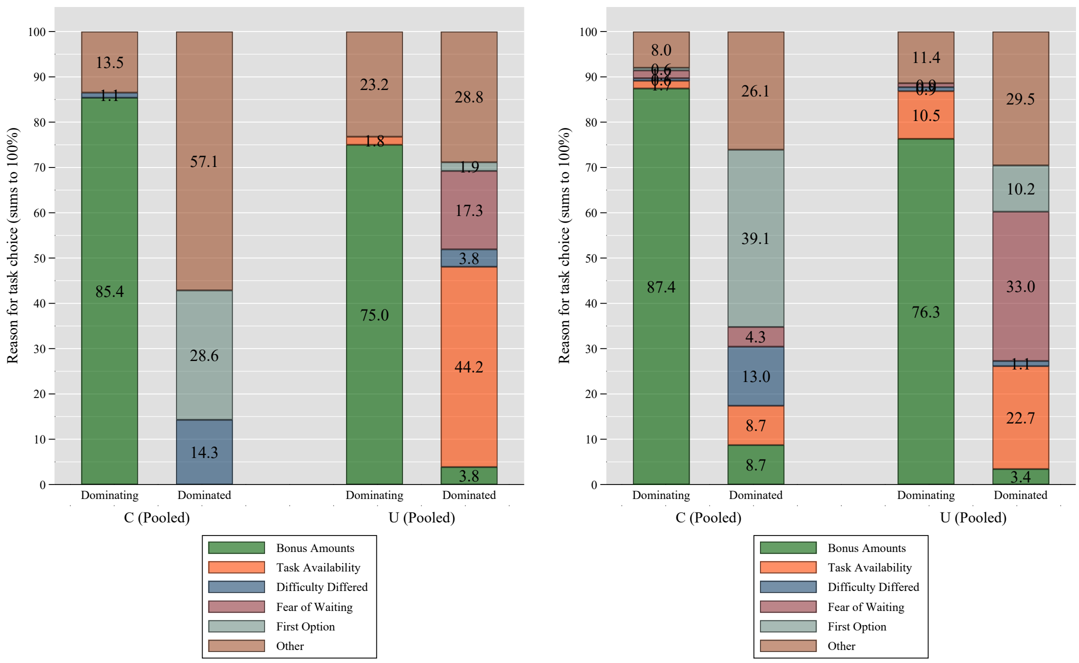
Table A6: Reasons for choice tests (Experiment IIB and Replication I, pooled)

Notes. * p < 0.1, ** p < 0.05, *** p < 0.01. All values reported are p-values χ^2 tests for multiple independent groups.

Treatments	C(Pooled)	U(Pooled)
Dominating (Fear of Waiting) vs. Dominated (Fear of Waiting)	0.277	0.000***
Dominating (First Option) vs. Dominated (First Option)	0.000***	0.000***

Table A7: Reasons for choice tests (Experiment IIB and Replication I, pooled)

Notes. * p < 0.1, ** p < 0.05, *** p < 0.01. All values reported are p-values χ^2 tests for two independent groups.



(a) Experiment IIB **(b)** Replication I

Figure A4: Reasons for choice (Experiment IIB and Replication I, disaggregated)

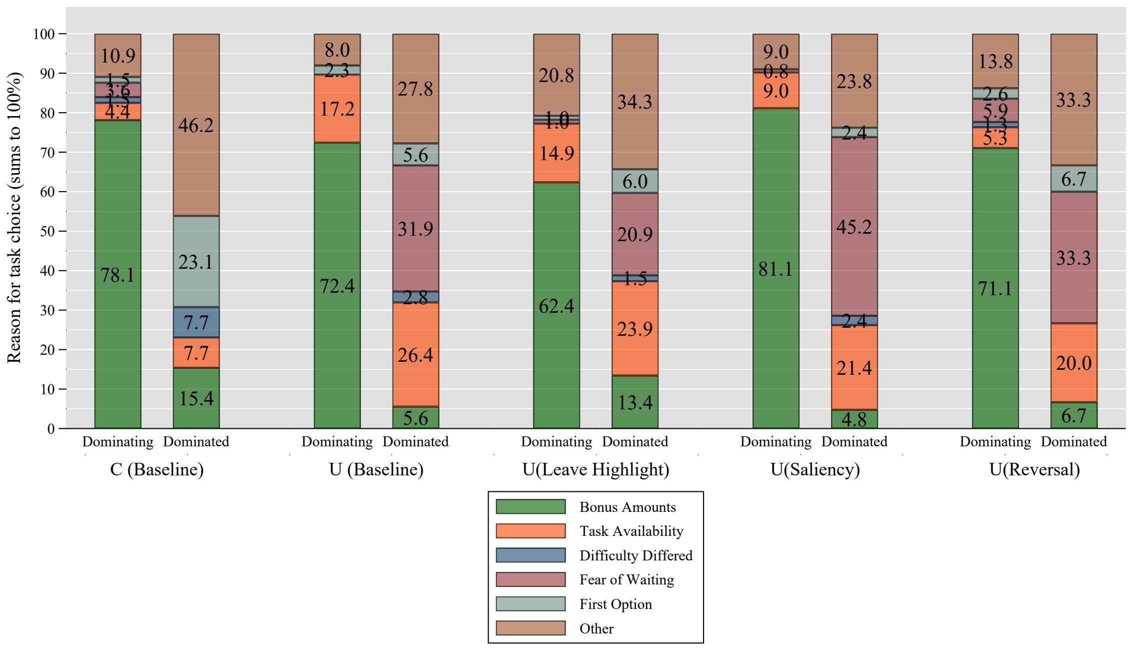


Figure A5: Reasons for choice (Replication II)

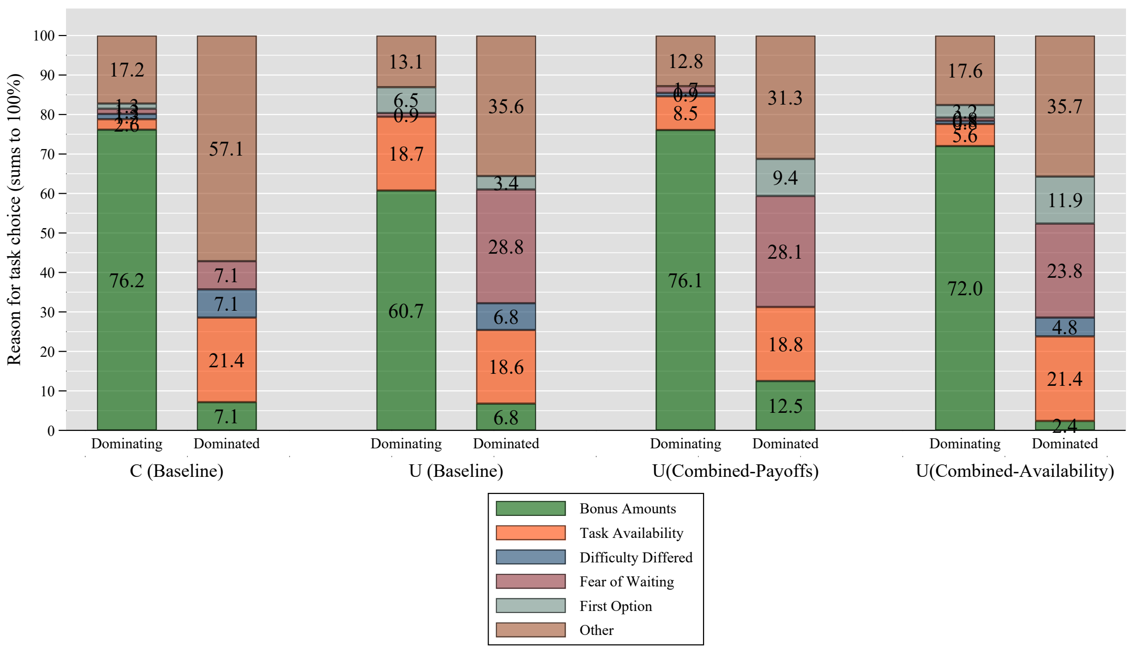


Figure A6: Reasons for choice (Replication III)

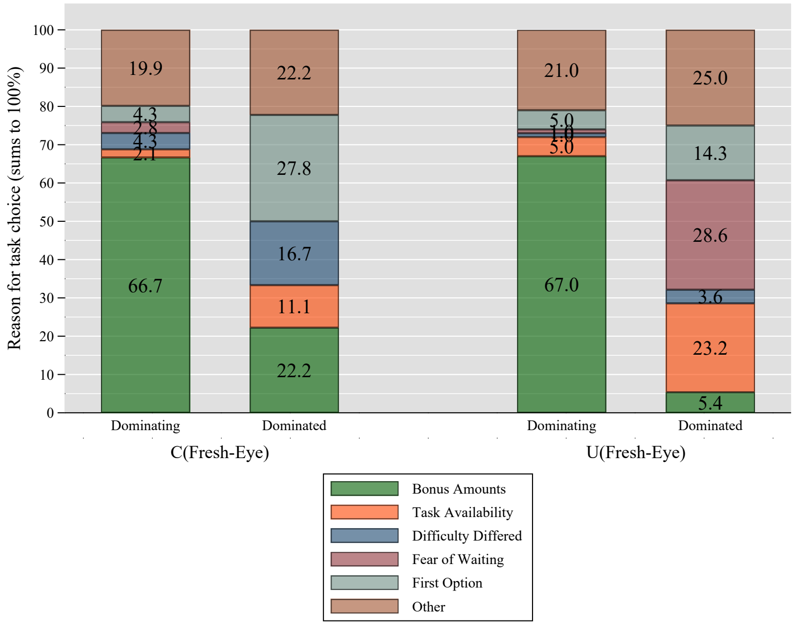


Figure A7: Reasons for choice (Replication IV)

Reversal treatment

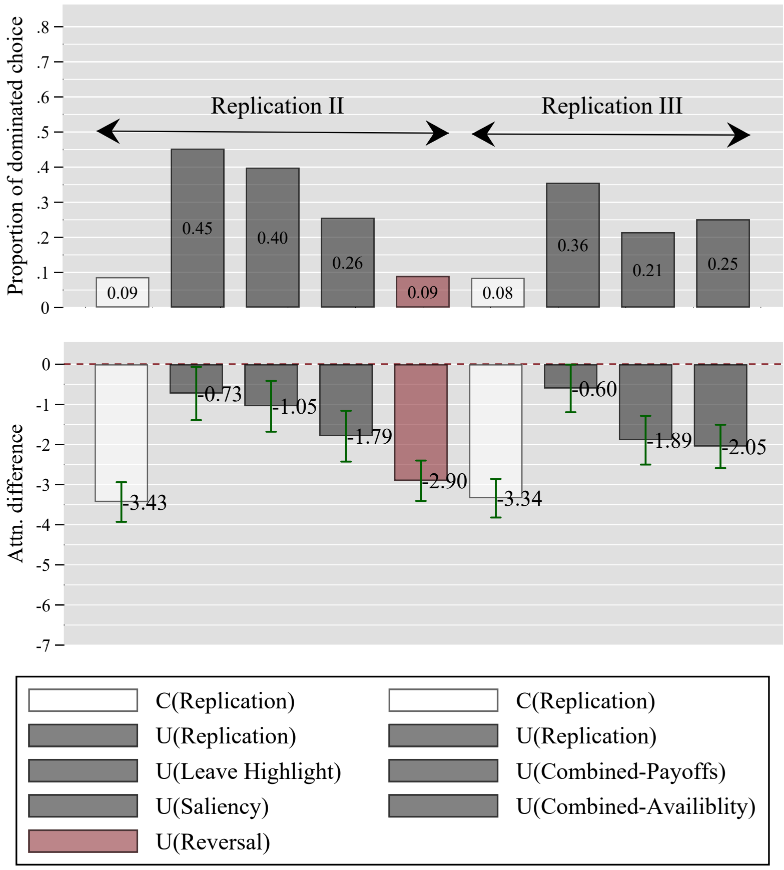


Figure A8: Dominated choice shares across conditions with Reversal treatment (Replications II and III)

Notes. Whiskers represent the 95% confidence. intervals. Attn. difference (−7 to +7) = Attn. expiration (1 to 8) - Attn bonus amounts (1 to 8).

Correlations across variables

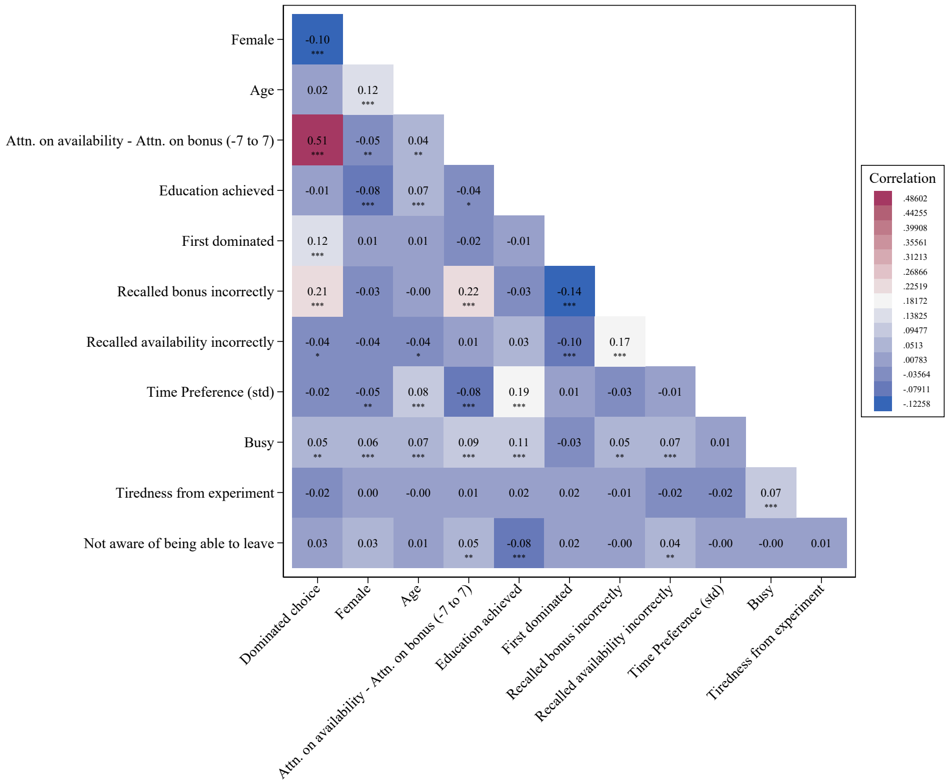


Figure A9: Correlations across variables

* p < 0.1, ** p < 0.05, *** p < 0.01.
Notes. Spearman's rank correlation coefficients (ρ) for RI, RII, RIII and RIV pooled and their associated significance levels. The blue and red shaded boxes represent negative and positive correlations, respectively with intensity of shaded region representing relative strength of correlation (as indicated by legend).

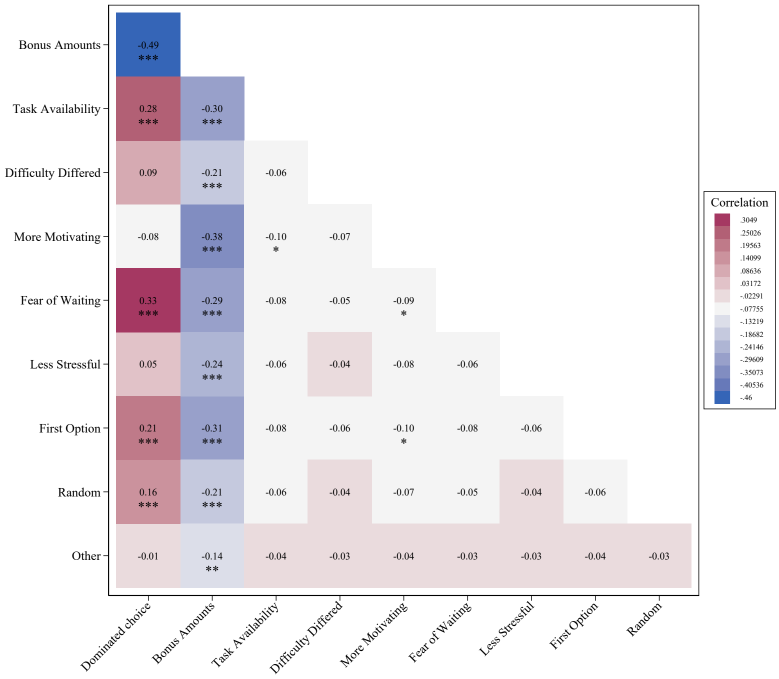


Figure A10: Correlations across reasons provided and choice

* p < 0.1, ** p < 0.05, *** p < 0.01.
Notes. Spearman's rank correlation coefficients (ρ) for EIIB, RI, RII, RIII and RIV pooled and their associated signifiance levels. The blue and red shaded boxes represent negative and positive correlations, respectively with intensity of shaded region representing relative strength of correlation (as indicated by legend).

Econometrics

	Choice					Attention
	(1) EIIB	(2) RI	(3) RII	(4) RIII	(5) RIV	(6) Pooled data
U(Original)	0.44*** (0.05)					
U(Pure Replication)		0.36*** (0.04)				
U(Replication)			0.39*** (0.04)			
U(Leave Highlight)			0.34*** (0.04)			
U(Saliency)			0.21*** (0.04)			
U(Reversal)			0.01 (0.03)			
U(Replication)				0.29*** (0.04)		
U(Combined-Payoffs)				0.15*** (0.04)		
U(Combined-Availiblity)				0.16*** (0.04)		
U(Fresh-Eye)					0.23*** (0.04)	
More attention on payoffs						-0.23*** (0.02)
More attention on availability						0.31*** (0.03)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	204	400	801	647	313	2374
Psuedo R^2	0.25	0.17	0.15	0.10	0.15	0.24
Log-lik	-91.76	-195.82	-389.04	-312.69	-144.67	-1024.06
χ^2	60.60	68.75	126.72	64.38	58.41	587.86

Table A8: Probit models (marginal effects)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Sample composition is the following. Column 1 includes data from Experiment IIB of Zhu et al. (2018): Column 2 includes data from Replication I: Column 3 includes data from Replication II: Column 4 includes data from Replication III: Column 5 includes data from Replication IV: and Column 6 includes all the data from Experiment IIB and Replications I, II, III, and IV pooled together. Dependent variable: Choice (1=Dominated, 0=Dominating). Each individual specification refers to the urgency treatments relative to the control condition in each of the experiments. Additional controls: gender, age, income, & comprehension (measured by ability to recall bonus amounts as well as availability). The last column refers to the independent variable of attention, split into: equal attention (base), more attention on payoffs and more attention on availability.

	Attention				
	(1) EIIB	(2) RI	(3) RII	(4) RIII	(5) RIV
U(Original)	0.34*** (0.05)				
U(Pure Replication)		0.28*** (0.04)			
U(Replication)			0.29*** (0.04)		
U(Leave Highlight)			0.28*** (0.04)		
U(Saliency)			0.20*** (0.04)		
U(Reversal)			0.03 (0.03)		
U(Replication)				0.27*** (0.04)	
U(Combined-Payoffs)				0.17*** (0.04)	
U(Combined-Availiblity)				0.13*** (0.04)	
U(Fresh-Eye)					0.16*** (0.04)
Additional controls	Yes	Yes	Yes	Yes	Yes
# Observations	204	396	808	644	313
Psuedo R^2	0.21	0.13	0.10	0.12	0.13
Log-lik	-90.28	-170.25	-387.00	-298.59	-145.24
χ^2	43.42	43.81	79.06	68.41	46.35

Table A9: Probit models (marginal effects)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Sample composition is the following. Column 1 includes data from Experiment IIB of Zhu et al. (2018): Column 2 includes data from Replication I: Column 3 includes data from Replication II: Column 4 includes data from Replication III: and Column 5 includes data from Replication IV. Dependent variable: Attention (1=Attention to time > Attention to bonus, 0=Otherwise). Each individual specification refers to the urgency treatments relative to the control condition in each of the experiments. Additional controls: gender, age, income, & comprehension (measured by ability to recall bonus amounts as well as availability).

	Choice (restricted)					Attention (restricted)
	(1) EIIB	(2) RI	(3) RII	(4) RIII	(5) RIV	(6) Pooled data
U(Original)	0.40*** (0.05)					
U(Pure Replication)		0.28*** (0.05)				
U(Replication)			0.26*** (0.05)			
U(Leave Highlight)			0.28*** (0.04)			
U(Saliency)			0.12*** (0.03)			
U(Reversal)			0.02 (0.02)			
U(Replication)				0.18*** (0.04)		
U(Combined-Payoffs)				0.07** (0.03)		
U(Combined-Availiblity)				0.08** (0.04)		
U(Fresh-Eye)					0.17*** (0.04)	
More attention on payoffs						-0.15*** (0.02)
More attention on availability						0.31*** (0.04)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	153	303	617	493	238	1805
Psuedo R^2	0.39	0.25	0.17	0.08	0.13	0.25
Log-lik	-46.84	-91.17	-219.04	-171.84	-82.09	-562.66
χ^2	43.00	62.32	85.42	36.02	36.38	335.96

Table A10: Probit models (marginal effects)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Sample composition is the following. Column 1 includes data from Experiment IIB of Zhu et al. (2018): Column 2 includes data from Replication I: Column 3 includes data from Replication II: Column 4 includes data from Replication III: Column 5 includes data from Replication IV: and Column 6 includes all the data from Experiment IIB and Replications I, II, III, and IV pooled together. Each sample has been restricted to those who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful in the 'reasons for choice' variable. We kept (iii) and (iv) because these are plausibly related to a sense of urgency. Dependent variable: Choice (1=Dominated, 0=Dominating). Each individual specification refers to the urgency treatments relative to the control condition in each of the experiments. Additional controls: gender, age, income, & comprehension (measured by ability to recall bonus amounts as well as availability). The last column refers to the independent variable of attention, split into: equal attention (base), more attention on payoffs and more attention on availability.

	Attention (restricted)				
	(1) EIIB	(2) RI	(3) RII	(4) RIII	(5) RIV
U(Original)	0.33***				
	(0.05)				
U(Pure Replication)		0.25***			
		(0.05)			
U(Replication)			0.24***		
			(0.05)		
U(Leave Highlight)			0.22***		
			(0.04)		
U(Saliency)			0.11***		
			(0.04)		
U(Reversal)			0.01		
			(0.03)		
U(Replication)				0.24***	
				(0.04)	
U(Combined-Payoffs)				0.12***	
				(0.04)	
U(Combined-Availiblity)				0.11***	
				(0.04)	
U(Fresh-Eye)					0.10**
					(0.04)
Additional controls	Yes	Yes	Yes	Yes	Yes
# Observations	153	303	624	502	238
Psuedo R^2	0.34	0.16	0.11	0.14	0.13
Log-lik	-45.90	-102.47	-242.71	-184.80	-86.68
χ^2	33.34	36.07	59.04	48.94	29.93

Table A11: Probit models (marginal effects)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Sample composition is the following. Column 1 includes data from Experiment IIB of Zhu et al. (2018): Column 2 includes data from Replication I: Column 3 includes data from Replication II: Column 4 includes data from Replication III: and Column 5 includes data from Replication IV. Each sample has been restricted to those who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful in the 'reasons for choice' variable. We kept (iii) and (iv) because these are plausibly related to a sense of urgency. Dependent variable: Attention (1=Attention to time > Attention to bonus, 0=Otherwise). Each individual specification refers to the urgency treatments relative to the control condition in each of the experiments. Additional controls: gender, age, income, & comprehension (measured by ability to recall bonus amounts as well as availability).

	Choice		Attention	
	(1) Pooled	(2) Pooled (Restricted)	(3) Pooled	(4) Pooled (Restricted)
C(Original)	-0.44***	-0.40***	-0.35***	-0.34***
	(0.05)	(0.06)	(0.05)	(0.06)
C(Pure Replication)	-0.40***	-0.38***	-0.36***	-0.31***
	(0.05)	(0.06)	(0.05)	(0.06)
U(Pure Replication)	-0.04	-0.12*	-0.08	-0.08
	(0.06)	(0.07)	(0.06)	(0.07)
C(Replication)	-0.42***	-0.38***	-0.36***	-0.30***
	(0.05)	(0.06)	(0.05)	(0.06)
U(Replication)	-0.04	-0.12*	-0.06	-0.06
	(0.06)	(0.07)	(0.06)	(0.07)
U(Leave Highlight)	-0.09	-0.10	-0.08	-0.08
	(0.06)	(0.07)	(0.06)	(0.07)
U(Saliency)	-0.22***	-0.27***	-0.14**	-0.18***
	(0.06)	(0.06)	(0.06)	(0.06)
U(Reversal)	-0.41***	-0.35***	-0.32***	-0.28***
	(0.05)	(0.06)	(0.05)	(0.06)
C(Replication)	-0.42***	-0.36***	-0.34***	-0.30***
	(0.05)	(0.06)	(0.05)	(0.06)
U(Replication)	-0.14**	-0.18***	-0.08	-0.07
	(0.06)	(0.07)	(0.06)	(0.07)
U(Combined-Payoffs)	-0.27***	-0.28***	-0.18***	-0.18***
	(0.06)	(0.06)	(0.06)	(0.06)
U(Combined-Availiblity)	-0.25***	-0.28***	-0.22***	-0.20***
	(0.06)	(0.06)	(0.06)	(0.06)
C(Fresh-Eye)	-0.40***	-0.36***	-0.29***	-0.25***
	(0.05)	(0.06)	(0.05)	(0.06)
U(Fresh-Eye)	-0.15**	-0.18***	-0.12**	-0.14**
	(0.06)	(0.07)	(0.06)	(0.07)
Additional controls	Yes	Yes	Yes	Yes
# Observations	2374	1805	2374	1826
Psuedo R^2	0.14	0.15	0.11	0.12
Log-lik	-1160.42	-634.79	-1115.44	-685.46
χ^2	355.57	222.10	232.33	164.07

Table A12: Probit models (marginal effects)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Sample composition is the following. Columns 1 and 2 includes all the data from Experiment IIB and Replications I, II, III, and IV pooled together: and Columns 2 and 3 includes all the data from Experiment IIB and Replications I, II, III, and IV pooled together as well as restricted. The restricted sample includes those who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful in the 'reasons for choice' variable. We kept (iii) and (iv) because these are plausibly related to a sense of urgency. Dependent variables: columns 1 and 2 - Choice (1=Dominated, 0=Dominating), and columns 3 and 4 - Attention (1=Attention to time > Attention to bonus, 0=Otherwise). Each individual specification refers to the treatments (pooled together) relative to the original urgency condition in each of the four experiments (EIIB, RI, RII, RIII and RIV). This allows us to see to what extent our new treatments attenuate the probability of dominated choice. Additional controls: gender, age, income, and comprehension (measured by ability to recall bonus amounts as well as availability). Columns 1 and 3 refer to the full sample pooled. Column 2 and 4 refer to the restricted sample who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful in the 'reasons for choice' variable.

	Choice			Choice (restricted)		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
U (Pooled)	0.38***	0.23***	0.31***	0.31***	0.20***	0.36***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
C (Baseline)	-0.01	-0.01	-0.03	0.03	0.02	0.05
	(0.02)	(0.02)	(0.05)	(0.02)	(0.02)	(0.04)
U (Baseline)	0.32***	0.18***	0.25***	0.23***	0.14***	0.26***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
U(Leave Highlight)	0.32***	0.19***	0.27***	0.29***	0.21***	0.34***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
U(Saliency)	0.19***	0.09***	0.18***	0.13***	0.07**	0.20***
	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.05)
U(Reversal)	0.01	-0.02	0.01	0.04*	0.03	0.11**
	(0.03)	(0.03)	(0.05)	(0.02)	(0.02)	(0.05)
U(Combined-Payoffs)	0.15***	0.06*	0.13**	0.12***	0.06**	0.17***
	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.06)
U(Combined-Availability)	0.17***	0.09***	0.19***	0.12***	0.07**	0.21***
	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.06)
C(Fresh-Eye)	0.03	-0.07**	-0.05	0.04*	-0.04	0.03
	(0.03)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)
U(Fresh-Eye)	0.27***	0.13***	0.20***	0.21***	0.10***	0.24***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
Attention		0.05***	0.03***		0.04***	0.01**
		(0.00)	(0.01)		(0.00)	(0.00)
U (Pooled) × Attention			0.04***			0.06***
			(0.01)			(0.01)
C (Baseline) × Attention			-0.01			0.01
			(0.01)			(0.01)
U (Baseline) × Attention			0.03***			0.03***
			(0.01)			(0.01)
U(Leave Highlight) × Attention			0.03***			0.04***
			(0.01)			(0.01)
U(Saliency) × Attention			0.03***			0.03***
			(0.01)			(0.01)
U(Reversal) × Attention			0.01			0.02**
			(0.01)			(0.01)
U(Combined-Payoffs) × Attention			0.02*			0.03***
			(0.01)			(0.01)
U(Combined-Availability) × Attention			0.03***			0.04***
			(0.01)			(0.01)
C(Fresh-Eye) × Attention			-0.01			0.01
			(0.01)			(0.01)
U(Fresh-Eye) × Attention			0.03***			0.04***
			(0.01)			(0.01)
Constant	0.07*	0.31***	0.25***	0.02	0.21***	0.09**
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
# observations	2374	2374	2374	1836	1836	1836
R^2	0.1	0.3	0.3	0.1	0.3	0.3

Table A13: OLS: all sessions (pooled)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Sample composition is the following. Columns 1 and 2 includes all the data from Experiment IIB and Replications I, II, III, and IV pooled together: and Columns 2 and 3 includes all the data from Experiment IIB and Replications I, II, III, and IV pooled together as well as restricted. The restricted sample includes those who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful in the 'reasons for choice' variable. We kept (iii) and (iv) because these are plausibly related to a sense of urgency. Dependent variable: Choice (1=Dominated, 0=Dominating). Each individual specification refers to the treatments (pooled together) relative to the original control condition in each of the four experiments (EIIB, RI, RII, RIII and RIV). We interact the treatments with relative attentional focus to investigate the interactions between the two. Additional controls: gender, age, income, and comprehension (measured by ability to recall bonus amounts as well as availability). Columns 1 and 3 refer to the full sample pooled. Column 2 and 4 refer to the restricted sample who stated: (i) bonus amounts, (ii) task availability, (iii) one option was more motivating, and (iv) one option was less stressful in the 'reasons for choice' variable.

	Start time (seconds)		
	(1) OLS	(2) OLS	(3) OLS
U(Pure Replication)	2.12 (2.42)	2.25 (2.55)	3.04 (3.94)
C(Replication)	0.08 (0.90)	0.07 (0.90)	0.09 (1.02)
U(Replication)	-0.02 (1.09)	0.12 (1.08)	-0.45 (0.93)
U(Leave Highlight)	-0.01 (1.11)	0.10 (1.10)	-0.58 (1.06)
U(Saliency)	1.49 (1.91)	1.56 (1.96)	2.34 (2.55)
U(Reversal)	-0.88 (0.79)	-0.88 (0.79)	-1.00 (0.89)
C(Replication)	-0.50 (0.80)	-0.51 (0.81)	-0.63 (0.90)
U(Replication)	0.39 (1.12)	0.49 (1.10)	-0.60 (0.94)
U(Combined-Payoffs)	2.56 (2.22)	2.61 (2.24)	2.00 (2.61)
U(Combined-Availiblity)	1.32 (1.23)	1.38 (1.23)	0.59 (1.24)
C(Fresh-Eye)	2.91 (1.94)	2.91 (1.94)	3.26 (2.19)
U(Fresh-Eye)	2.51 (2.14)	2.60 (2.20)	2.91 (3.08)
Dominated choice		-0.38 (0.84)	
Dominated			-1.25 (1.24)
U(Pure Replication) × Dominated			-1.20 (4.16)
C(Replication) × Dominated			-0.44 (1.58)
U(Replication) × Dominated			1.91 (2.05)
U(Leave Highlight) × Dominated			2.32 (2.29)
U(Saliency) × Dominated			-2.59 (2.86)
U(Reversal) × Dominated			0.96 (1.51)
C(Replication) × Dominated			1.04 (1.35)
U(Replication) × Dominated			3.64 (2.50)
U(Combined-Payoffs) × Dominated			3.21 (4.62)
U(Combined-Availiblity) × Dominated			3.55 (3.35)
C(Fresh-Eye) × Dominated			-3.21 (2.54)
U(Fresh-Eye) × Dominated			-0.37 (3.81)
Constant	2.38 (2.33)	2.36 (2.34)	2.73 (2.54)
Additional controls	Yes	Yes	Yes
N observations	2170	2170	2170
R-squared	0.0	0.0	0.0

Table A14: OLS: all sessions (pooled)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: start time (seconds). Additional controls: gender, age, income, and comprehension (measured by ability to recall bonus amounts as well as availability).

	Correlates of choice		
	(1) ME	(2) ME	(3) ME
Female	-0.08***	-0.07***	-0.07***
	(0.02)	(0.02)	(0.02)
Age	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
> $50,000	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)
Tertiary education	0.02	0.03*	0.03
	(0.02)	(0.02)	(0.02)
Recalled bonus incorrectly		0.18***	0.18***
		(0.02)	(0.02)
Recalled availability incorrectly		0.00	-0.00
		(0.02)	(0.02)
Busy			0.01**
			(0.01)
English ability (1 to 8)			-0.04***
			(0.01)
Time Preference (std)			-0.02**
			(0.01)
Additional controls	No	No	No
Observations	2123	2123	2123
Psuedo R^2	0.10	0.13	0.14
Log-lik	-1076.58	-1036.96	-1026.63
χ^2	227.04	302.19	318.57

Table A15: OLS: all sessions (pooled)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: Choice (1=Dominated, 0=Dominating). Each individual specification refers to the treatments (pooled together) relative to the original control condition in each of the four experiments (EIIB, RI, RII, RIII and RIV). Controls: gender (female), age, income (>$50,000), education (tertiary), ability to recall bonus amounts as well as availability, perceived busyness, perceived english ability and a non-incentivised measure of time preferences.

APPENDIX B

Chapter II

B.1 Theory

B.1.1 A formal expression

In the following, we present a formal expression of the different solutions to the bracketing in time allocation problem.

Proposition: The following two statements are equivalent;

(i) Statement 1: $\bar{b} = \frac{T}{n}$ uniquely solves the Broad Bracket Problem (BBP), and $\bar{b} = 0$ uniquely solves the Narrow Bracket Problem (NBP).

(ii) Statement 2: $\frac{T}{n}(2-\mu) > R > \frac{\mu T}{n}$, where $R \equiv \left(\frac{\alpha}{\beta}\right)\left(\frac{p}{k}\right)$. (Condition C*)

Proof: Let $M_b(\bar{b})$ be the maximand of the BBP expressed as a function of $\bar{b}$; and $M_n(\bar{b})$ be the maximand of the NBP expressed likewise. Note that $M_b(\bar{b}) = M_n(\bar{b}) + (1-\mu)\frac{\alpha kT}{pn}$. Thus, the two functions of $\bar{b}$ have the same derivatives. Setting the first derivative to zero and noting the sign of the second derivative yields:

$$\frac{\partial M_b}{\partial \bar{b}} = \frac{\partial M_n}{\partial \bar{b}} = -\alpha(\frac{k}{p}) + 2\beta(\frac{k}{p})^2\bar{b} = 0$$

$$\frac{\partial^2 M_b}{\partial \bar{b}^2} = \frac{\partial^2 M_n}{\partial \bar{b}^2} = 2\beta(\frac{k}{p})^2 > 0$$

Hence, $M_b(\bar{b})$ and $M_n(\bar{b})$ are both quadratics with a global minimum at $\bar{b} = \frac{R}{2}$, where $R \equiv \left(\frac{\alpha}{\beta}\right)\left(\frac{p}{k}\right)$. Consequently, neither BBP nor NBP has an interior solution.

To characterise their corner solutions, first note that Statement 1 requires $\frac{T}{n} > \frac{R}{2} > 0$ (to allow some of the upward-sloping segment of the broad bracketing maximand to be feasible and, likewise, some of the downward-sloping segment of the narrow bracketing maximand). The assumptions of the model imply $R > 0$ and the first inequality of Statement 2 implies $\frac{T}{n} > \frac{R}{2}$. Hence, if either Statement holds, then $\frac{T}{n} > \frac{R}{2} > 0$. Then, by symmetry of quadratic forms (and noting the feasible ranges of $\bar{b}$ in BBP and NBP, respectively):

$\bar{b} = \frac{T}{n}$ uniquely solves BBP *iff*:

$$\left[\frac{T}{n} - \frac{R}{2}\right] > \left[\frac{R}{2} - \frac{(1-\mu)T}{n}\right] \tag{B.1}$$

Rearranging B.1, yields:

$$\frac{T}{n}\left[2-\mu\right] > R$$

$\bar{b} = 0$ uniquely solves NBP *iff*:

$$\left[\frac{R}{2} - 0\right] > \left[\frac{\mu T}{n} - \frac{R}{2}\right] \tag{B.2}$$

Rearranging B.2, yields:

$$R > \frac{\mu T}{n}$$

The re-arranged implications of B.1 and B.2 jointly imply that, if Statement 1 holds, then so does Statement 2. Conversely, if Statement 2 holds, then B.1 and B.2 also hold and, therefore, so does Statement 1.

■

It may help the reader to note that, though (as shown above) the left-hand-sides (LHS) of B.1 and B.2 are strictly positive whenever either Statement holds, the same is not true of the corresponding right-hand-sides (RHS). Instead, B.1 (resp. B.2) can hold *either* because its RHS is negative or because its RHS is positive but small enough. The relevant 'iff' clause of the proof gives a necessary and sufficient condition because it captures both possibilities.
Figure 2.1 of Chapter 2 illustrates a case where both RHSs are positive but small enough. This is an interesting case because the Proposition holds in it even though, for each of BBP and NBP, both ends of the feasible range of $\bar{b}$ are local maxima. The figure illustrates how, in each case, there is nevertheless only one global maximum when condition C* holds. It is the end of the feasible range that is further than the other from the global minimum, in line with B.1 for BBP and with B.2 for NBP.

B.1.2 Consequentiality of losses

We would like losses from narrow bracketing to be consequential. Exactly how consequential is an open question that we explore below.
We can define $L = BBM$ (evaluated at $\bar{b} = \frac{T}{n}$) − BBM (evaluated at $\bar{b} = \frac{(1-\mu)T}{n}$). Note: The loss is defined by the broad bracketing maximand, evaluated at the levels of $\bar{b}$ that ultimately results from the different types of bracketing behaviour. Under C*, a NB chooses $\bar{b} = 0$ in NBP, but ends up with $\bar{b} = \frac{(1-\mu)T}{n}$. If C* fails, the above formula does not apply and losses from NB = 0. Apart from C*, there is no other way the two forms of bracketing can disagree as $\mu > 2 - \mu$ is impossible.

$$L = \left[\frac{\alpha kT}{pn} - \frac{\alpha k}{p}(\frac{T}{n}) + \beta(\frac{k}{p})(\frac{T}{n})^2\right] - \left[\frac{\alpha kT}{pn} + \frac{\alpha k}{p}(\frac{(1-\mu)T}{n}) - \beta(\frac{k}{p})(\frac{(1-\mu)T}{n}^2)\right]$$
$$L = \mu(\frac{kT}{S})\left[(2-\mu)(\frac{kT}{s})\beta - \alpha\right]$$

Where $S = np$ (as defined in Chapter 3). Moreover, with $\mu(\frac{kT}{S}) > 0$ and $\left[(2-\mu)(\frac{kT}{s})\beta - \alpha\right] > 0$ (guaranteed by the first inequality in C*). L is increasing in $\frac{kT}{S}$. That is, faster agents lose more from narrow bracketing. Further, L is increasing in β and decreasing in α.

$$\frac{\partial L}{\partial \mu} = \mu\frac{kT}{S}\left[-\frac{kT\beta}{S}\right] + \left[(2-\mu)(\frac{kT}{S}\beta - \alpha)\right]\frac{kT}{S}$$
$$\frac{\partial L}{\partial \mu} = \mu\frac{kT}{S}\left[\beta(\frac{kT}{S})2(1-\mu) - \alpha\right]$$

$$\frac{\partial L^2}{\partial \mu^2} = 2\beta(\frac{kT}{S})^2(-1) < 0$$

$\therefore$ L is maximised w.r.t. μ when $2(1-\mu)(\frac{kT}{S}) = \frac{\alpha}{\beta}$

$$1 - \mu = \frac{\alpha}{\beta}(\frac{S}{kT})(\frac{1}{2})$$
$$\mu = [1 - \frac{\alpha}{\beta}(\frac{S}{kT})(\frac{1}{2})]$$

C* guarantees $(2-\mu)(\frac{kT}{S}) > \frac{\alpha}{\beta}$ and so $2 > 2-\mu > \frac{\alpha}{\beta}(\frac{S}{kT}) > 0$. Hence, the L-maximising μ satisfies $0 < \mu < 1$. The intuition for this is simple. Very small μ, implies that hardly any time is affected by narrow bracketing. Hence, L is very small. Very large μ, means that a lot of time is affected by narrow bracketing. But if C* holds, 'as much A as possible' is optimal over a time interval only slightly less than $\frac{T}{n}$. Though 'B only' is optimal over the whole-time interval, it cannot 'win by much'.

Finally, we can define $L\%$ to be the following:

$$L\% = \frac{L}{\text{Maximal attainable earnings}} = \frac{\mu(\frac{kT}{S})\left[(2-\mu)(\frac{kT}{S})\beta - \alpha\right]}{\frac{\alpha kT}{pn} - \frac{\alpha k}{p}(\frac{T}{n}) + \beta(\frac{k}{p})^2(\frac{T}{n})^2}$$
$$L\% = \mu\left[(2-\mu) - (\frac{\alpha}{\beta})(\frac{S}{kT})\right] > 0$$

Note: Maximising L (or $L\%$) is not necessary, but we do want these terms to be 'consequential' enough. The expression for L assumes optimal allocation of time across components within each high-level activity and 'correct' solutions of broad/narrow bracketing problems. Real subjects can lose money in ways not captured by the formulae (and this is indeed what we see). Suppose we decide we would like to keep $L\%$ above some threshold of consequentialness. That is, we want $L\% \geq z$ where $0 < z < 1$ (z is the threshold). Then we need:

$$L\% = \mu\left[(2-\mu) - (\frac{\alpha}{\beta})(\frac{S}{kT})\right] \geq z$$
$$L\% = \mu\left[(2-\mu) - \omega\right] \geq z$$

Where $\omega \equiv (\frac{\alpha}{\beta})(\frac{S}{kT})$. Therefore, with the knowledge that C* is $(\frac{kT}{S})[2-\mu] > \frac{\alpha}{\beta} > \mu(\frac{kt}{S})$, rearranging to $[2-\mu] > \frac{\alpha}{\beta}(\frac{S}{kT}) > \mu$. Now note that this implies $2 > [2-\mu] > \omega > \mu > 0$. Moreover, $2-\mu > \omega \rightarrow 2-\omega > \mu$. Hence, C* requires $min\{2-\omega, \omega\} > \mu$. As a result, $\frac{\partial L\%^2}{\partial \mu^2} = -2 < 0$. $L\%$ is maximised by the same μ that maximises L. This means that $\mu = \frac{2-\omega}{2}$. Its maximised value is $\frac{2-\omega}{2}[\frac{2-\omega}{1} - \frac{2-\omega}{2}]$, yielding $[\frac{2-\omega}{4}]^2$. Suppose we now select candidate values of $\frac{\alpha}{\beta}, S, T$ and that we hypothesise a value for k. This gives us a 'test' value for ω and as a result, $min\{2-\omega, \omega\}$. Thus, we can perform two types of investigation.

(i) Consistency with C*: We have an upper bound on 'useable' values of μ, given our test value of ω.

- $min\{2-\omega, \omega\} \leq 1$ for all $\omega > 0$
- $min\{2-\omega, \omega\} < 0$ for all $\omega > 2$
- $min\{2-\omega, \omega\}$ is larger the closer ω is to 1

(ii) Consistency with the desideration that $L\%$ be 'consequential': Given the test value of ω, we can ask: (a) what is the value of μ that would maximise $L\%$?; (b) what is the resulting maximal possible $L\%$? The answers are given by the formulae above. Given the answer to (b), we can ask: (c) is this 'consequential' enough? and (d) does the answer to (a) meet the upper bound on μ? If the answer to either (c) or (d) is no, we need to reconsider the parameters that gave the test value of ω. If the answer to both (c) and (d) is yes, then we may want to consider values of μ below the one that answer (a). Why? Because generally low μ helps keep the range of speeds (k) at which C* holds wide, helping more subjects to satisfy it. How far could we go? It depends how far below the answer to (b) $L\%$ could fall and still be consequential. Suppose we think any $L\%$ close to some level of z would be consequential enough. Note: We will only reach this point in the decision tree if: $\frac{(2-\omega)^2}{4} > z$. Thus we can impose $\frac{(2-\omega)^2}{4} > z > 0$, which then requires $L\% = \mu[(2-\mu)-\omega] \geq z$. What is the range of μ satisfying this? Define $h(y) \equiv L\% - z = \mu[(2-\mu)-\omega] - z = -\mu^2 + (2-\omega)\mu - z$. $h(y)$ is a quadratic, with a maximum at $\mu = \frac{2-\omega}{2}$, where is takes value $\frac{(2-w)^2}{4} - z > 0$. Thus, for μ 'close-enough' to $\frac{(2-w)}{2}$, $h(y) \geq 0$. To do this, solve $h(y) = -\mu^2 + (2-\omega)\mu - z = 0$. The solutions are:

$$\mu = \frac{-(2-\omega) \pm \sqrt{(2-\omega)^2 - 4z}}{-2}$$
$$\mu = \frac{(2-\omega) \pm \sqrt{(2-\omega)^2 - 4z}}{2}$$

Note: the square root term is positive and well defined. Hence, the range of values of at which $L\%$ is consequential, as defined by z, is:

$$\frac{(2-\omega) - \sqrt{(2-\omega)^2 - 4z}}{2} < \mu < \frac{(2-\omega) + \sqrt{(2-\omega)^2 - 4z}}{2}$$

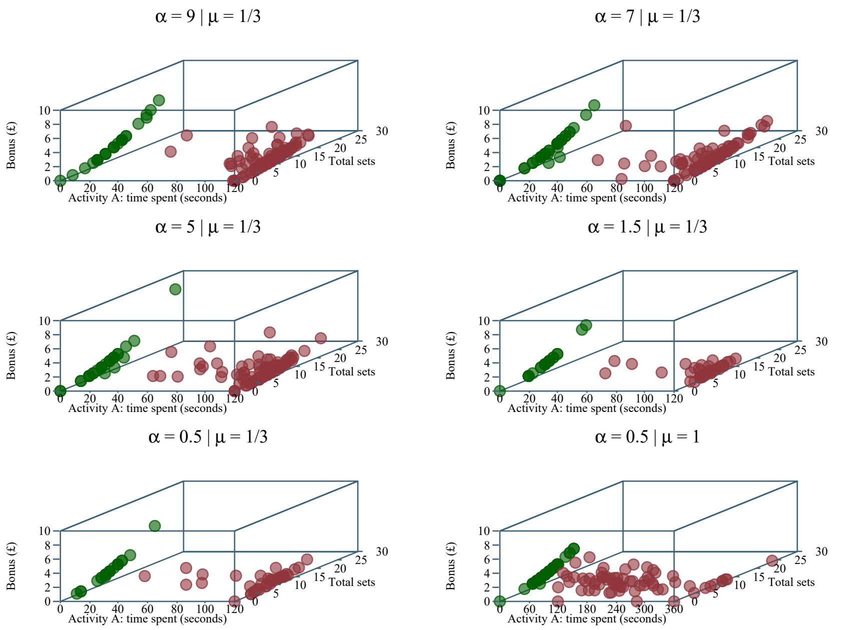
Suppose that we decide it would be acceptably consequential if the typical subject's $L\%$ was 20% of potential earnings. This implies that $z = 0.2$. Assuming $\omega = 1$ (which is expanded on below), the range of μ at which typical subject's $L\%$ would be at-least 0.2 is:

$$\frac{(2-1) - \sqrt{(2-1)^2 - 0.8}}{2} < \mu < \frac{(2-1) + \sqrt{(2-1)^2 - 0.8}}{2}$$
$$0.275 < \mu < 0.725$$

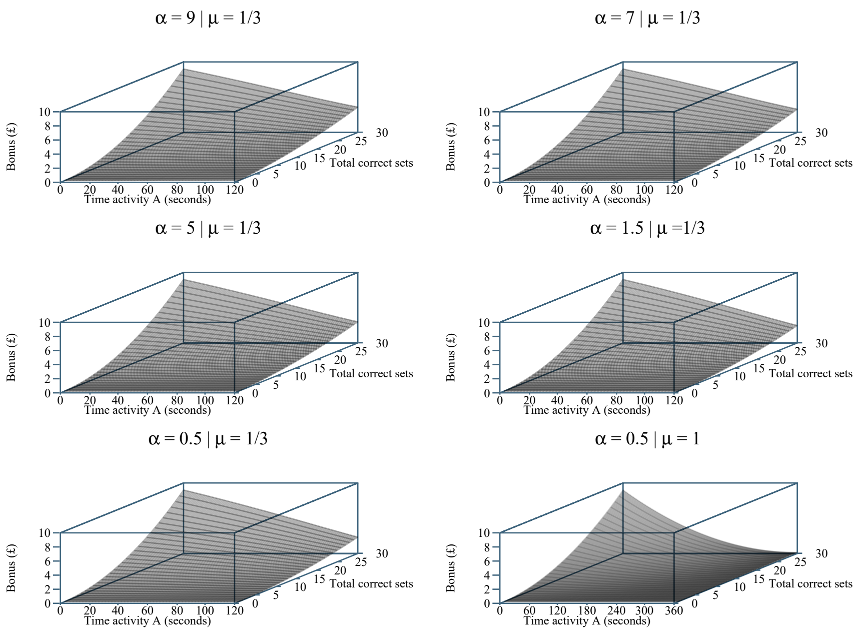
So, trying to keep μ low to hold number of subjects who met C*, a suitable candidate would be to select $= \frac{1}{3}$

B.1.3 Three-dimensional graphics

As you recall in Section 2.3 above, while we can set μ, T, n, p, α, and β parameters ourselves (exogenously determined), we cannot control the latent speed of a subject, k (endogenously determined). This highlights the fact that speed matters for whether C* holds. In this exercise (Figure B1a), we use our observed data from across the various sessions and plot time spent on activity A, the correct sets produced and the respective bonus on the x, y, and z-axis, accordingly. We then use this data to simulate another dataset that fills in all the missing data from the sessions (Figure B1b), which now looks akin to a folded sheet of paper. These graphs show how these variables interact with one-another as we change our two key parameters α and μ in our set-up. Decreasing α flattens the right-hand side of the graphs, reflecting the increased opportunity costs from working on activity A (while holding constant the earnings from spending all time on activity B). Changing μ has a dramatic impact on the earnings from activity A in this case because $\alpha = 0.5$. This means that for someone who makes 30 sets that are all in activity A, would only earn 15 ECU. If they spent these 30 sets in activity B, they would earn 900 ECU.



(a) Observed data



(b) Simulated data

Figure B1: Three-dimensional graphics

Notes. Time spent on activity A (x-axis), the correct sets produced (y-axis) and the respective bonus (z-axis). Sample sizes in observed data are N={140, 142, 142, 59, 57, 121} for $\alpha = \{9, 7, 5, 1.5, 0.5\}$ and $\mu = 1$ (pooled) respectively. Sample sizes in simulated data are N=3,630 for $\alpha = \{9, 7, 5, 1.5, 0.5\}$ and $N = 10,831$ for $\mu = 1$.

B.2 Instructions

[The experiment was programmed on LIONESS LAB (Giamattei et al., 2020) (all material available on upon request). The Two dotted lines represent a new screen in the programme. Comments in footnotes and square brackets were not seen by readers. Curly brackets indicate treatment differences.]

....
....

Welcome to this study

In this study we will ask you to complete a task.
The study consists of 3 stages: (Stage 1) Instructions, (Stage 2) Task, and (Stage 3) End of Study Survey.
You will be paid a fee of £2.00 for completing all three stages of the study plus a bonus payment depending on your performance in the task.
Please read all instructions carefully.
Note: supported devices are a laptop or PC. Your submission will be returned if you are using a mobile or tablet.

....
....

In studies like ours, there are sometimes participants who do not carefully read the instructions. This means that there are random answers which compromise the results of research studies. To show that you read our questions carefully, please enter 'Turquoise' as your answer to the next question.
Based on the text above, what is your favourite colour?

(Red)
(Yellow)
(Green)
(Turquoise)
(Other)

....
....

Stage 1: Instructions and comprehension questions

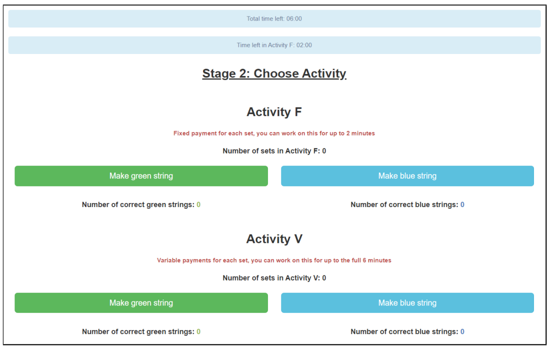
Please read the instructions carefully.

In this study, you must choose how to allocate your time across two activities: Activity F and Activity V. These activities involve the same task but differ in their payment and in their timing - more details on this below.

The task: You will be given sequences of randomly generated 6-letter strings and are required to type as many of them as you can in reverse order. As an example: If you see 'rlgows' for one string, the correct answer will be 'swoglr'.

For each activity, there will be green and blue strings. Two green and two blue together form a set. You will be paid for the number of correctly solved sets in each activity. As an example, imagine that you correctly solve 6 green and 8 blue strings in one of the activities. This gives you 3 complete sets in that activity (for which you will get paid).

The screenshot below shows how your environment will look.



Below we explain how the two activities differ with respect to payment and timing.

1) Payment:

Payment: You get paid through Experimental Currency Units (ECUs), which are later converted into a Prolific bonus (£) at a pre-defined exchange rate (100 ECU = £1.00).[1]

Activity A

Every correctly solved set is worth:	{9, 7, 5, 1.5, 0.5} ECU

Activity B

Your first correctly solved set is worth:	1 ECU
Your second correctly solved set is worth:	3 ECU
Your third correctly solved set is worth:	5 ECU
Your fourth correctly solved set is worth:	7 ECU
Your fifth correctly solved set is worth:	9 ECU
Your sixth correctly solved set is worth:	11 ECU

...and so on, with each successive set of correctly solved strings in Activity V worth 2 ECU more than the previous one.[2]

2) Timing:

In the experiment, you have a total of 6 minutes. You can freely choose how to allocate your time across the two activities. However, the time that you can spend working on each of them differs.

- Activity F: You can work on this activity for up to 2 minutes.
- Activity V: You can work on this activity for up to the full 6 minutes.

[1]Note: $1 = 1^2, 1 + 3 = 2^2, 1 + 3 + 5 = 3^2, 1 + 3 + 5 + 7 = 4^2$ and so on. The quadratic payoffs for activity B are implemented because $\forall x \geq 1, x^2 = (x-1)^2 + 2(x-1) + 1, x^2 - (x-1)^2 = 2x - 1$ The value of x sets = $2x - 1$. The x^{th} sets has incremental value of $2x - 1$ which rises by 2 as x rises by 1.

[2]The difference between experiments 1 and 2 in Study 1 is that the price sequence of activity B only went up to the fourth unit. In all remaining sessions, we implemented sets up to the sixth unit.

There will be two timers: (1) There will be a 2-minute timer that will run down only while you are working on Activity F. Once the 2 minutes have passed, you can only work on Activity V. (2) There will be a 6-minute timer that will run down only while you are working on either Activity F or Activity V. Once the 6 minutes have passed, you cannot work any more on either activity.
To summarise: You can spend between 0 and 2 minutes on Activity F, and between 4 minutes and 6 minutes on Activity V, but NO MORE THAN 6 MINUTES IN TOTAL between the two. Furthermore, you can switch freely between the two activities, subject to the availability of Activity F (as described above).
[Note that the wording in the timing section was slightly different in the $\mu = 1$ treatments. It read the following: *'In the experiment, you have a total of 6 minutes. You can freely choose how to allocate your time across the two activities. There will be a 6-minute timer that will run down only while you are working on either Activity F or Activity V. Once the 6 minutes have passed, you cannot work any more on either activity. To summarise: You can spend anywhere between 0 and 6 minutes on either activity, but NO MORE THAN 6 MINUTES IN TOTAL between the two. Furthermore, you can switch freely between the two activities.'*]

Comprehension questions

In order to move on please answer the following questions:

1. The underlying task (typing assignment) for producing sets is the same in Activity F as in Activity V.

   (Yes)[3]
   (No)

2. If you correctly reverse 8 green and 4 blue strings in an activity, you will be paid for [...] in that activity.

   (3 sets)
   (2 sets)
   (8 sets)

3. If you correctly reverse 8 green strings in Activity F and 4 blue strings Activity V, you will be paid for [...].

   (0 sets)
   (2 sets)
   (8 sets)

4. Imagine that you had already correctly solved six sets in Activity F. How much would it add to your earnings (in ECUs) if you solved a seventh set for that activity?

   {9, 7, 5, 1.5, 0.5} ECU
   (49 ECU)
   (13 ECU)

5. Imagine that you solve exactly seven sets in Activity F, how much would you earn in total from this activity?[4]

[3] Correct responses are in green.
[4] The correct answer for questions 4 and 5 depends on what the value of α was set to.

(63 ECU)
{63, 49, 35, 10.5, 3.5} ECU
(7 ECU)

6. Imagine that you had already correctly solved six sets in Activity V. How much would it add to your earnings (in ECUs) if you solved a seventh set for that activity?

(7 ECU)
(49 ECU)
(13 ECU)

7. Imagine that you solve exactly seven sets in Activity V, how much would you earn in total from this activity?

(49 ECU)
(63 ECU)
(13 ECU)

8. In this experiment, you can work on Activity F for up to 2 minutes and Activity V for up to the full 6 minutes. In total, you have:

(6 minutes)
(8 minutes)
(None of the above)

9. Will the timer on an activity run down while you are not working on it?

(Yes)
(No)

. . . .
. . . .

Practice string production

Correctly type up as many reverse text sequences as possible. This is just for practice, not for payment. When you feel ready, please click 'Continue' to proceed.
To submit a reversed text sequence, either press the 'Submit' button or your 'Enter' key.

Total strings attempted: 0
Number of correct strings: 0

Text sequence: **ylqhcv**

Your answer: [input]

. . . .
. . . .

Congratulations, you have completed the comprehension questions and can now move on to Stage 2.

In Stage 2, you will begin the task. You need to click on 'Make green string' and 'Make blue string' buttons to reverse strings and subsequently make as many sets of strings as possible.
Remember that two green and two blue together in the same activity form a set. You will be paid for the number of correctly solved sets in each activity.
Two green and two blue strings from different activities DO NOT form a set.
When you are ready, please click 'Continue' to proceed to Stage 2, where you will begin working on the activities of your choice.

Task loading

Stage 2: Choose activity

[Task environment]

Congratulations, you have completed Stage 2.
When you are ready, please click 'Continue' to proceed to Stage 3, where you will begin answering the end of study survey.

Stage 3: Questionnaire
Part 1/4

[In this section, subjects complete a battery of questions pertaining to (i) reasons for task choice, (ii) perceived attention on timing and payments, (iii) perceived busyness in the task, (iv) perceived urgency of activity A, (v) numeric ability (Lipkus et al. 2001; Bitterly et al. 2020), (vi) perceived maths ability (Koch & Nafziger, 2019), (vii) an alternate form of the cognitive reflection test (Thomson & Oppenheimer, 2016), (viii) lottery isolation (Tversky & Kahneman, 1985), (ix) an unincentivised measure of risk preferences (Falk et al., 2018), and (x) general demographics (e.g. age, gender and education).]

1. Please briefly explain how you decided which of the activities to work on in this study.

   (open ended response)

2. While deciding which of the activities to spend time on, to what extent did you pay attention to the payment?

(not at all ← 1 to 8 → very much)

3. While deciding which of the activities to spend time on, to what extent did you pay attention to the time that could be spent on them?

(not at all ← 1 to 8 → very much)

4. How urgent did Activity F feel? (This activity had a fixed payment for each set, and you could work on it for up to 2 {6} minutes)

(not urgent at all ← 1 to 8 → very urgent)

5. In general, how busy did you feel during this experiment?[5]

(not busy at all ← 1 to 8 → very busy)

6. Overall, did you correctly reverse more or less strings in the task than you originally expected to be able to do in the full 6 minutes?

(expected to do much less ← 1 to 8 → expected to do much more)

7. Do you think you divided your time optimally (to earn the highest number of ECUs) between Activities F and V?

(Yes)
(No)
(Not sure)

8. If you were to do the task again, how much time would you spend on Activity F? (This activity had a fixed payment for each set, and you could work on it for up to 2 {6} minutes)

(0 to {120, 360} seconds)

. . . .
. . . .

Part 2/4

[Related to numeric ability (Bitterly et al., 2020; Lipkus et al., 2001)]

1. The government of Moneyland discovered a flaw in their currency, the Gar. They need to reprint the Gars in circulation. The number of Gars that the government prints every day increases:
   - On day 1, they print 1 new Gar
   - On day 2, they print 3 new Gars
   - On day 3, they print 5 new Gars

[5] In Chapter 3, we tweaked this question slightly to *'In general, how busy did you feel during the task?'*. We did this because we wanted a more specific measure of perceived busyness while choosing to allocate time. With the benefit of hindsight, this question could be construed as 'busyness' in the string-reversal task (i.e., a measure of within component busyness) and not more generally across the various activities worked.

And so on… with each successive day the government increasing their printing by 2 Gars. If the government continues printing at this rate, how many Gars IN TOTAL will they have printed by the end of day 7 (including the printing on day 7)?

(numeric response)

2. Consider the following diseases and the numbers of cases faced for each disease from January through April. Each is equally dangerous, but the rate of transmission differs from disease to disease. Following the current growth paths, Which disease will be the most serious threat in May?

Month	Disease A	Disease B	Disease C	Disease D
January	100	100	400	800
February	500	200	700	1200
March	900	400	1100	1500
April	1300	800	1600	1700
May	?	?	?	?

(Disease A)
(Disease B)
(Disease C)
(Disease D)

3. Please indicate your answer on a scale from 0 to 10, where 0 means 'does not describe me at all' and 10 means 'describes me perfectly'.

I am good at maths.

(does not describe me at all ← 0 to 10 → describes me perfectly)

. . . .

Part 3/4

[Cognitive reflection test 2 (Thomson & Oppenheimer, 2016) and lottery isolation (Tversky & Kahneman, 1985)]

1. If you're running a race and you pass the person in second place, what place are you in?

(First)
(Second)
(Third)

2. A farmer had 15 sheep and all but 8 died. How many are left?

(numeric response)

3. How many cubic metres of dirt are there in a hole that is 3m deep x 3m wide x 3m long?

(numeric response)

4. Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

   Decision (1): Choose between (before answering, read Decision (2)):

   (A) a sure gain of £240
   (B) 25% chance to gain £1000, and 75% chance to gain nothing

   (A)
   (B)

   Decision (2): Choose between:

   (C) a sure loss of £750
   (D) 75% chance to lose £1000, and 25% chance to lose nothing

   (C)
   (D)

. . . .
. . . .

Part 4/4

[An unincentivised measures of risk-taking from the *Global Evidence on Economic Preferences* (Falk et al., 2018). The authors measure these preferences through a combination of two survey measures, one quantitative and another qualitative. The quantitative consists of a series of five interdependent hypothetical binary choices employing the staircase method. The qualitative, on the other hand, is a participant's self-assessed level of a subject's willingness to take risks in general.]

1. Please tell me, in general, how willing or unwilling you are to take risks, using a scale from 0 to 10, where 0 means you are 'completely unwilling to take risks' and 10 means you are 'very willing to take risks'.

   (completely unwilling to take risks ← 0 to 10 → very willing to take risks)

2. Please imagine the following situation. You can choose between a sure payment of a particular amount of money, OR a draw, where you would have an equal chance (50/50 chance) of getting £150 or getting nothing. We will present you with five different situations.

   Situation 1:
   What would you prefer: a draw with a 50 percent chance of receiving £150 and the same 50 percent chance of receiving nothing, OR the amount of £80 as a sure payment?
   .
   .
   .

Demographics

1. What is your gender?

(Male)
(Female)
(Prefer not to say)

2. What is your age?

(numeric input)

3. What is your race?

(White)
(Mixed)
(Asian or Asian British)
(Black or Black British)
(Chinese)
(Any other ethnic group)
(Prefer not to say)

4. What is your native language?

(text input)

5. What's your country of primary citizenship?

(text input)

6. What is the highest degree or level of school you have completed?

(No formal qualifications)
(Secondary school/GCSE)
(College/A levels)
(Undergraduate degree (BA/BSc/other))
(Graduate degree (MA/MSc/MPhil/other))
(Doctorate degree (PhD/MD/other))
(Prefer not to say)

7. What is your household pre-tax income last year?

(Less than £10,000)
(£10,000-£19,999)
(£20,000-£49,999)
(£50,000-£79,999)
(£80,000-£149,999)
(More than £150,000)
(Prefer not to say)

8. Including yourself, how many people currently live in your household?

(text input)

9. Any feedback regarding this study is highly appreciated. Please leave comments below:

(text input)

. . . .
. . . .

Feedback

Below you can find an overview of your performance in the experiment, and your subsequent earnings:
Number of Activity F sets that you created: [x], earning you: {9,7,5,1.5,0.5}([x]) = [x] ECU.
Number of Activity V sets that you created: [x], earning you: $[x]^2$ = [x] ECU.
Total ECU earned, [x] ECU
Exchange rate: 100 ECU = £1.00.
YOUR BONUS: £[x]
YOUR FIXED WAGE: £[x]
IN TOTAL, YOU RECEIVE £[x]

. . . .
. . . .

B.3 Power calculations

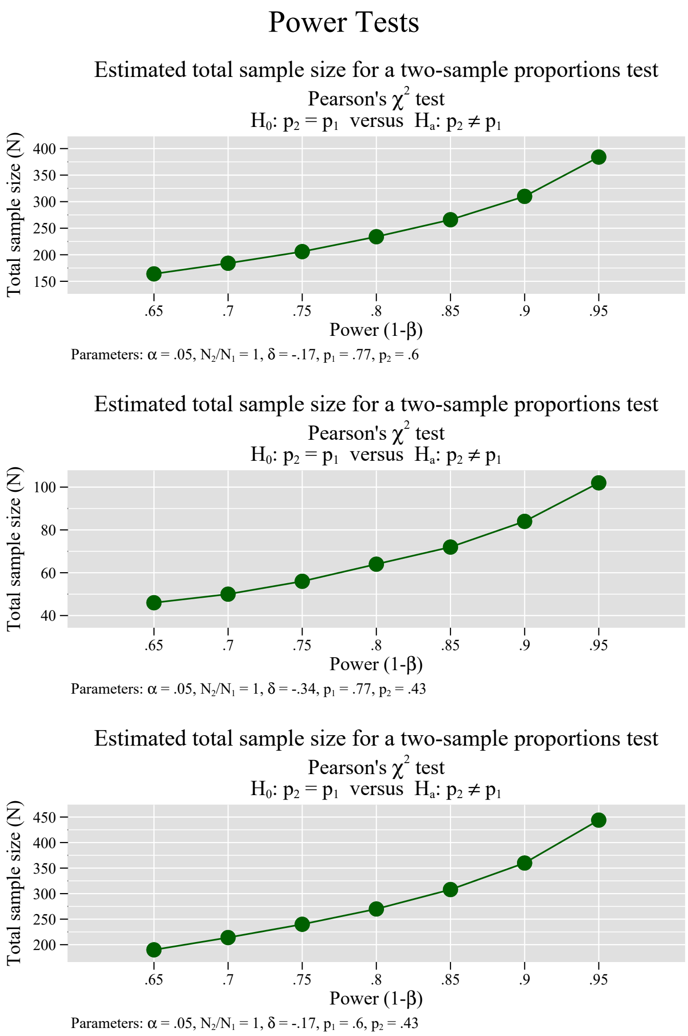


Figure B2: Power calculations for $\alpha = \{9, 7, 5\}$

Notes. These power calculations are for $\alpha = \{9, 7, 5\}$ and show what power we would obtain under various total sample sizes (N). The test for each is a two-sample χ^2 test. In the $\alpha = 9$ treatment in Study 1, 0.77 of the sample follows NB behaviour. Given this, we can make an assumption by how much we might expect the proportion of NB to fall across $\alpha = \{9, 7, 5\}$. Comparing $\alpha = \{9, 7\}$, as a *rule-of-thumb*, we might expect the proportion of NBs to decrease to $0.77(\frac{7}{9}) = 0.60$. Between $\alpha = \{9, 5\}$, we might expect $0.77(\frac{5}{9}) = 0.43$ (likewise for $\alpha = \{9, 7\}$, $0.60(\frac{7}{9}) = 0.43$).

B.4 Experiments

B.4.1 String performance from Chapter 1

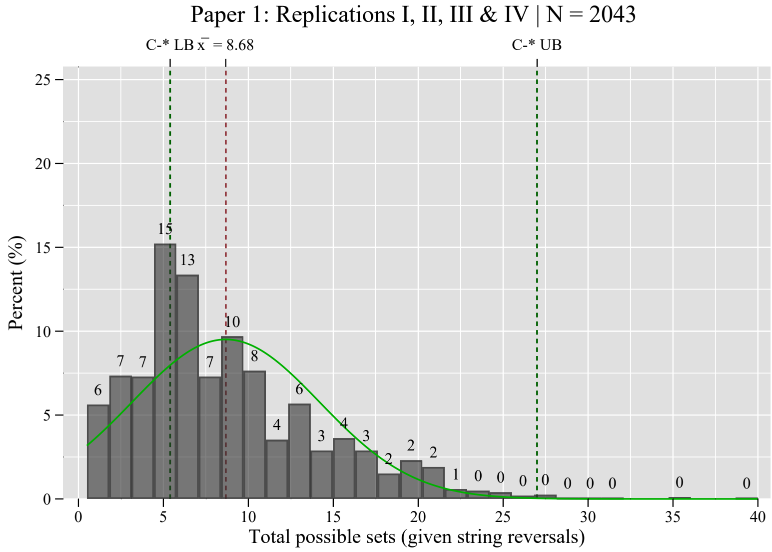


Figure B3: Performance for Chapter 1

Notes. $\bar{x}$ represents the mean. The performance is from the 3-minute unincentivised string-reversal task across Replications I, II, III and IV. We used Python to sort through the text-entry responses, generating three variables: (i) total *attempted* string-reversals, (ii) total *correct* string-reversals, and (iii) total *incorrect* string-reversals. To obtain the number of 'possible' sets produced in 6-minutes, we simply multiplied the total number of correct string-reversals over the 3-minutes by two (as the task in Chapter 2 is 6-minutes) and then divided by four (as one set requires four correctly reversed strings).

B.4.2 Summary statistics

Study 1: summary statistics

	Mean	SD	Min	Max	N
Performance					
Total strings	44.92	14.16	23.00	79.00	83.00
Possible sets	10.88	3.58	5.00	19.00	83.00
Total sets	9.87	4.36	0.00	19.00	83.00
Time allocation					
Time A (> 20 seconds)	0.89	0.31	0.00	1.00	83.00
Earnings					
Earnings (narrow) £ p/h	7.65	3.33	1.45	16.56	65.00
Earnings (broad) £ p/h	9.04	3.75	4.70	16.29	9.00
Demographics					
Female	0.66	0.48	0.00	1.00	82.00
Age	35.69	12.83	18.00	65.00	83.00
Tertiary education	0.51	0.50	0.00	1.00	83.00
Comprehension					
Comprehension attempts	0.47	0.50	0.00	1.00	83.00
Attention					
Attention timings	6.34	1.96	1.00	8.00	83.00
Attention payments	4.82	2.42	1.00	8.00	83.00
Attention difference	1.52	2.59	-5.00	7.00	83.00
Cognition					
Perceived maths ability	5.02	2.85	0.00	10.00	83.00
Numeric ability	1.33	0.65	0.00	2.00	83.00
Cognitive reflection	1.69	0.74	0.00	3.00	81.00
Lottery isolation	0.55	0.50	0.00	1.00	83.00
Other					
Risk taking (standardised)	0.00	0.79	-2.19	1.94	83.00
Expectation	5.23	1.62	1.00	8.00	83.00

Table B1: Summary statistics for Study 1

Study 2: summary statistics

	Mean	SD	Min	Max	N
$\alpha = 9$					
Performance					
Total strings	48.12	16.34	0.00	98.00	140.00
Possible sets	11.69	4.07	0.00	24.00	140.00
Total sets	10.38	4.76	0.00	24.00	140.00
Time allocation					
Time A (> 20 seconds)	0.81	0.39	0.00	1.00	140.00
Earnings					
Earnings (narrow) £ p/h	8.69	3.44	2.89	19.05	100.00
Earnings (broad) £ p/h	13.27	6.71	4.02	33.06	24.00
Demographics					
Female	0.64	0.48	0.00	1.00	140.00
Age	36.01	11.25	19.00	68.00	140.00
Tertiary education	0.60	0.49	0.00	1.00	140.00
Comprehension					
Comprehension attempts	0.45	0.50	0.00	1.00	140.00
Attention					
Attention timings	6.47	1.84	1.00	8.00	140.00
Attention payments	5.52	2.34	1.00	8.00	140.00
Attention difference	0.95	2.82	-7.00	7.00	140.00
Cognition					
Perceived maths ability	5.86	2.46	0.00	10.00	140.00
Numeric ability	1.45	0.66	0.00	2.00	140.00
Cognitive reflection	1.77	0.84	0.00	3.00	137.00
Lottery isolation	0.69	0.47	0.00	1.00	140.00
Other					
Activity A (top of screen)	0.49	0.50	0.00	1.00	140.00
Risk taking (standardised)	0.03	0.72	-1.89	2.48	140.00
Expectation	4.97	1.64	1.00	8.00	140.00
$\alpha = 7$					
Performance					
Total strings	47.75	16.88	1.00	103.00	142.00
Possible sets	11.54	4.23	0.00	25.00	142.00
Total sets	10.13	5.08	0.00	23.00	142.00
Time allocation					
Time A (> 20 seconds)	0.73	0.44	0.00	1.00	142.00
Earnings					
Earnings (narrow) £ p/h	7.99	2.60	2.50	14.82	87.00
Earnings (broad) £ p/h	11.29	3.50	5.12	20.98	32.00
Demographics					
Female	0.69	0.46	0.00	1.00	142.00
Age	37.46	12.36	18.00	67.00	142.00
Tertiary education	0.62	0.49	0.00	1.00	141.00
Comprehension					
Comprehension attempts	0.47	0.50	0.00	1.00	142.00
Attention					
Attention timings	6.37	2.01	1.00	8.00	141.00
Attention payments	5.39	2.32	1.00	8.00	141.00
Attention difference	0.98	2.72	-7.00	7.00	141.00
Cognition					
Perceived maths ability	5.61	2.57	0.00	10.00	142.00
Numeric ability	1.47	0.63	0.00	2.00	142.00
Cognitive reflection	1.79	0.82	0.00	3.00	139.00
Lottery isolation	0.75	0.44	0.00	1.00	142.00
Other					
Activity A (top of screen)	0.50	0.50	0.00	1.00	142.00
Risk taking (standardised)	-0.01	0.81	-2.01	2.29	142.00
Expectation	4.93	1.68	1.00	8.00	141.00
$\alpha = 5$					
Performance					
Total strings	44.59	14.44	12.00	115.00	140.00
Possible sets	10.81	3.60	3.00	28.00	140.00
Total sets	9.78	4.09	0.00	28.00	140.00
Time allocation					
Time A (> 20 seconds)	0.80	0.40	0.00	1.00	140.00
Earnings					
Earnings (narrow) £ p/h	7.65	2.57	2.89	15.98	88.00
Earnings (broad) £ p/h	8.68	2.59	4.22	13.11	22.00
Demographics					
Female	0.71	0.46	0.00	1.00	138.00
Age	36.59	13.57	18.00	71.00	140.00
Tertiary education	0.68	0.47	0.00	1.00	140.00
Comprehension					
Comprehension attempts	0.46	0.50	0.00	1.00	140.00
Attention					
Attention timings	6.24	1.90	1.00	8.00	140.00
Attention payments	5.07	2.34	1.00	8.00	140.00
Attention difference	1.16	2.74	-7.00	7.00	140.00
Cognition					
Perceived maths ability	5.57	2.47	0.00	9.00	140.00
Numeric ability	1.49	0.65	0.00	2.00	140.00
Cognitive reflection	1.88	0.74	0.00	3.00	139.00
Lottery isolation	0.67	0.47	0.00	1.00	140.00
Other					
Activity A (top of screen)	0.50	0.50	0.00	1.00	140.00
Risk taking (standardised)	-0.02	0.79	-2.01	1.65	140.00
Expectation	5.13	1.67	1.00	8.00	140.00
Total					
Performance					
Total strings	46.82	15.97	0.00	115.00	422.00
Possible sets	11.35	3.98	0.00	28.00	422.00
Total sets	10.09	4.66	0.00	28.00	422.00
Time allocation					
Time A (> 20 seconds)	0.78	0.41	0.00	1.00	422.00
Earnings					
Earnings (narrow) £ p/h	8.14	2.95	2.50	19.05	275.00
Earnings (broad) £ p/h	11.16	4.83	4.02	33.06	78.00
Demographics					
Female	0.68	0.47	0.00	1.00	420.00
Age	36.69	12.41	18.00	71.00	422.00
Tertiary education	0.63	0.48	0.00	1.00	421.00
Comprehension					
Comprehension attempts	0.46	0.50	0.00	1.00	422.00
Attention					
Attention timings	6.36	1.92	1.00	8.00	421.00
Attention payments	5.33	2.34	1.00	8.00	421.00
Attention difference	1.03	2.75	-7.00	7.00	421.00
Cognition					
Perceived maths ability	5.68	2.50	0.00	10.00	422.00
Numeric ability	1.47	0.65	0.00	2.00	422.00
Cognitive reflection	1.81	0.80	0.00	3.00	415.00
Lottery isolation	0.70	0.46	0.00	1.00	422.00
Other					
Activity A (top of screen)	0.50	0.50	0.00	1.00	422.00
Risk taking (standardised)	-0.00	0.78	-2.01	2.48	422.00
Expectation	5.01	1.66	1.00	8.00	421.00

Table B2: Summary statistics for Study 2

Study 3: summary statistics

	Mean	SD	Min	Max	N
$\alpha = 1.5$					
Performance					
Total strings	46.20	13.41	0.00	87.00	59.00
Possible sets	11.20	3.25	0.00	21.00	59.00
Total sets	10.39	3.29	0.00	21.00	59.00
Time allocation					
Time A (> 20 seconds)	0.64	0.48	0.00	1.00	59.00
Earnings					
Earnings ($TimeA$ > 20 seconds) £ p/h	8.39	2.33	3.54	12.52	38.00
Earnings ($TimeA \leq$ 20 seconds) £ p/h	11.33	4.86	4.16	21.09	20.00
Demographics					
Female	0.61	0.49	0.00	1.00	59.00
Age	31.29	14.18	18.00	70.00	59.00
Tertiary education	0.60	0.49	0.00	1.00	58.00
Comprehension					
Comprehension attempts	0.39	0.49	0.00	1.00	59.00
Attention					
Attention timings	6.00	2.09	1.00	8.00	59.00
Attention payments	5.59	2.46	1.00	8.00	59.00
Attention difference	0.41	3.16	-7.00	7.00	59.00
Cognition					
Perceived maths ability	5.58	2.65	0.00	10.00	59.00
Numeric ability	1.31	0.77	0.00	2.00	59.00
Cognitive reflection	1.77	0.78	0.00	3.00	57.00
Lottery isolation	0.59	0.50	0.00	1.00	59.00
Other					
Activity A (top of screen)	0.49	0.50	0.00	1.00	59.00
Risk taking (standardised)	0.01	0.74	-1.59	2.09	59.00
Expectation	4.88	1.74	2.00	8.00	59.00
$\alpha = 0.5$					
Performance					
Total strings	48.84	12.90	19.00	93.00	57.00
Possible sets	11.86	3.25	4.00	23.00	57.00
Total sets	10.74	3.98	0.00	23.00	57.00
Time allocation					
Time A (> 20 seconds)	0.63	0.49	0.00	1.00	57.00
Earnings					
Earnings ($TimeA$ > 20 seconds) £ p/h	8.47	3.75	1.82	18.36	35.00
Earnings ($TimeA \leq$ 20 seconds) £ p/h	13.75	9.37	4.73	45.64	21.00
Demographics					
Female	0.74	0.44	0.00	1.00	57.00
Age	30.61	9.93	18.00	56.00	57.00
Tertiary education	0.51	0.50	0.00	1.00	57.00
Comprehension					
Comprehension attempts	0.46	0.50	0.00	1.00	57.00
Attention					
Attention timings	6.81	1.53	3.00	8.00	57.00
Attention payments	5.53	2.42	1.00	8.00	57.00
Attention difference	1.28	2.89	-5.00	7.00	57.00
Cognition					
Perceived maths ability	5.89	2.70	0.00	10.00	57.00
Numeric ability	1.28	0.73	0.00	2.00	57.00
Cognitive reflection	1.68	0.99	0.00	3.00	56.00
Lottery isolation	0.77	0.42	0.00	1.00	57.00
Other					
Activity A (top of screen)	0.51	0.50	0.00	1.00	57.00
Risk taking (standardised)	0.00	0.82	-1.48	2.11	57.00
Expectation	4.58	1.56	1.00	8.00	57.00
Total					
Performance					
Total strings	47.50	13.17	0.00	93.00	116.00
Possible sets	11.53	3.26	0.00	23.00	116.00
Total sets	10.56	3.64	0.00	23.00	116.00
Time allocation					
Time A (> 20 seconds)	0.64	0.48	0.00	1.00	116.00
Earnings					
Earnings ($TimeA$ > 20 seconds) £ p/h	8.43	3.07	1.82	18.36	73.00
Earnings ($TimeA \leq$ 20 seconds) £ p/h	12.57	7.52	4.16	45.64	41.00
Demographics					
Female	0.67	0.47	0.00	1.00	116.00
Age	30.96	12.23	18.00	70.00	116.00
Tertiary education	0.56	0.50	0.00	1.00	115.00
Comprehension					
Comprehension attempts	0.42	0.50	0.00	1.00	116.00
Attention					
Attention timings	6.40	1.87	1.00	8.00	116.00
Attention payments	5.56	2.43	1.00	8.00	116.00
Attention difference	0.84	3.05	-7.00	7.00	116.00
Cognition					
Perceived maths ability	5.73	2.67	0.00	10.00	116.00
Numeric ability	1.29	0.75	0.00	2.00	116.00
Cognitive reflection	1.73	0.89	0.00	3.00	113.00
Lottery isolation	0.68	0.47	0.00	1.00	116.00
Other					
Activity A (top of screen)	0.50	0.50	0.00	1.00	116.00
Risk taking (standardised)	0.01	0.78	-1.59	2.11	116.00
Expectation	4.73	1.65	1.00	8.00	116.00

Table B3: Summary statistics for Study 3a

	Mean	SD	Min	Max	N
$\mu = 1$					
Performance					
Total strings	47.84	12.27	24.00	98.00	58.00
Possible sets	11.62	3.08	6.00	24.00	58.00
Total sets	10.95	3.88	0.00	24.00	58.00
Time allocation					
Time A (> 20 seconds)	0.57	0.50	0.00	1.00	58.00
Earnings					
Earnings ($TimeA > 20$ seconds) £ p/h	7.34	2.40	2.96	13.21	32.00
Earnings ($TimeA \leq 20$ seconds) £ p/h	13.17	3.96	7.02	23.20	24.00
Demographics					
Female	0.74	0.44	0.00	1.00	57.00
Age	29.33	9.61	19.00	61.00	58.00
Tertiary education	0.50	0.50	0.00	1.00	58.00
Comprehension					
Comprehension attempts	0.45	0.50	0.00	1.00	58.00
Attention					
Attention timings	6.17	1.93	1.00	8.00	58.00
Attention payments	6.36	1.93	1.00	8.00	58.00
Attention difference	-0.19	2.68	-7.00	7.00	58.00
Cognition					
Perceived maths ability	5.14	2.72	0.00	10.00	58.00
Numeric ability	1.21	0.72	0.00	2.00	58.00
Cognitive reflection	1.63	0.84	0.00	3.00	57.00
Lottery isolation	0.60	0.49	0.00	1.00	57.00
Other					
Activity A (top of screen)	0.52	0.50	0.00	1.00	58.00
Risk taking (standardised)	-0.04	0.85	-1.73	1.49	58.00
Expectation	4.86	1.70	1.00	8.00	58.00
$\mu = 1$ (Highlight)					
Performance					
Total strings	45.02	12.25	0.00	75.00	62.00
Possible sets	10.92	3.07	0.00	18.00	62.00
Total sets	10.52	3.49	0.00	18.00	62.00
Time allocation					
Time A (> 20 seconds)	0.56	0.50	0.00	1.00	62.00
Earnings					
Earnings ($TimeA > 20$ seconds) £ p/h	7.82	2.81	1.76	13.15	33.00
Earnings ($TimeA \leq 20$ seconds) £ p/h	10.73	4.40	2.91	17.98	27.00
Demographics					
Female	0.74	0.44	0.00	1.00	61.00
Age	31.87	9.94	18.00	59.00	62.00
Tertiary education	0.52	0.50	0.00	1.00	62.00
Comprehension					
Comprehension attempts	0.48	0.50	0.00	1.00	62.00
Attention					
Attention timings	5.90	2.12	1.00	8.00	62.00
Attention payments	6.00	2.17	1.00	8.00	62.00
Attention difference	-0.10	2.99	-7.00	7.00	62.00
Cognition					
Perceived maths ability	5.87	2.59	1.00	10.00	62.00
Numeric ability	1.50	0.67	0.00	2.00	62.00
Cognitive reflection	1.72	0.81	0.00	3.00	58.00
Lottery isolation	0.65	0.48	0.00	1.00	62.00
Other					
Activity A (top of screen)	0.50	0.50	0.00	1.00	62.00
Risk taking (standardised)	0.02	0.84	-1.82	2.15	62.00
Expectation	4.73	1.78	1.00	8.00	62.00
Total					
Performance					
Total strings	46.38	12.29	0.00	98.00	120.00
Possible sets	11.26	3.08	0.00	24.00	120.00
Total sets	10.72	3.67	0.00	24.00	120.00
Time allocation					
Time A (> 20 seconds)	0.57	0.50	0.00	1.00	120.00
Earnings					
Earnings($TimeA > 20$ seconds) £ p/h	7.58	2.61	1.76	13.21	65.00
Earnings ($TimeA \leq 20$ seconds) £ p/h	11.88	4.34	2.91	23.20	51.00
Demographics					
Female	0.74	0.44	0.00	1.00	118.00
Age	30.64	9.82	18.00	61.00	120.00
Tertiary education	0.51	0.50	0.00	1.00	120.00
Comprehension					
Comprehension attempts	0.47	0.50	0.00	1.00	120.00
Attention					
Attention timings	6.03	2.02	1.00	8.00	120.00
Attention payments	6.17	2.05	1.00	8.00	120.00
Attention difference	-0.14	2.83	-7.00	7.00	120.00
Cognition					
Perceived maths ability	5.52	2.67	0.00	10.00	120.00
Numeric ability	1.36	0.71	0.00	2.00	120.00
Cognitive reflection	1.68	0.82	0.00	3.00	115.00
Lottery isolation	0.62	0.49	0.00	1.00	119.00
Other					
Activity A (top of screen)	0.51	0.50	0.00	1.00	120.00
Risk taking (standardised)	-0.01	0.84	-1.82	2.15	120.00
Expectation	4.79	1.74	1.00	8.00	120.00

Table B4: Summary statistics for Study 3b

Data pooled: summary statistics

	$\alpha = \{9, 7, 5\}$	$\alpha = \{1.5, 0.5\}$	$\mu = 1$
Performance			
Total strings	46.51 (15.69)	47.50 (13.17)	46.38 (12.29)
Possible sets	11.27 (3.92)	11.53 (3.26)	11.26 (3.08)
Total sets	10.06 (4.61)	10.56 (3.64)	10.72 (3.67)
Time allocation			
Time A (> 20 seconds)	0.80 (0.40)	0.64 (0.48)	0.57 (0.50)
Earnings			
Earnings ($TimeA > 20$ seconds) £ p/h	8.39 (3.23)	8.43 (3.07)	7.58 (2.61)
Earnings ($TimeA \leq 20$ seconds) £ p/h	11.04 (4.74)	12.57 (7.52)	11.88 (4.34)
Demographics			
Female	0.68 (0.47)	0.67 (0.47)	0.74 (0.44)
Age	36.53 (12.48)	30.96 (12.23)	30.64 (9.82)
Tertiary education	0.61 (0.49)	0.56 (0.50)	0.51 (0.50)
Comprehension			
Comprehension attempts	0.46 (0.50)	0.42 (0.50)	0.47 (0.50)
Attention			
Attention timings	6.36 (1.92)	6.40 (1.87)	6.03 (2.02)
Attention payments	5.24 (2.36)	5.56 (2.43)	6.17 (2.05)
Attention difference	1.11 (2.73)	0.84 (3.05)	-0.14 (2.83)
Cognition			
Perceived maths ability	5.57 (2.57)	5.73 (2.67)	5.52 (2.67)
Numeric ability	1.45 (0.65)	1.29 (0.75)	1.36 (0.71)
Cognitive reflection	1.78 (0.80)	1.70 (0.90)	1.65 (0.82)
Lottery isolation	0.68 (0.47)	0.68 (0.47)	0.62 (0.49)
Other			
Risk taking (standardised)	-0.10 (0.78)	0.22 (0.75)	0.21 (0.81)
Expectation	5.05 (1.66)	4.73 (1.65)	4.79 (1.74)

Table B5: Summary statistics for Studies 1-3 (pooled)

B.4.3 Study 1: Supplementary results

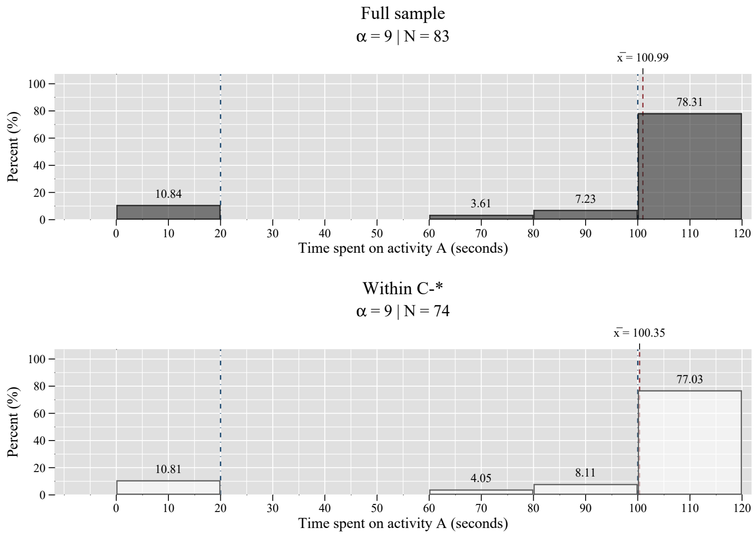


Figure B4: Time on activity A for $\alpha = 9$ (actual sets created)

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. Within C* Sample restricted to total correct sets made. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off.

B.4.4 Study 2: Supplementary results

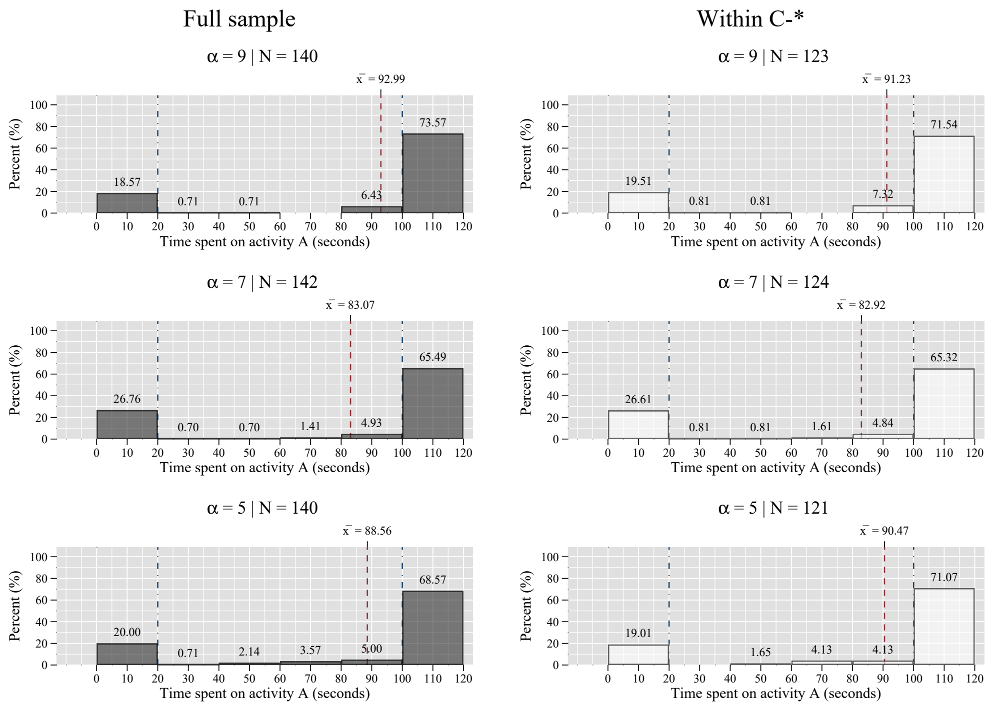


Figure B5: Time on activity A for $\alpha = \{9, 7, 5\}$ (actual sets created)

Notes. $\overline{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. Within C* Sample restricted to total correct sets made. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off.

	Full sample			High comprehension	
	(1) ME	(2) ME	(3) ME	(4) ME	(5) ME
TREATMENTS					
Base. $\alpha = 9$					
$\alpha = 7$	-0.07	-0.07	-0.04	-0.16**	-0.11
	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)
$\alpha = 5$	-0.01	-0.01	-0.03	-0.08	-0.09
	(0.05)	(0.05)	(0.05)	(0.07)	(0.08)
COMPREHENSION OF INSTRUCTIONS					
0 - 1 comprehension mistakes		-0.11***	-0.08*		
		(0.04)	(0.04)		
Additional controls	No	No	Yes	No	Yes
# Observations	393	393	389	193	190
Psuedo R^2	0.01	0.02	0.10	0.02	0.11
Log-lik	-202.70	-199.29	-180.04	-110.39	-98.21
χ^2	2.44	8.69	37.38	4.06	20.39

Table B6: Marginal effects: treatments and comprehension

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: A binary measure taking on a value of 1 if $Time_A > 20$ seconds, and 0 otherwise. Independent measures include the three treatments: $\alpha = \{9, 7, 5\}$ with $\alpha = 9$ as the base; and comprehension as a binary measure (0-1 mistakes). Additional controls include age, gender, tertiary education, subject performance in terms of total strings correctly reversed, activity order, and expectations about performance. The 'Full sample' is restricted to those who generated more than zero sets and had a speed classification within or above C*. On the other hand, the 'High comprehension' refers to those in the full sample with comprehension attempts ranging from 0-1 (instead of 2-3).

Closeness to optimality

Figure B6 provides an additional metric regarding how close subjects are to what would be considered optimal behaviour with respect to monetary maximisation. We can determine how close a subject's behaviour is to 'optimal' by evaluating their possible sets generated and how they spent their time. This measure is less restrictive as it includes those who fall outside of C*. To elaborate, for subjects who have a very slow string-reversal speed (below C*), then their optimal behaviour is to spend as much time on activity A as possible. However, for subjects with sufficiently high string-reversal speed (within or above C*), their optimal behaviour is to spend no time on activity A at all. Therefore, the measure is calculated as follows:

$$\text{Speeds in or above C* range: closeness-to-optimality} = \left(\frac{120 - Time_A}{120}\right)$$

$$\text{Speeds below C* range: closeness-to-optimality} = \left(\frac{Time_A}{120}\right)$$

Those who completely misunderstood the nature of the weakest link structure across the three treatments are excluded. This estimate ranges from 0 (fully suboptimal) to 1 (fully optimal). The averages for this measure across the three treatments are 0.24, 0.31, and 0.27 respectively, showing that across the treatments, behaviour is far from optimal. Akin to our results above, several two-sample Wilcoxon rank-sum tests show that mean optimality is not statistically different across the three levels of α. These results highlight the severity of suboptimal behaviour in our set-up as well as its apparent 'stickiness' across the treatments.

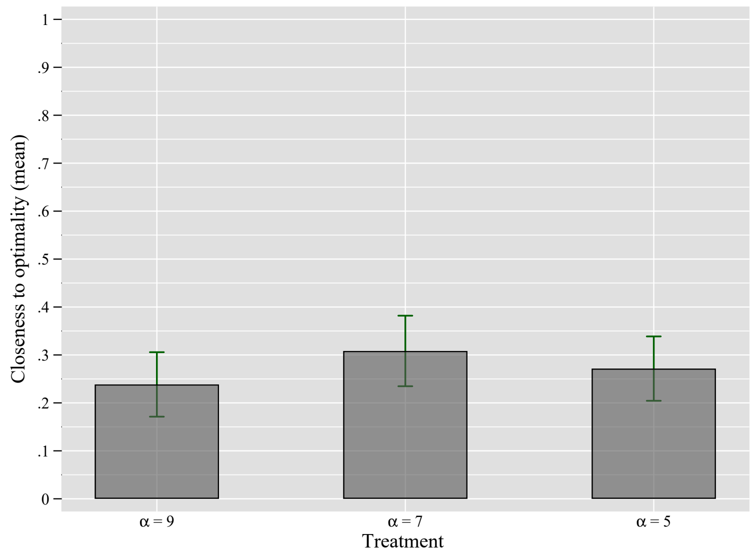


Figure B6: Closeness to optimality estimates for $\alpha = \{9, 7, 5\}$

Notes. Whiskers represent the 95% confidence intervals.

B.4.5 Study 3: Supplementary results

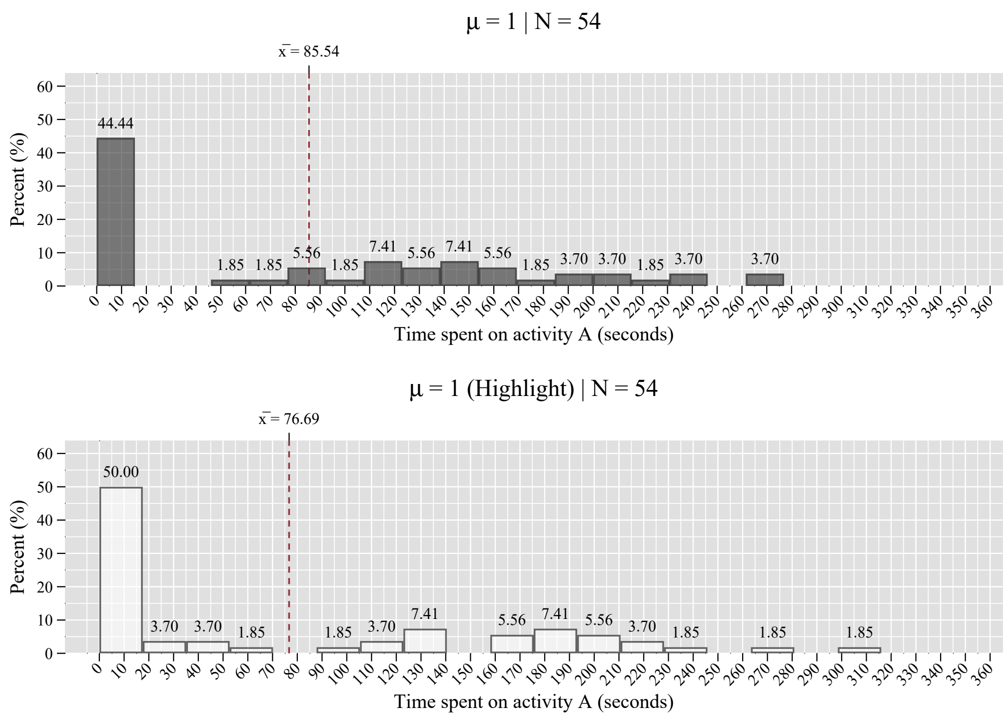


Figure B7: Time on activity A for $\mu = 1$ (excluding those who spent 360 seconds)

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. Note that there are no dash-dots as in the previous time on activity A figures, since we have moved away from the concept of bracketing as outlined in our theoretical framework.

B.4.6 Studies pooled: Supplementary results

Correlations across variables

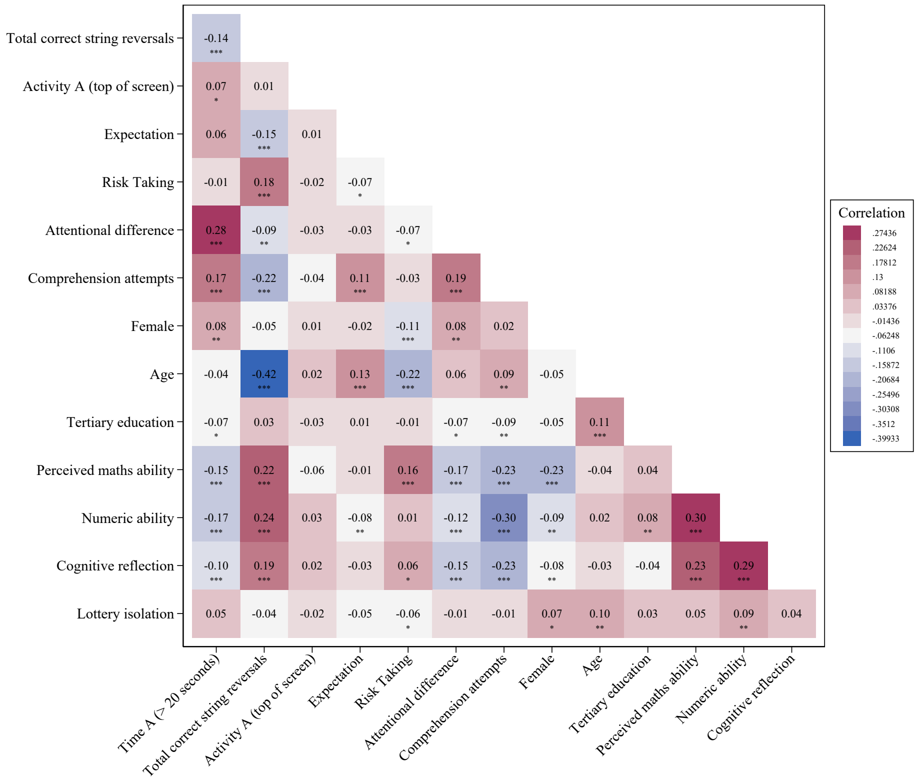


Figure B8: Correlations across variables

* p < 0.1, ** p < 0.05, *** p < 0.01.
Notes. Spearman's rank correlation coefficients (ρ) for Studies 1-3 pooled and their associated signifiance levels. The data has been restricted to those who (i) were within or above C^* for $\alpha = \{9, 7, 5\}$ and (ii) understood how to generate at-least one set for all treatments. The blue and red shaded boxes represent negative and positive correlations, respectively with intensity of shaded region representing relative strength of correlation (as indicated by legend).

Order of activities on screen

The order of activities was randomised on the choose-activity screen.[6] That is, activity A was either displayed at the top of the screen with activity B at the bottom or vice versa. Read et al. (1999b) discuss the importance of *cognitive inertia*, where people might respond to elements in their environment in the order that they are presented. This indicates that the order of the activity on screen might have an impact in how people choose to narrow bracket. Indeed, we saw a strong effect of ordering on dominated choice in Chapter 1.
To analyse ordering effects, we timestamped where a subject is at any given moment throughout their time spent working on the tasks. Utilising this information, we construct an overall time allocation profile for each subject. Figure B9 shows how the order of activity affects the allocation of time across all the experiment pooled together. The *y-axis* represents the proportion of subjects in activity A at any given second and the *x-axis* is the time spent working from 0 to 360 seconds.[7]

[6]This was not the case in our first two experiments, where we wanted to test the existence of narrow bracketing
[7]Note that subject are free to choose as much time as they would like to in the choose-activity screen. While we do record this information, we exclude it for the purposes of the current analysis, as our key interest is how subjects allocate their 360 seconds of available working time.

Thus, the graphs show how the proportion of subjects in activity A changes over the task duration for all experiments, on average. The top figure is when activity A is displayed at the top of the screen, and the bottom figure is when activity B is displayed at the top. See Figure B10 for results disaggregated over the seven experiments.

The figure reveals two insights. The first is that the time profile for how subjects spend their time is vastly different when activity A is displayed at the top rather than at the bottom ($p < 0.01$). More specifically, subjects appear to be working in the order that options are presented to them. When activity A is at the top of the screen, the majority of subjects start working on the components of activity of A. However, when activity B is at top of the page, the opposite is true. The second is that although the time profiles for whether activity A is at the top of the screen or at the bottom are different, the overall time spent on activity A is not different across the two. To elaborate, the total 'area' under the two lines are not statistically different to one-another. This is reflected in Table B7 which shows OLS estimates for several regressions at different 'blocks'

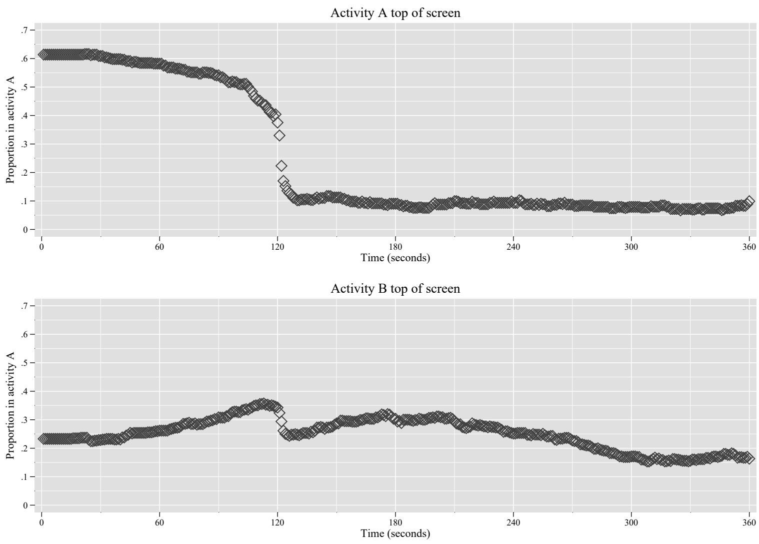


Figure B9: Activity ordering effects for all sessions

	Time A in blocks						Time A overall
	(1) 1-60	(2) 61-120	(3) 121-180	(4) 181-240	(5) 241-300	(6) 301-360	(7) 1-360
Activty A (top of screen)	0.40***	0.22***	-0.20***	-0.22***	-0.14***	-0.09***	-0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	626	626	626	626	626	626	626
R^2	0.24	0.15	0.18	0.20	0.14	0.15	0.12

Table B7: Ordinary least squares: activity ordering effects for all sessions

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.

Notes. Dependent variable: Proportion of subjects in activity A (A1+A2) across blocks of 60-second intervals. Independent measures include: a binary measure of whether activity A was positioned on the top of the screen; and treatments with the $\alpha = 9$ as the base. Additional controls include age, gender, tertiary education, subject performance in terms of sets generated and comprehension. The sample is restricted to those who generated more than 0 sets and had a speed classification within or above $C- *$ since it is optimal for these subjects to spend no time on activity A at all. The data for Study 1 has been excluded from this analysis since we did not randomise the order of activities on screen.

of time spent working. More specifically, we regress the order of activity A on screen with the proportion of subjects working in activity A in 60-second intervals up to 360 seconds. From 1-120 seconds, we see that there is a significantly higher proportion of subjects in activity A when it is first on screen (columns 1 and 2). This then reverses thereafter (columns 3-6), making the overall effect null (column 7).

Taken together, this data highlights an interesting behavioural phenomenon in that subjects are, to a large extent, responding to the environment in the order in which the options are presented to them. More specifically, the order of options on the screen affects the sequence in which people choose to work. Since the absolute measure of time spent on activity A is what we use in our analysis, this is not something that we need to worry about when comparing treatment differences. This behaviour might have important implications in certain classes of environments. For instance, imagine a world where the clock is ticking, and one is working on one of any two tasks (say activity A or B). At any given moment in time, if their ability to work on activity A or B might get stopped, meaning that they cannot come back to it, then the order that the activity is presented has a dramatic impact on their time allocation between the two. In our current set-up, it does not have such an impact because both activities A and B are available for the full 360 seconds and subjects are aware of this. However, in life, when the availabilities of tasks are much more haphazard with fleeting opportunities to work on some tasks, this phenomenon may have powerful implications for the overall outcomes of certain time-allocation decision environments.

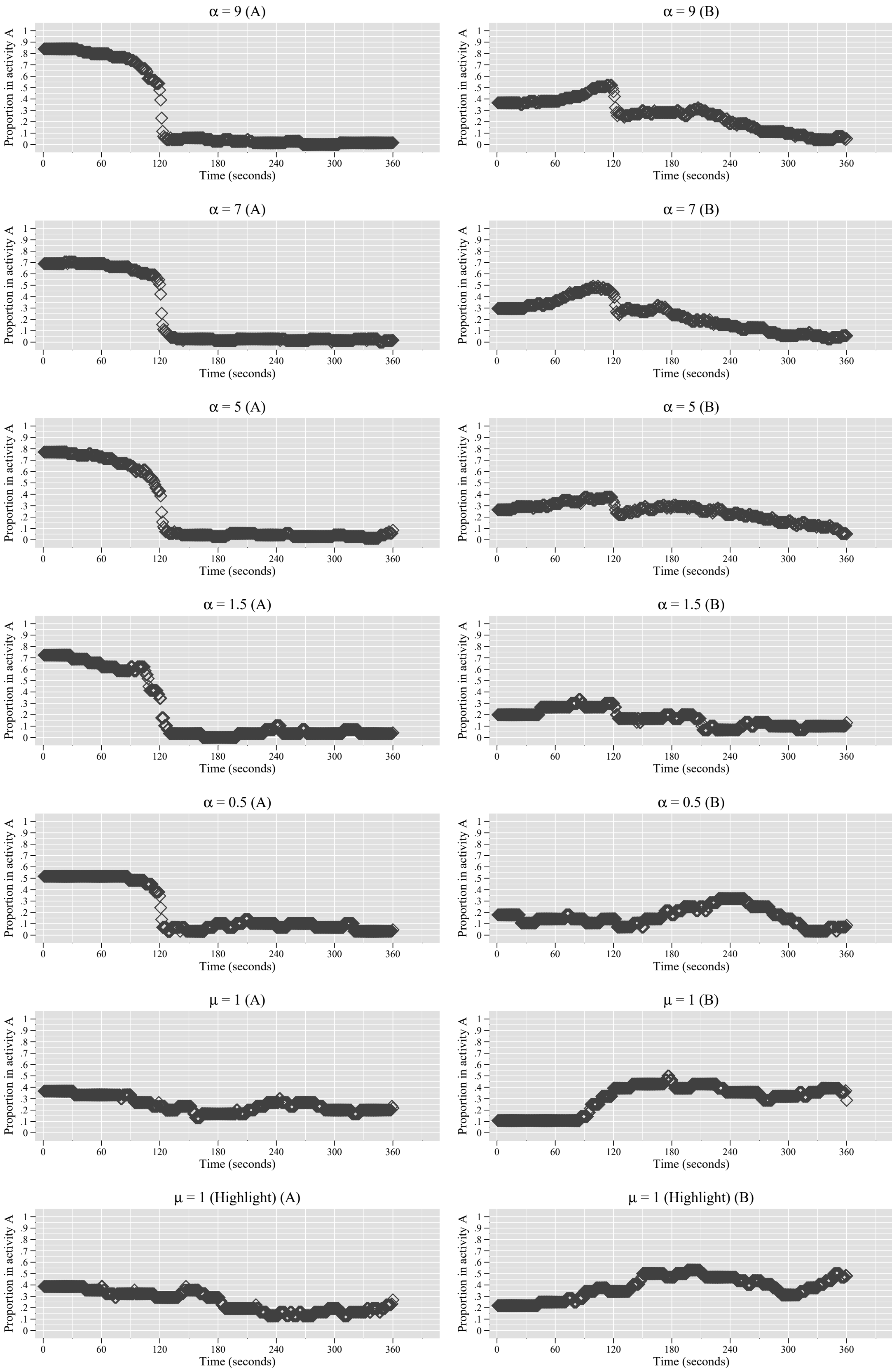


Figure B10: Activity ordering effects for all experiments (disagregated)

Expectations

If subjects systematically underestimate their performance they think they can do, then it is conceivable that they are 'rationally' spending time on activity A (given their beliefs). In other words, one might expect subjects who spend time on activity A ($Time_A > 20$ seconds) do so because they anticipate being able to construct fewer sets than those who spend no time on activity A ($Time_A \leq 20$ seconds), on average. Figure B11 provides subjects perceptions about whether they correctly reversed more or less strings in the task than they originally expected. The results show no difference in expectations across types, suggesting that expectations do not drive the behaviour we observe. Further strengthening this point, we also find no differences in risk preferences in the expected direction (Falk et al., 2018). If activity B was seen as more 'risky', we should see more risk-averse subjects choosing to work on activity A.

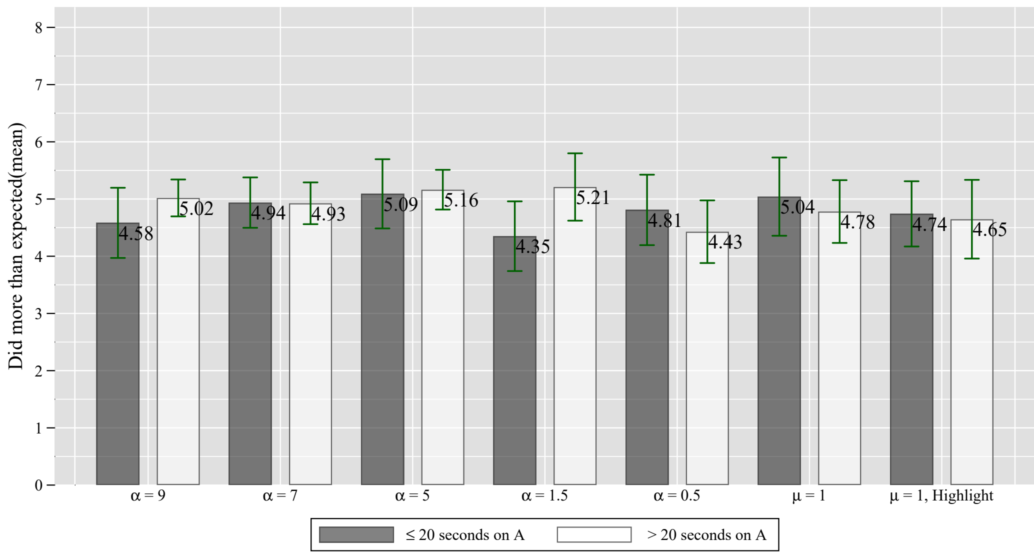


Figure B11: Expectations across sessions

Comprehension

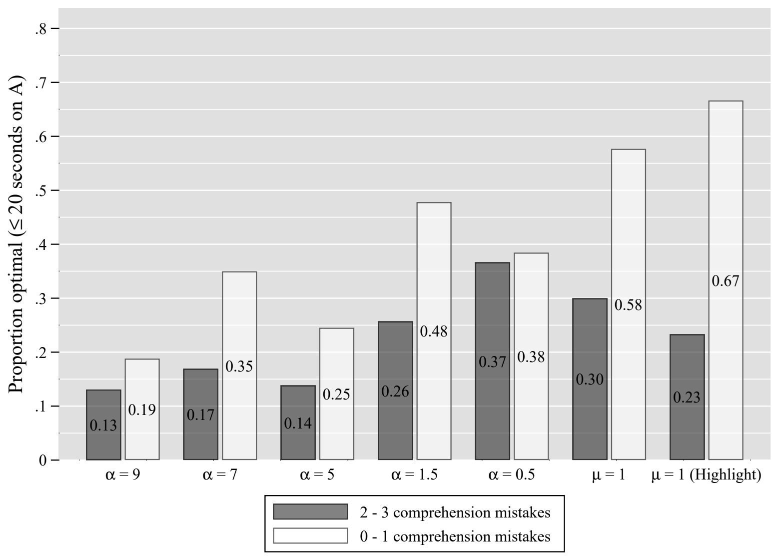


Figure B12: Optimal behaviour across all sessions and comprehension (disagregated)

	Full sample			High comprehension	
	(1) ME	(2) ME	(3) ME	(4) ME	(5) ME
TREATMENTS					
Base. $\alpha = \{1.5, 0.5\}$					
$\alpha = \{9, 7, 5\}$	0.16***	0.17***	0.21***	0.17**	0.23***
	(0.04)	(0.04)	(0.04)	(0.07)	(0.07)
$\mu = 1$	-0.07	-0.06	-0.08	-0.18**	-0.23**
	(0.05)	(0.05)	(0.05)	(0.09)	(0.09)
COMPREHENSION OF INSTRUCTIONS					
Comprehension (0-1 mistakes)		-0.16***	-0.14***		
		(0.03)	(0.03)		
Additional controls	No	No	Yes	No	Yes
# observations	702	702	694	336	331
Psuedo R^2	0.04	0.07	0.12	0.07	0.13
Log-lik	-388.26	-376.87	-351.47	-201.76	-185.14
χ^2	33.09	52.69	85.30	27.47	48.52

Table B8: Marginal effects: comprehension

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: A binary measure taking on a value of 1 if $Time_A > 20$ seconds, and 0 otherwise. We construct this binary measure (instead of the continuous measure of time on activity A) since in the $\mu = 1$ sessions, time on activity A ranges from 0 to 360 seconds, which is not comparable to the other sessions. Thus, the proportion of optimal choice within the sample is a much more accurate metric to measure optimal choice through. Independent measures include the studies grouped into three treatments: $\alpha = \{9, 7, 5\}$, $\alpha = \{1.5, 0.5\}$ (the base), and $\mu = 1$; and comprehension as a binary measure (0-1 mistakes). Additional controls include age, gender, tertiary education, subject performance in terms of total strings correctly reversed, activity order, and expectations about performance. The 'Full sample' is restricted to those who generated more than zero sets and had a speed classification within or above C*. On the other hand, the 'High comprehension' refers to those in the full sample with comprehension attempts ranging from 0-1 (instead of 2-3).

Attentional focus

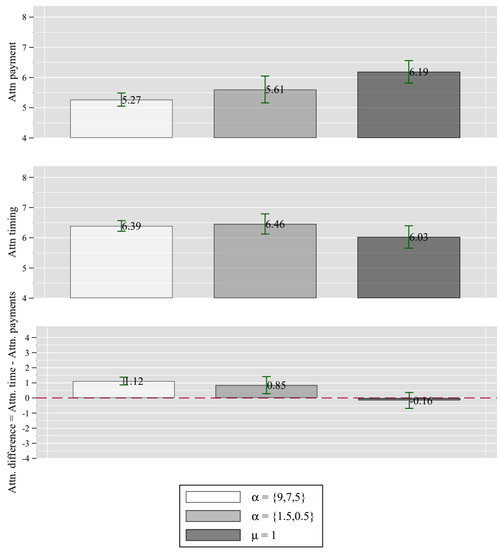


Figure B13: Attentional focus and perceived urgency (aggregated)

Notes. Whiskers represent the 95% confidence intervals.

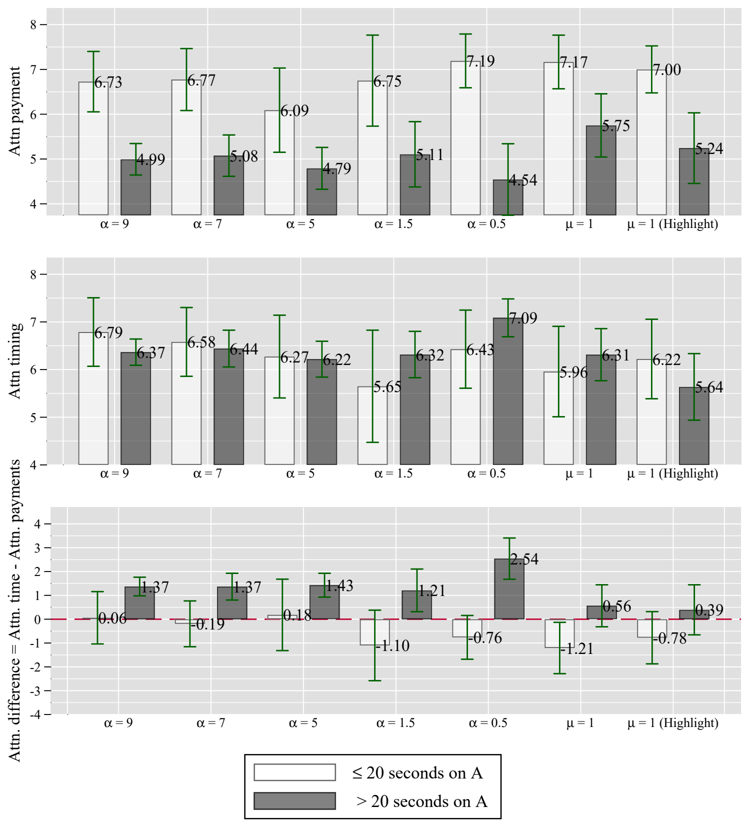


Figure B14: Attentional focus and perceived urgency (disaggregated across all treatments)

Notes. Whiskers represent the 95% confidence intervals.

	Attn. payment	Attn. timing	Attn. difference	$Time_A > 20$ seconds		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
TREATMENTS						
Base. $\alpha = \{1.5, 0.5\}$						
$\alpha = 9, 7, 5$	-0.39	-0.03	0.36			
	(0.24)	(0.20)	(0.31)			
$\mu = 1$	0.55*	-0.44*	-0.99**			
	(0.28)	(0.26)	(0.39)			
ATTENTION						
Attention to Payment				-0.06***		
				(0.01)		
Attention to Time					0.01	
					(0.01)	
Attentional difference						0.04***
						(0.01)
Constant	5.79***	5.96***	7.17***	1.29***	0.92***	0.69***
	(0.67)	(0.55)	(0.82)	(0.12)	(0.13)	(0.12)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
N observations	694	694	694	694	694	694
R-squared	0.12	0.02	0.08	0.14	0.06	0.12

Table B9: Ordinary least squares: attentional focus

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Notes. Dependent variables: Column 1 —Attention to payment (1 to 8); Column 2 —Attention to timing (1 to 8); Column 3 —Attentional difference (−7 to +7); and Columns 4, 5 and 6 —the binary measure of time spent on activity A ($Time_A > 20$ seconds). Independent measures include the studies grouped into (i) three treatments (Columns 1, 2 and 3): $\alpha = \{9, 7, 5\}$, $\alpha = \{1.5, 0.5\}$ (the base); and $\mu = 1$; (ii) attentional focus metrics (Columns 4,5 and 6); attention to payment (1 to 8); attention to timing (1 to 8); and attentional difference (−7 to +7) Additional controls include age, gender, tertiary education, subject performance in terms of total strings correctly reversed, activity order, and expectations about performance.

	More attention on time		Perceived urgency of A		Perceived busyness	
	(1) 1-120	(2) 1-360	(3) 1-120	(4) 1-360	(5) 1-120	(6) 1-360
Attentional difference	0.00 (0.01)	0.01*** (0.00)				
Perceived urgency			0.01 (0.01)	0.03*** (0.00)		
Perceived busyness					0.03* (0.02)	0.02*** (0.00)
Constant	0.60*** (0.13)	0.44*** (0.05)	0.54*** (0.14)	0.29*** (0.05)	0.41** (0.17)	0.30*** (0.05)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
N observations	455	632	455	632	455	632
R-squared	0.01	0.11	0.02	0.19	0.02	0.12

Table B10: Ordinary least squares: urgency and attention

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variables: Columns 1, 3 and 5; the proportion of subjects in activity A from 0-120 seconds: Columns 2, 4 and 6; the proportion of subjects in activity A from 0-360 seconds. The independent measures in Columns 1 and 2; Attentional difference (-7 to +7). Columns 3 and 4; Perceived urgency of activity A (1 to 8). Columns 5 and 6; Perceived urgency of activity A (1 to 8). Additional controls include ordering of activities on screen, age, gender, tertiary education, treatments and subject performance in terms of sets generated. The sample has been restricted to only those that have spent > 20 second in activity A in columns 1, 3 and 5 and includes the full sample in columns 2, 4 and 6. All samples are for subjects within C*.

Demographics and cognition

	Full sample				
	(1) maths	(2) NA	(3) CRT	(4) LI	(5) RT
DEMOGRAPHICS					
Female	-1.29*** (0.19)	-0.13** (0.05)	-0.12* (0.07)	0.08* (0.04)	-0.48** (0.23)
Age	-0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.00** (0.00)	-0.05*** (0.01)
Tertiary education	0.11 (0.19)	0.11** (0.05)	-0.08 (0.06)	0.03 (0.04)	-0.06 (0.22)
Constant	6.92*** (0.40)	1.36*** (0.11)	1.92*** (0.13)	0.52*** (0.07)	11.10*** (0.44)
Additional controls	Yes	Yes	Yes	Yes	Yes
N observations	695	695	695	694	695
R-squared	0.06	0.03	0.01	0.01	0.08

Table B11: Ordinary least squares: demographics and cognition

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: Perceived maths ability (0 to 10) (Column 1); Numeric ability (0 to 2) (Column 2); Cognitive reflection test (0 to 3) (Column 3); Lottery isolation (binary) (Column 4); Risk taking (standardised) (Column 5). Additional controls include the treatment groups pooled together.

Other factors

Figure B15 shows that most subjects do not believe their behaviour is optimal. This result holds for even those whose behaviour is optimal (≤ 20 seconds on activity A). That said, a large part of this is driven by the fact that many answer 'not sure' to the question, indicating substantive uncertainty.

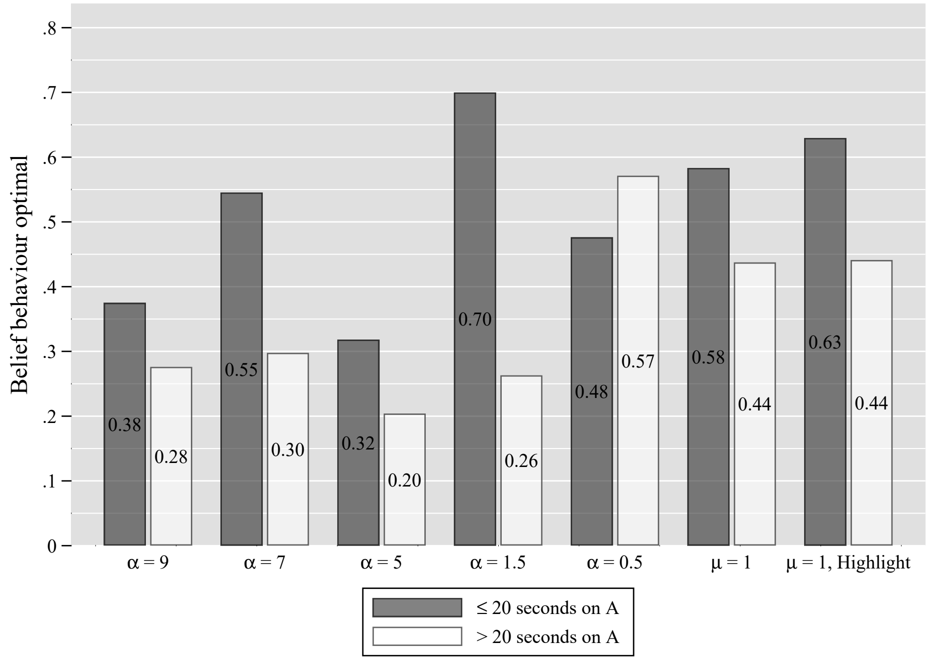
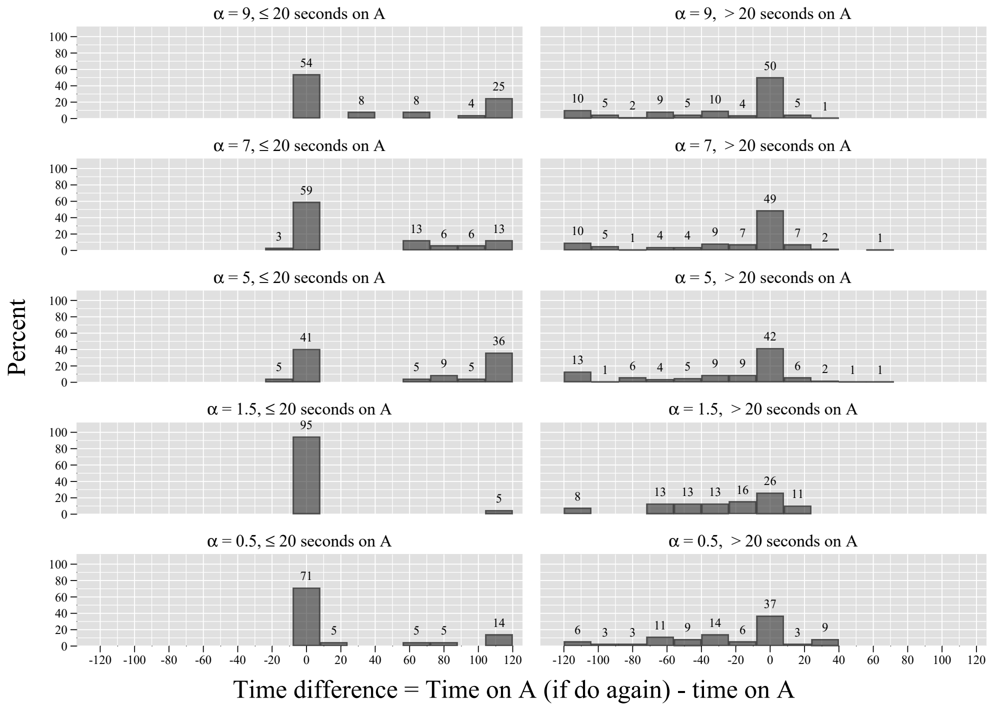
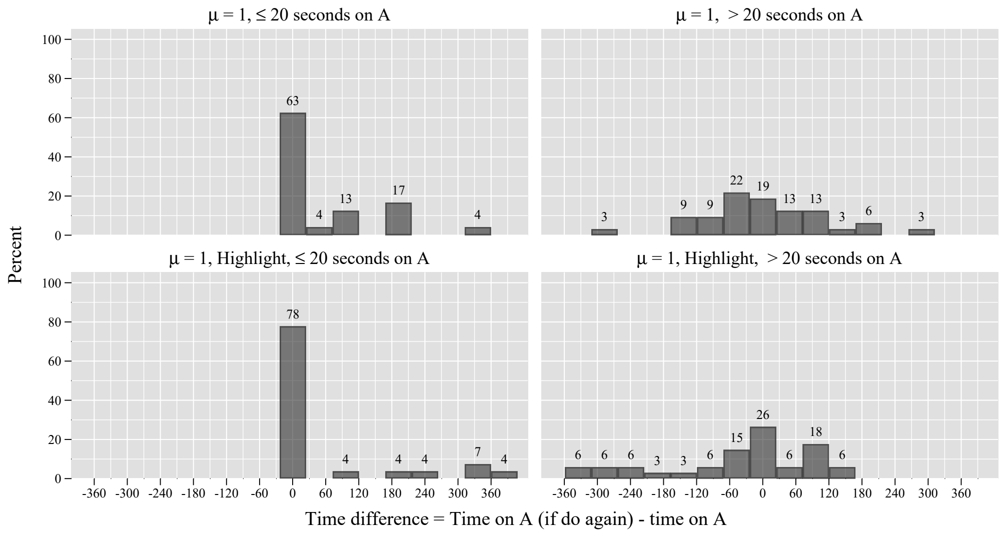


Figure B15: Belief about optimal behaviour across sessions

We also ask subjects what they would do if they did the task again. This information is shown in Figure B16. Across all sessions, while we find that those who were suboptimal (spending $Time_A > 20$ seconds) would spend more time on activity B, there is a very small proportion fully swinging the other way (stating that they would spend ≤ 20 seconds on activity A). For those that are optimal, a substantial proportion states that they will continue to spend no time on activity A. Overall, this metric indicates that behaviour is quite sticky, with the majority of subjects stating that they would have the same resource allocation in a second iteration of the task.



(a) $\alpha = \{9, 7, 5, 1.5, 0.5\}$



(b) $\mu = 1$

Figure B16: Time if do task again across sessions

APPENDIX C

Chapter III

C.1 New instructions

[The experiment was programmed on LIONESS LAB (Giamattei et al., 2020) (all material available on upon request). Comments in footnotes and square brackets were not seen by readers. We highlight text in red to help facilitate comparison across treatment differences. Note that instructions in the rest of the experiment (i.e., not the main instructions) were identical to our work in Chapter 2. Thus, we only show the core changes across instructions.]

C.1.1 Baseline condition

Stage 1: Instructions and comprehension questions

Please read the instructions carefully.

In this study, you must choose how to allocate your time across two activities: Activity F and Activity V. These activities involve the same underlying string-reversal task but differ in their payment and in their time limits - more details on this below.

The task: You will be given sequences of randomly generated 6-letter strings and are required to type as many of them as you can in reverse order. As an example: If you see 'rlgows' for one string, the correct answer will be 'swoglr'.

For each activity, there are two different coloured strings. In each activity two strings of each colour form a set.
In Activity F, the colours are $colour1$ and $colour2$:

2 $colour1$ strings + 2 $colour2$ strings = 1 set in Activity F

In Activity V, the colours are $colour3$ and $colour4$:

2 $colour3$ strings + 2 $colour4$ strings = 1 set in Activity V

You will be paid for the number of correctly solved sets in each activity. As an example, imagine that you correctly solve 6 $colour1$ and 8 $colour2$ strings in Activity F, then you will be paid for 3 sets in Activity F. Similarly, if you correctly solve 6 $colour3$ and 8 $colour4$ strings in Activity V, then you will be paid for 3 sets in Activity V.
Note that different coloured strings from different activities do not form sets. For example:

2 $colour1$ strings (from Activity F) + 2 $colour3$ strings (from Activity V) = 0 sets

Please CLICK ON THE BUTTON below which shows an example of how your Choose Activity screen will look like. You will notice that the buttons to reverse the coloured strings have been disabled in this example. However, in the actual task you will be able to click on them and reverse the coloured strings of your choosing.

[Link to task environment]

Below we explain how the two activities differ with respect to payment and time limits.

1) Payment:

In the task, you earn Points for completing sets. At the end of the study, your Points will be converted into a Prolific bonus (£) at a pre-defined exchange rate (100 Points = £1.00).
In Activity F the Points you earn for each set are fixed. In Activity V the Points you earn for each set are variable. More specifically, each set you complete in Activity F is worth the same number of Points, but how many Points a set in Activity V is worth rises with the number of sets you complete in that activity. For each activity, the value of each set completed (in Points) is given below.

Activity A

Every correctly solved set is worth:	7 Points

Activity B

Your first correctly solved set is worth:	1 Points
Your second correctly solved set is worth:	3 Points
Your third correctly solved set is worth:	5 Points
Your fourth correctly solved set is worth:	7 Points
Your fifth correctly solved set is worth:	9 Points
Your sixth correctly solved set is worth:	11 Points

...and so on, with each successive set of correctly solved strings in Activity V worth 2 Points more than the previous one.

2) Time limits:

In the task, you have a total of 6 minutes. You can freely choose how to allocate your time across the two activities. However, the availability and time that you can spend working on each of them differs.

- Activity F: This activity is available for the full 6 minutes, but you can only work on it for up to 2 minutes in total.
- Activity V: This activity is available for the full 6 minutes, and you can work on it for up to 6 minutes in total.

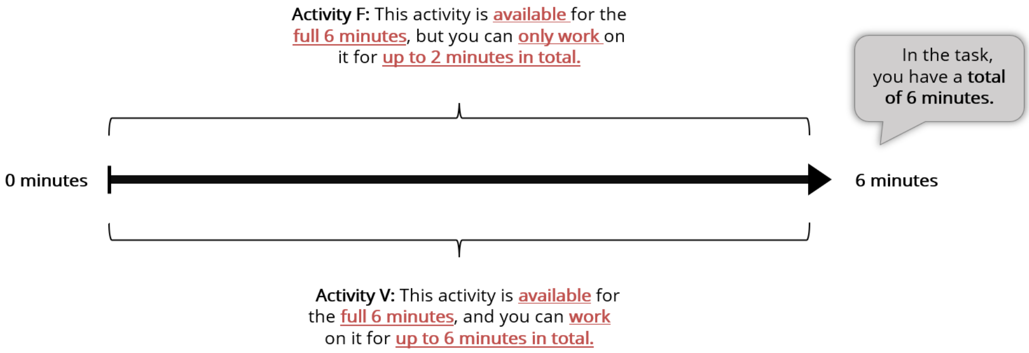
There will be two timers:

1. A 2-minute timer will run down only while you are working on Activity F. Once the 2 minutes on this timer have passed, you cannot work anymore on Activity F.
2. A 6-minute timer will run down while you are working on either Activity F or Activity V. Once the 6 minutes on this timer have passed, you cannot work anymore on either activity.

The timers will not run down when you are on the Choose Activity screen.

Note: You have 6 minutes of working time in total. The least time you can spend in Activity F is 0 minutes, so the most you can spend in Activity V is 6 minutes. The most time you can spend in Activity F is 2 minutes, so the least you can spend in Activity V is 4 minutes. You can switch freely between the activities whenever you like, subject to the time limits.

Please LOOK AT THE FIGURE BELOW to make sure you understand the time limits on both activities.



IMPORTANT NOTE: You do not have to work on both activities. How you choose to allocate your time will affect your payment. **Pay attention to the payment rules and time limits of the activities above.**

Comprehension questions

In order to move on please answer the following questions:

1. The underlying string-reversal task for producing sets is the same in Activity F as in Activity V.

   (Yes)[1]
   (No)

2. If you correctly reverse 8 $colour1$ and 4 $colour2$ strings in Activity F, you will be paid for [...] in Activity F.

   (3 sets)
   (2 sets)
   (8 sets)

3. If you correctly reverse 8 $colour3$ and 4 $colour4$ strings in Activity V, you will be paid for [...] in Activity V.

   (3 sets)
   (2 sets)
   (8 sets)

4. If you correctly reverse 8 $colour2$ and 4 $colour4$ strings in Activity V, you will be paid for [...].

   (0 sets)
   (2 sets)
   (8 sets)

[1]Correct responses are in green.

5. Imagine that you had already correctly solved six sets in Activity F. How much would it add to your earnings (in Points) if you solved a seventh set for that activity?

{7} Points
(49 Points)
(13 Points)

6. Imagine that you solve exactly seven sets in Activity F, how much would you earn in total from this activity?

(63 Points)
{49} Points
(7 Points)

7. Imagine that you had already correctly solved six sets in Activity V. How much would it add to your earnings (in Points) if you solved a seventh set for that activity?

(7 Points)
(49 Points)
(13 Points)

8. Imagine that you solve exactly seven sets in Activity V, how much would you earn in total from this activity?

(49 Points)
(63 Points)
(13 Points)

9. In this experiment, you can work on Activity F for up to 2 minutes and Activity V for up to the full 6 minutes. In total, you have:

(6 minutes)
(8 minutes)
(None of the above)

10. Will the 2-minute timer run down when you are working on Activity V?

(Yes)
(No)

C.1.2 Urgency condition

Stage 1: Instructions and comprehension questions

Please read the instructions carefully.

[Everything is the same, except the instructions regarding the time-limits.]

2) Time limits:

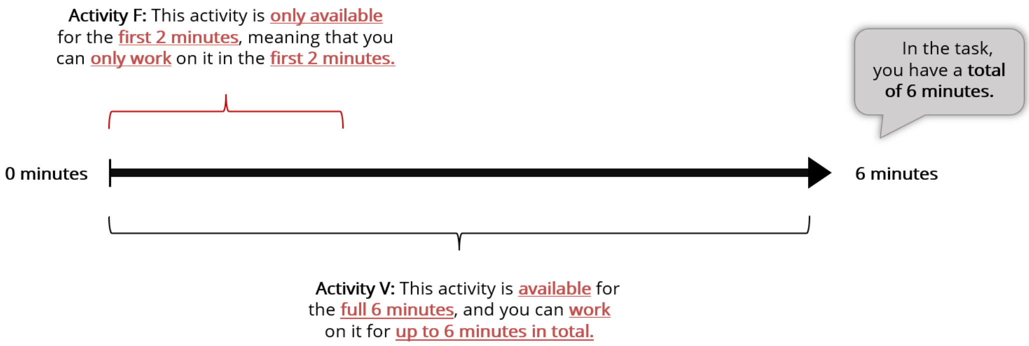
In the task, you have a total of 6 minutes. You can freely choose how to allocate your time across the two activities. However, the availability and time that you can spend working on each of them differs.

- Activity F: This activity is only available for the first 2 minutes, meaning that you can only work on it in the first 2 minutes.
- Activity V: This activity is available for the full 6 minutes, and you can work on it for up to 6 minutes in total.

There will be two timers:

1. A 2-minute timer will run down for the first 2 minutes while you are working on either Activity F or Activity V. Once the 2 minutes on this timer have passed, you cannot work anymore on Activity F.
2. A 6-minute timer will run down while you are working on either Activity F or Activity V. Once the 6 minutes on this timer have passed, you cannot work anymore on either activity.

The timers will not run down when you are on the Choose Activity screen.
Note: You have 6 minutes of working time in total. The least time you can spend in Activity F is 0 minutes, so the most you can spend in Activity V is 6 minutes. The most time you can spend in Activity F is 2 minutes, so the least you can spend in Activity V is 4 minutes. You can switch freely between the activities whenever you like, subject to the time limits.
Please LOOK AT THE FIGURE BELOW to make sure you understand the time limits on both activities.



IMPORTANT NOTE: You do not have to work on both activities. How you choose to allocate your time will affect your payment. **Pay attention to the payment rules and time limits of the activities above.**

Comprehension questions

In order to move on please answer the following questions:

[All comprehension questions are the same, except the answer to question 10 above is now 'Yes'.]

C.1.3 Busyness condition

Stage 1: Instructions and comprehension questions

Please read the instructions carefully.

In this study, you must choose how to allocate your time across two activities: Activity F and Activity V. These activities involve the same underlying string-reversal task but differ in their payment and in their time limits - more details on this below.

The task: You will be given sequences of randomly generated 6-letter strings and are required to type as many of them as you can in reverse order. As an example: If you see 'rlgows' for one string, the correct answer will be 'swoglr'.

For each activity, there are four different coloured strings. In each activity, one string of each colour forms a set.

In Activity F, the colours are $colour1$, $colour2$, $colour3$, and $colour4$:

1 $colour1$ string + 1 $colour2$ string + 1 $colour3$ string + 1 $colour4$ string = 1 set in Activity F

In Activity V, the colours are $colour5$, $colour6$, $colour7$, and $colour8$:

1 $colour5$ string + 1 $colour6$ string + 1 $colour7$ string + 1 $colour8$ string = 1 set in Activity V

You will be paid for the number of correctly solved sets in each activity. As an example, imagine that you correctly solve 3 $colour1$, 4 $colour2$, 3 $colour3$, and 4 $colour4$ strings in Activity F, then you will be paid for 3 sets in Activity F. Similarly, if you correctly solve 3 $colour5$, 4 $colour6$, 3 $colour7$, and 4 $colour8$ strings in Activity V, then you will be paid for 3 sets in Activity V. par
Note that different coloured strings from different activities do not form sets. For example:

1 $colour1$ string + 1 $colour3$ string (both from Activity F) + 1 $colour5$ string + 1 $colour7$ string (both from Activity V) = 0 sets

Please CLICK ON THE BUTTON below which shows an example of how your Choose Activity screen will look like. You will notice that the buttons to reverse the coloured strings have been disabled in this example. However, in the actual task you will be able to click on them and reverse the coloured strings of your choosing.

[Link to task environment]

Below we explain how the two activities differ with respect to payment and time limits.

1) Payment:

In the task, you earn Points for completing sets. At the end of the study, your Points will be converted into a Prolific bonus (£) at a pre-defined exchange rate (100 Points = £1.00).
In Activity F the Points you earn for each set are fixed. In Activity V the Points you earn for each set are variable. More specifically, each set you complete in Activity F is worth the same number of Points, but how many Points a set in Activity V is worth rises with the number of sets you complete in that activity. For each activity, the value of each set completed (in Points) is given below.

Activity A

Every correctly solved set is worth:	7 Points

Activity B

Your first correctly solved set is worth:	1 Points
Your second correctly solved set is worth:	3 Points
Your third correctly solved set is worth:	5 Points
Your fourth correctly solved set is worth:	7 Points
Your fifth correctly solved set is worth:	9 Points
Your sixth correctly solved set is worth:	11 Points

...and so on, with each successive set of correctly solved strings in Activity V worth 2 Points more than the previous one.

2) Time limits:

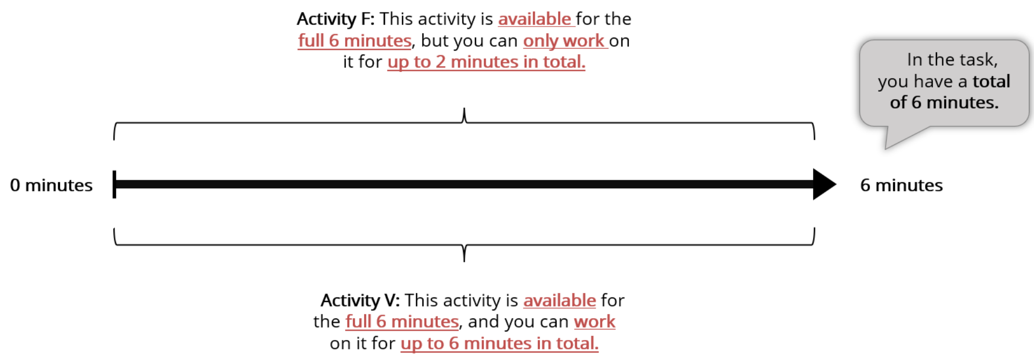
In the task, you have a total of 6 minutes. You can freely choose how to allocate your time across the two activities. However, the availability and time that you can spend working on each of them differs.
Activity F: This activity is available for the full 6 minutes, but you can only work on it for up to 2 minutes in total.

- Activity F: This activity is available for the full 6 minutes, but you can only work on it for up to 2 minutes in total.
- Activity V: This activity is available for the full 6 minutes, and you can work on it for up to 6 minutes in total.

There will be two timers:

1. A 2-minute timer will run down only while you are working on Activity F. Once the 2 minutes on this timer have passed, you cannot work anymore on Activity F.
2. A 6-minute timer will run down while you are working on either Activity F or Activity V. Once the 6 minutes on this timer have passed, you cannot work anymore on either activity.

The timers will not run down when you are on the Choose Activity screen.
Note: You have 6 minutes of working time in total. The least time you can spend in Activity F is 0 minutes, so the most you can spend in Activity V is 6 minutes. The most time you can spend in Activity F is 2 minutes, so the least you can spend in Activity V is 4 minutes. You can switch freely between the activities whenever you like, subject to the time limits.
Please LOOK AT THE FIGURE BELOW to make sure you understand the time limits on both activities.



IMPORTANT NOTE: **You do not have to work on both activities.** How you choose to allocate your time will affect your payment. **Pay attention to the payment rules and time limits of the activities above.**

Comprehension questions

In order to move on please answer the following questions:

All comprehension questions are the same, except for questions 2, 3 and 4 regarding set generation.

2. If you correctly reverse 4 $colour1$, 2 $colour2$, 4 $colour3$ and 2 $colour4$ strings in Activity F, you will be paid for [...] in Activity F.

   (3 sets)
   (2 sets)
   (8 sets)

3. If you correctly reverse 4 $colour5$, 2 $colour6$, 4 $colour7$, and 2 *colour*8 strings in Activity V, you will be paid for [...] in Activity V.

   (3 sets)
   (2 sets)
   (8 sets)

4. If you correctly reverse 4 $colour2$ and 4 $colour4$ strings in Activity F, and 2 $colour6$ and 2 $colour8$ strings Activity V, you will be paid for [...].

   (0 sets)
   (2 sets)
   (8 sets)

C.2 Power calculations

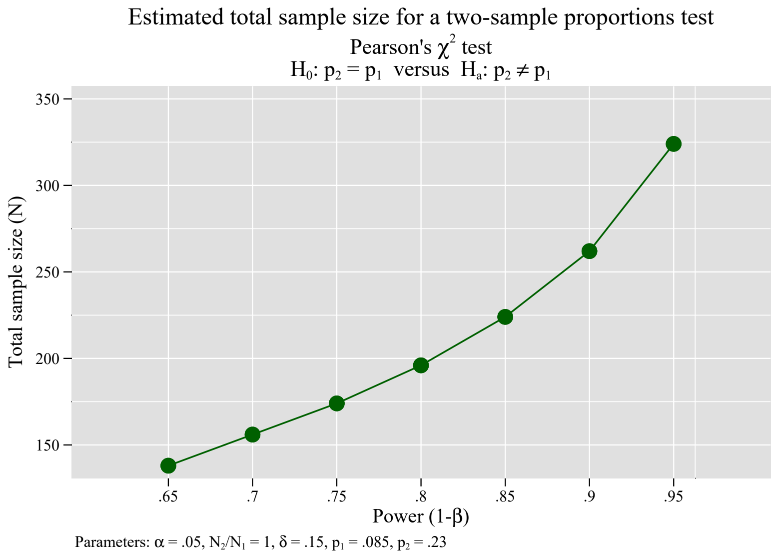
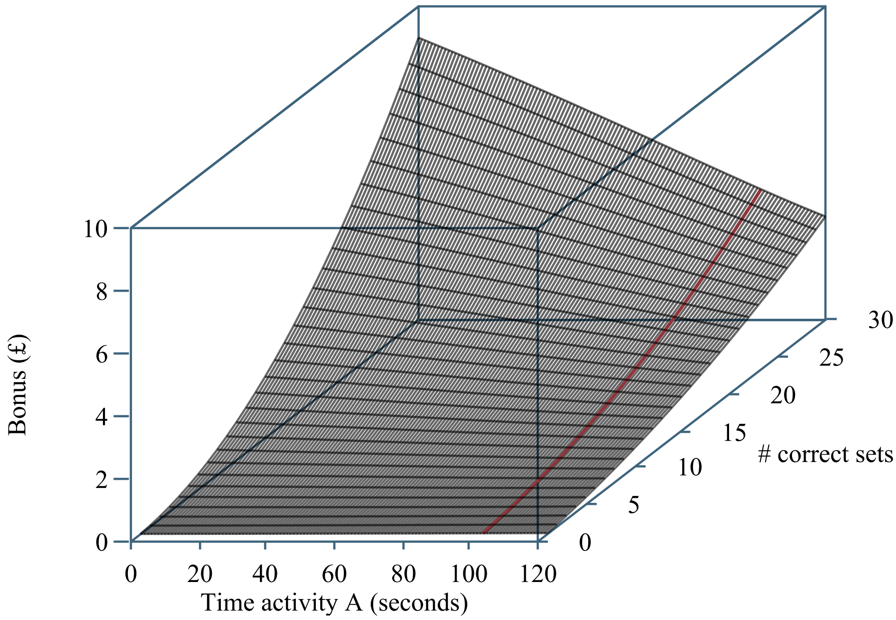


Figure C1: Power calculations

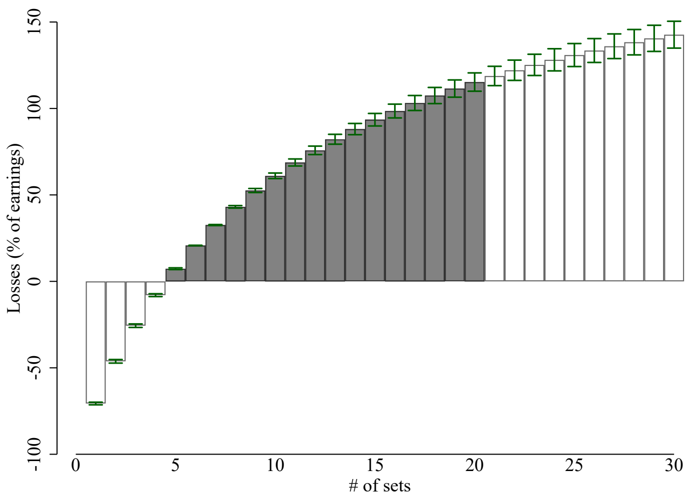
Notes. This power calculation is based off our mere urgency effect replication work (see Chapter 1) and shows what power we would obtain under various total sample sizes (N). The test is for a two-sample χ^2 test with the dominated choice proportions we obtained in Replication III —we take the average dominated choice share between 'Combined-Payoffs' and 'Combined-Availability' (see Figure 1.3). With 90% power, we would need a sample size of just over $N = 250$. However, this is 2x2 design. As a result, we have simply doubled the sample size required.

C.3 Losses associated with time spent on activity A

To give a flavour for the differences in earning across time spent on activity A (i.e., bracketing behaviour), while controlling for varying performance abilities of the subject that we now know from our previous work, Figure C2a shows a 3-dimensional graphic with regards to time spent on activity A, sets produced and the respective bonus (on the x-, y- and z-axis, respectively). Akin to a contorted sheet of paper, the figure shows how these three elements interact with one-another. One can see that earnings are quadratically increasing with the number of sets produced, with especially sharp curvature when no time is spent on activity A (implying that all time is spent on activity B). Figure C2b makes losses more tangible by showing the losses as a percentage of total earnings across the different sets generated for agents spending ≥ 100 seconds on activity A. The dark-grey shaded bars represented individuals within C^* and the white-shaded bars those outside. Note that for individuals within C^*, classified as narrow bracketers, this figure shows that losses are indeed substantial and are rising quadratically as the number of sets produced increases. For individuals below C^*, losses are negative for low levels of performance because subjects have an interest in spending maximal amount of time on activity A.



(a) Three-dimensional graphics



(b) Losses from time spent on activity A bracketing

Figure C2: Trade-offs between time spent, performance, and payoffs

Notes. The data has been generated, given the distribution of performance from Chapter 2. Specifically, we observed a range between 0 to 30 sets produced across the three studies. With this in mind, a data point has been generated for each interaction across three metrics: (i) Time spent on activity A (x-axis), (ii) sets produced (y-axis) and (iii) the respective bonus (z-axis). In Figure (a), a surface plot of this data has been drawn. The red line at $Time_A = 100$ seconds represents the 100-second threshold at which we empirically classify narrow bracketing behaviour. For those within C*, spending more than 100 seconds on activity A is classified as narrow bracketing. Thus, this line refers to the restricted sample in Figure (b), which shows the average losses from time spent on activity A across varying levels of set generation. The white bars represent those outside C*, while the grey bars represent those within C*. Note that the negative losses under low levels of set generation are because these individuals should spend time on activity A (in terms of income maximisation). Moreover, losses can go above 100% since these are losses as a proportion of earnings (and losses can be larger than the earnings themselves).

C.4 Improved comprehension relative to Chapter 2

There were several departures from the studies in Chapter 3 relative to Chapter 2 in terms of the instructions and the display of the decision-screen. This was because one of our aims was to improve subject comprehension, something that we saw was significantly positively correlated with narrow bracketing behaviour.

In Chapter 3, one of the key changes already outlined in the main text is to do with the colours of the strings. Specifically string colours were always unique for each string, in the sense that *each* component has a colour that is unique to any other component in activity A and B. This contrasts to Chapter 2, where there were only two green and two blue strings. The rationale for this change was to create a clear separation between the components that generate sets in the two activities, and as a result, to reduce possible confusion.

We also added a comprehension question to further distil this distinction. Another key change is that we highlighted the difference between (i) *availability*, and (ii) *time that can be spent working* on the activities. In all the conditions, activity A can be worked on for up to 2 minutes in total. However, its availability differs across treatments. More precisely, activity A is only available in the first 2 minutes in the localised condition, yet it is available for up to 6 minutes in the delocalised condition, with the constraint that not more than 2 minutes can be spent on it. To make this feature as salient as possible, we added infographics that differed across treatments, in addition to the changes we made in the wording of the instructions.[2]

On the decision screen itself, we provide subjects with the number of 'Points' they generate in real time. We felt that this would allow them to see what sets, and therefore, what Points they were accumulating for each activity. This is to minimise any mistakes of thinking they might be in one activity, but are in fact, in the other. We also added a link that allows subjects to go back to the instructions in case they felt like they needed to refresh their knowledge of the set-up. Lastly, we chose not to randomise the order of the activities on screen in this experiment. This was for the important reason that in our previous studies, we saw that subjects followed a pattern of working in the order that options were presented to them on screen (i.e., top to bottom). While this did not interact with the treatment manipulations in Chapter 2, it would for Chapter 3. Specifically, because we vary the time at which certain tasks are available, if subjects naturally tend to work on the tasks in the order that they appear, then the same behaviour has different implications across treatments.

For a precise distinction between the sets of instructions, see Section B.2 for Chapter 2 and Section C.1 for Chapter 3. Figure C3 shows improved comprehension over experiments.

[2] Another minor change is that we moved from the concept of Experimental Currency Units (ECUs) into 'Points' as we felt this was less complicated terminology

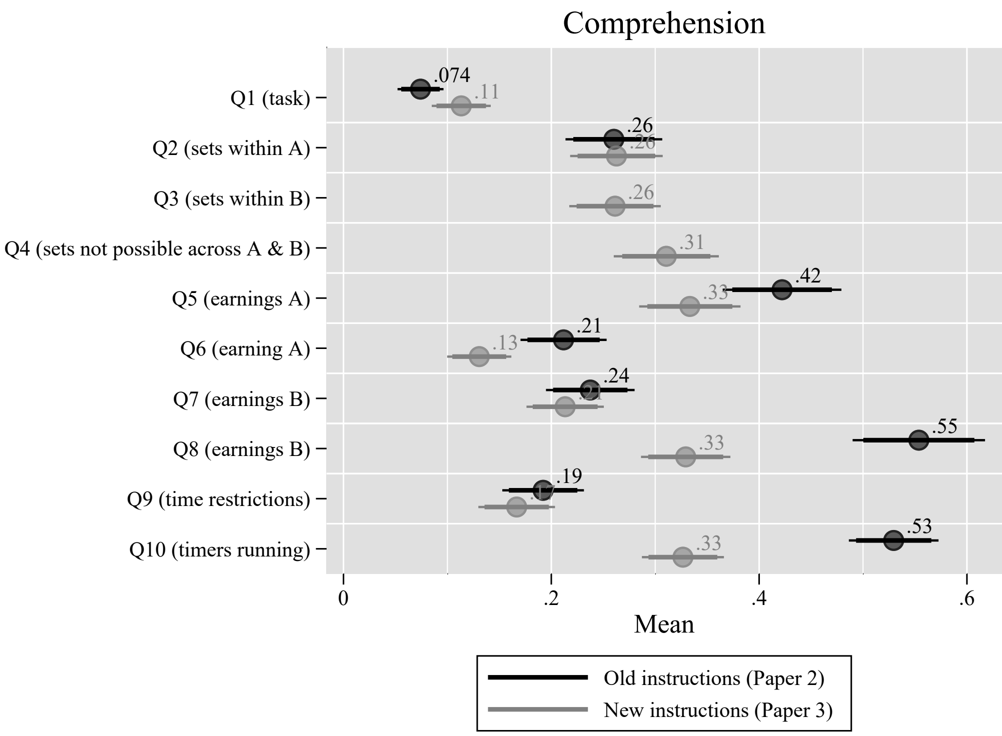


Figure C3: Comprehension across instructions

C.5 Experiments

C.5.1 Summary statistics

	Mean	SD	Min	Max	N
Baseline					
Performance					
Total strings	44.62	14.97	0.00	91.00	177.00
Possible sets	10.81	3.73	0.00	22.00	177.00
Total sets	9.09	4.88	0.00	21.00	177.00
Time allocation					
Time A (> 20 seconds)	0.58	0.49	0.00	1.00	177.00
Earnings					
Earnings (narrow) £ p/h	8.39	3.45	2.16	21.74	82.00
Earnings (broad) £ p/h	11.83	5.63	3.31	34.06	66.00
Demographics					
Female	0.53	0.50	0.00	1.00	175.00
Age	39.10	13.95	18.00	73.00	177.00
Tertiary Education	0.65	0.48	0.00	1.00	176.00
Comprehension					
Comprehension attempts	0.68	0.47	0.00	1.00	177.00
Attention					
Attention timings	6.46	1.94	1.00	8.00	177.00
Attention payments	5.80	2.28	1.00	8.00	177.00
attn_diffr	0.66	2.84	-7.00	7.00	177.00
Cognition					
Perceived maths ability	5.74	2.75	0.00	10.00	177.00
Numeric ability	1.65	0.52	0.00	2.00	177.00
Cognitive reflection	1.80	0.84	0.00	3.00	177.00
Lottery isolation	0.62	0.49	0.00	1.00	177.00
Other					
Activity A (top of screen)	1.00	0.00	1.00	1.00	177.00
Risk taking (standardised)	-0.00	0.77	-1.68	2.01	177.00
Expectation	4.97	1.80	1.00	8.00	177.00
Urgent					
Performance					
Total strings	46.40	15.76	0.00	96.00	174.00
Possible sets	11.23	3.92	0.00	24.00	174.00
Total sets	9.61	4.87	0.00	23.00	174.00
Time allocation					
Time A (> 20 seconds)	0.60	0.49	0.00	1.00	174.00
Earnings					
Earnings (narrow) £ p/h	8.73	3.64	1.88	21.27	83.00
Earnings (broad) £ p/h	11.94	6.00	3.96	29.85	65.00
Demographics					
Female	0.53	0.50	0.00	1.00	173.00
Age	39.16	13.05	18.00	74.00	174.00
Tertiary Education	0.65	0.48	0.00	1.00	173.00
Comprehension					
Comprehension attempts	0.68	0.47	0.00	1.00	174.00
Attention					
Attention timings	6.72	1.70	1.00	8.00	174.00
Attention payments	5.56	2.22	1.00	8.00	174.00
attn_diffr	1.17	2.91	-7.00	7.00	174.00
Cognition					
Perceived maths ability	5.76	2.60	0.00	10.00	174.00
Numeric ability	1.42	0.69	0.00	2.00	174.00
Cognitive reflection	1.84	0.85	0.00	3.00	174.00
Lottery isolation	0.74	0.44	0.00	1.00	174.00
Other					
Activity A (top of screen)	1.00	0.00	1.00	1.00	174.00
Risk taking (standardised)	-0.07	0.83	-2.08	1.73	173.00
Expectation	5.23	1.64	1.00	8.00	174.00
Busy					
Performance					
Total strings	44.61	15.65	16.00	108.00	186.00
Possible sets	10.79	3.94	4.00	27.00	186.00
Total sets	7.97	5.03	0.00	26.00	186.00
Time allocation					
Time A (> 20 seconds)	0.58	0.50	0.00	1.00	186.00
Earnings					
Earnings (narrow) £ p/h	8.53	3.55	3.07	18.54	81.00
Earnings (broad) £ p/h	11.18	5.06	3.95	27.41	67.00
Demographics					
Female	0.52	0.50	0.00	1.00	185.00
Age	38.54	12.34	20.00	72.00	186.00
Tertiary Education	0.62	0.49	0.00	1.00	186.00
Comprehension					
Comprehension attempts	0.66	0.47	0.00	1.00	186.00
Attention					
Attention timings	6.46	1.89	1.00	8.00	186.00
Attention payments	5.63	2.32	1.00	8.00	186.00
attn_diffr	0.83	2.70	-7.00	7.00	186.00
Cognition					
Perceived maths ability	5.67	2.72	0.00	10.00	186.00
Numeric ability	1.59	0.58	0.00	2.00	186.00
Cognitive reflection	1.76	0.87	0.00	3.00	186.00
Lottery isolation	0.66	0.48	0.00	1.00	186.00
Other					
Activity A (top of screen)	1.00	0.00	1.00	1.00	186.00
Risk taking (standardised)	0.07	0.85	-1.92	2.98	186.00
Expectation	5.02	1.80	1.00	8.00	186.00
Total					
Performance					
Total strings	45.19	15.46	0.00	108.00	537.00
Possible sets	10.94	3.86	0.00	27.00	537.00
Total sets	8.87	4.97	0.00	26.00	537.00
Time allocation					
Time A (> 20 seconds)	0.59	0.49	0.00	1.00	537.00
Earnings					
Earnings (narrow) £ p/h	8.55	3.54	1.88	21.74	246.00
Earnings (broad) £ p/h	11.65	5.55	3.31	34.06	198.00
Demographics					
Female	0.53	0.50	0.00	1.00	533.00
Age	38.92	13.09	18.00	74.00	537.00
Tertiary Education	0.64	0.48	0.00	1.00	535.00
Comprehension					
Comprehension attempts	0.67	0.47	0.00	1.00	537.00
Attention					
Attention timings	6.55	1.85	1.00	8.00	537.00
Attention payments	5.66	2.27	1.00	8.00	537.00
attn_diffr	0.88	2.82	-7.00	7.00	537.00
Cognition					
Perceived maths ability	5.72	2.68	0.00	10.00	537.00
Numeric ability	1.55	0.61	0.00	2.00	537.00
Cognitive reflection	1.80	0.85	0.00	3.00	537.00
Lottery isolation	0.67	0.47	0.00	1.00	537.00
Other					
Activity A (top of screen)	1.00	0.00	1.00	1.00	537.00
Risk taking (standardised)	0.00	0.82	-2.08	2.98	536.00
Expectation	5.07	1.75	1.00	8.00	537.00

Table C1: Summary statistics

C.5.2 Supplementary results

Performance

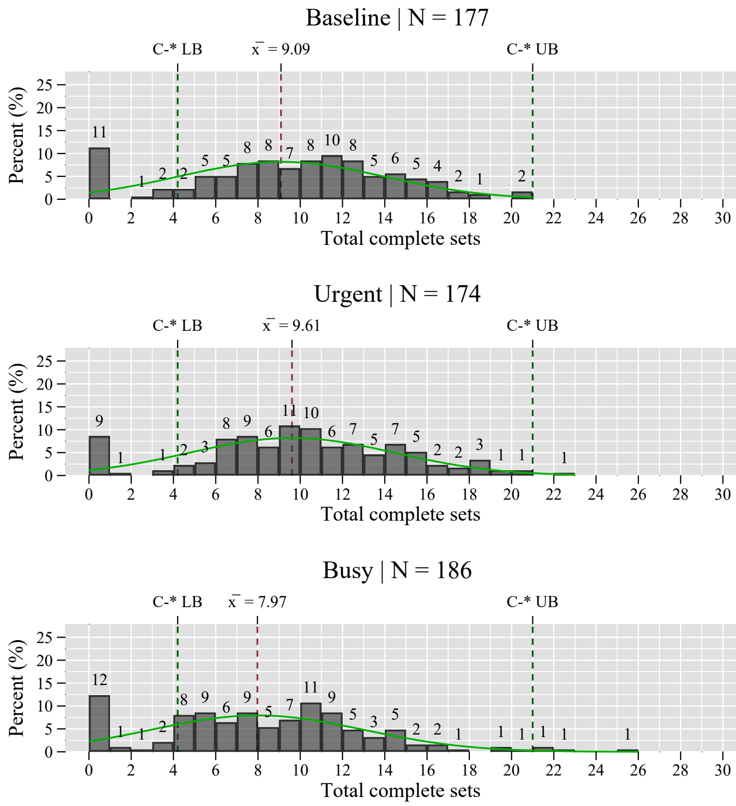


Figure C4: Performance for experiment (actual sets produced)

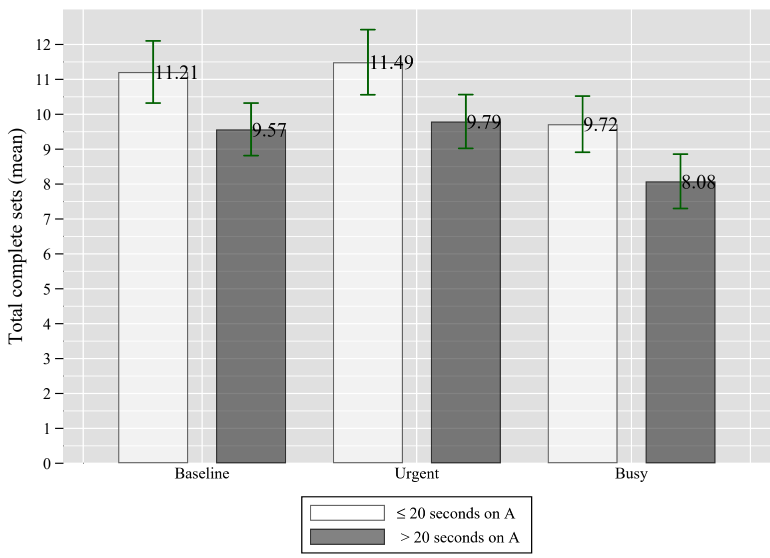


Figure C5: Total complete sets across bracketing behaviour

Time spent on activity A

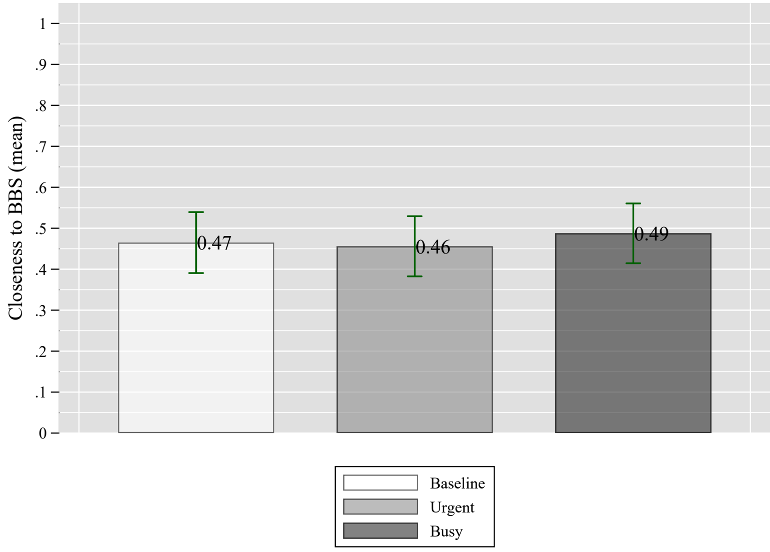


Figure C6: Closeness to optimality estimates

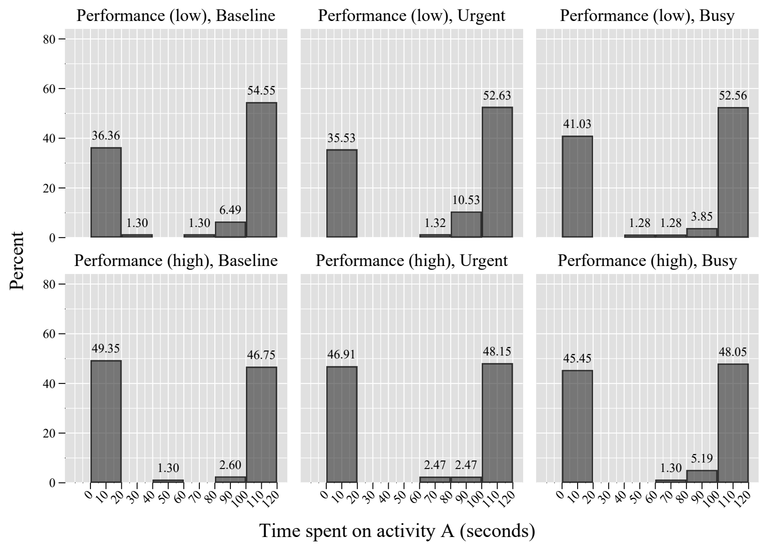


Figure C7: Time on activity A for experiments across performance

Notes. $\overline{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off. The sample size is restricted to those within C* and who produced at-least one set.

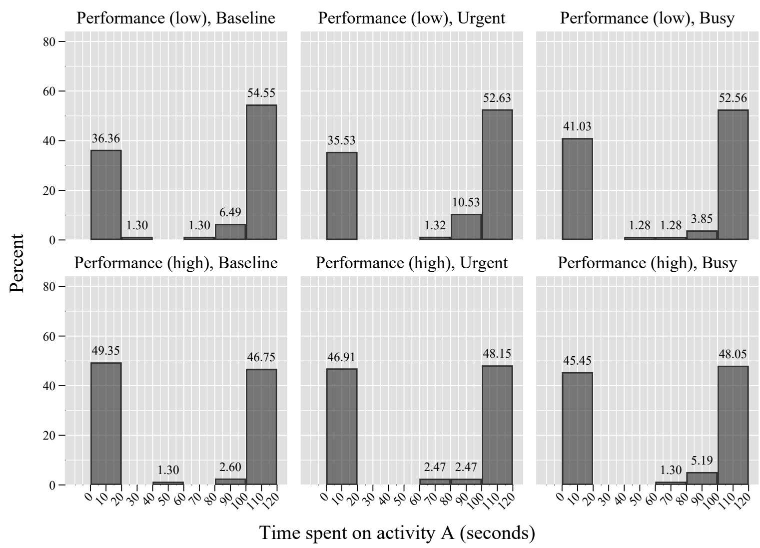


Figure C8: Time on activity A for experiments across comprehension

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off. The sample size is restricted to those within C* and who produced at-least one set.

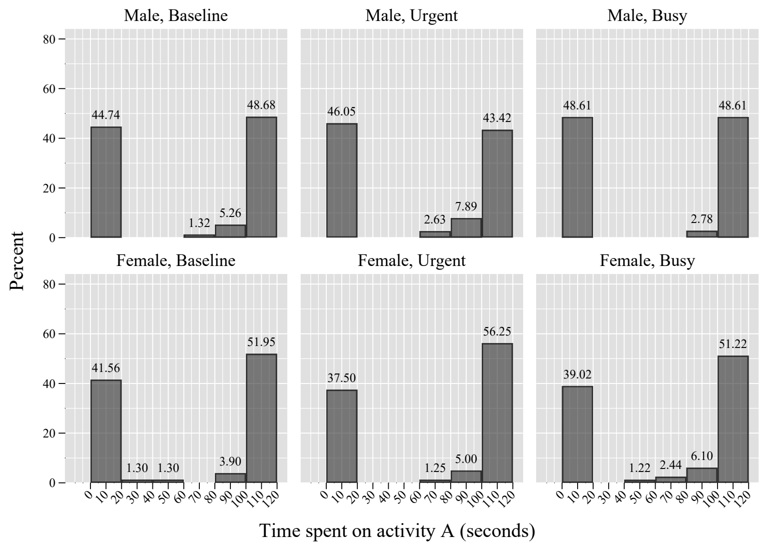


Figure C9: Time on activity A for experiments across gender

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off. The sample size is restricted to those within C* and who produced at-least one set.

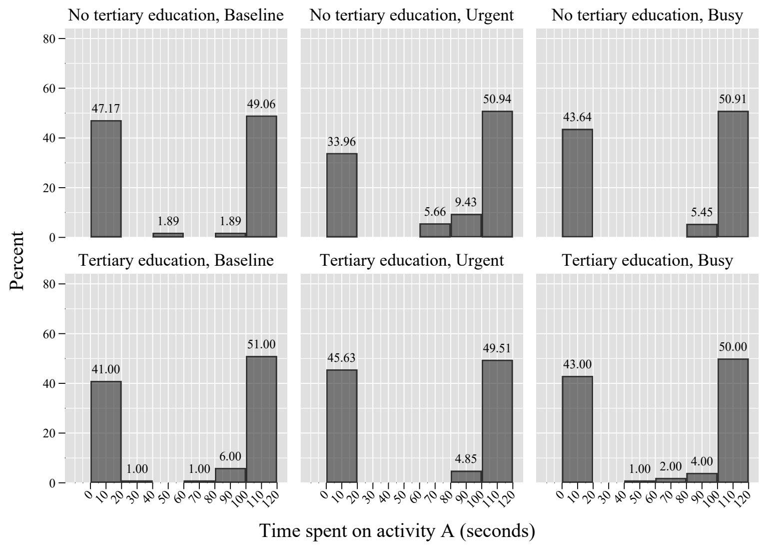


Figure C10: Time on activity A for experiments across tertiary education

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off. The sample size is restricted to those within C* and who produced at-least one set.

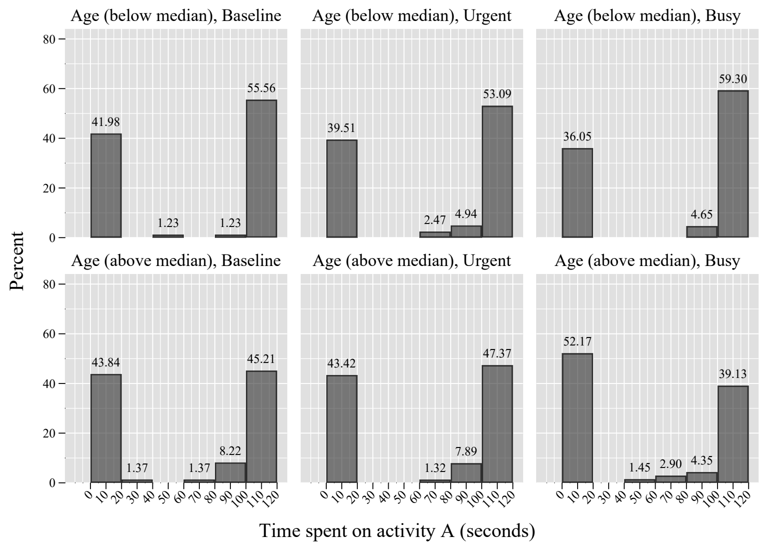


Figure C11: Time on activity A for experiments across age

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off.

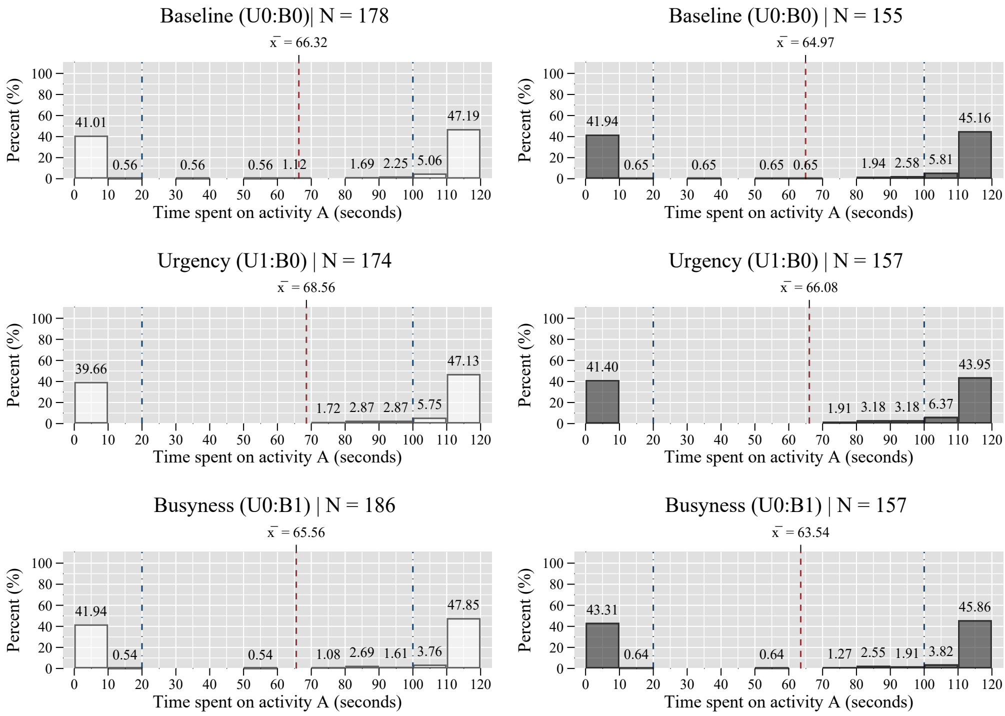


Figure C12: Time on activity A for experiments (finer bins)

Notes. $\bar{x}$ represents the mean. The bin size has been set to 12, meaning that each bin represents 10-second intervals. The left dash-dot line represents our BB cut-off and the right dash-dot line our NB cut-off.

Tunnelling

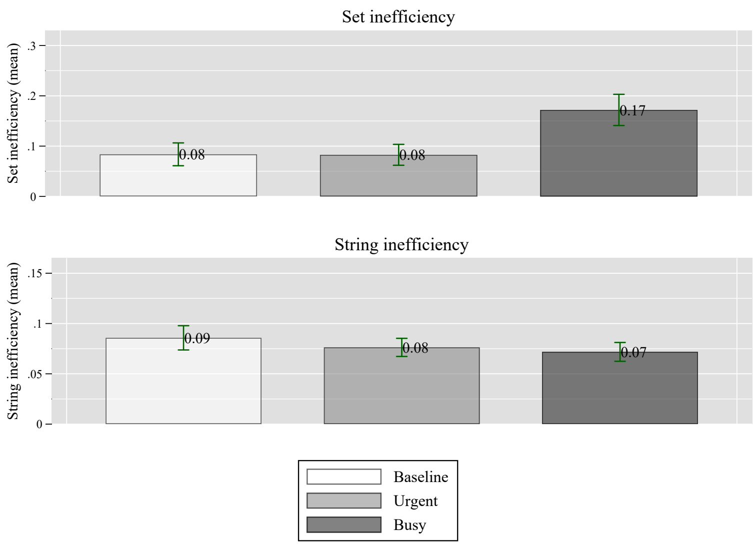


Figure C13: Inefficiency rates across treatments

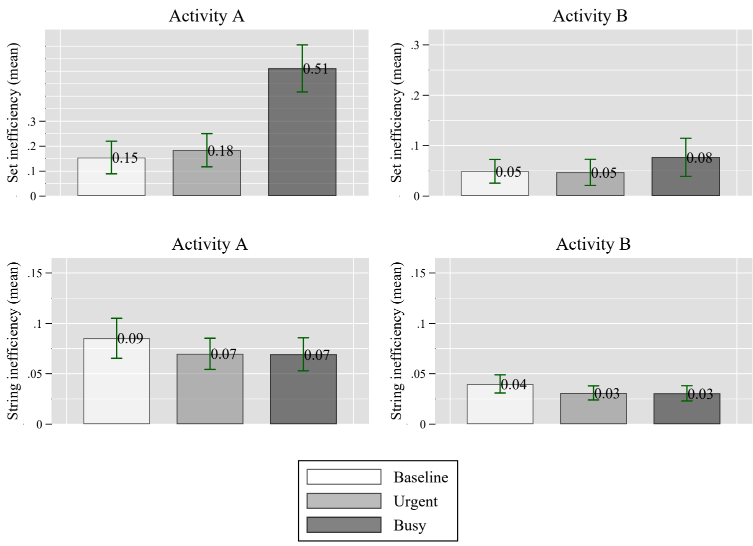


Figure C14: Inefficiency rates across activities A & B for narrow bracketers

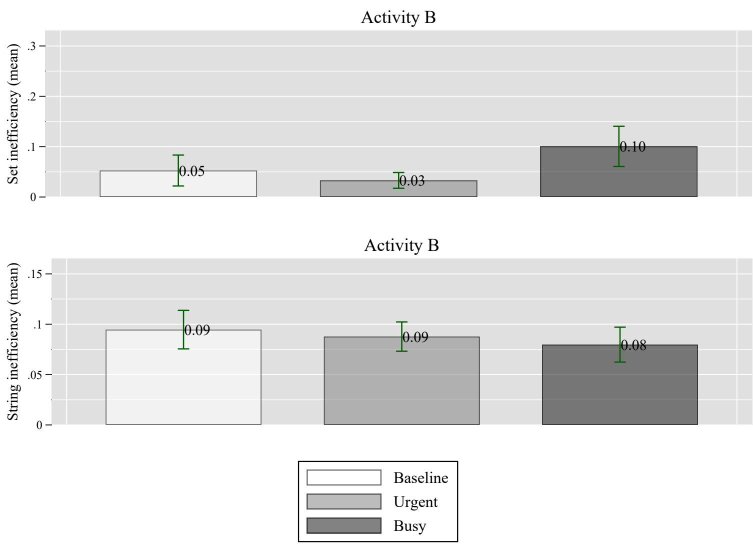


Figure C15: Inefficiency rates across activity B for broad bracketers

	Set inefficiency			String inefficiency		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
TREATMENTS						
Base. Baseline						
Urgent	0.00	0.00	-0.02	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Busy	0.05***	0.05***	0.02	-0.02**	-0.02**	-0.02*
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
$Time_A$ > 20 seconds		0.03***	0.00		-0.03***	-0.03**
		(0.01)	(0.02)		(0.01)	(0.01)
INTERACTIONS						
Urgent × $Time_A$ > 20 seconds			0.03			-0.00
			(0.03)			(0.02)
Busy × $Time_A$ > 20 seconds			0.06**			0.01
			(0.03)			(0.02)
Constant	0.36***	0.32***	0.34***	0.13***	0.16***	0.16***
	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	465	465	465	465	465	465
R^2	0.48	0.49	0.50	0.06	0.09	0.09

Table C2: Ordinary least squares: inefficiency rates across treatments

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variables: Columns 1 and 2 —Set inefficiency (0 to 1); Columns 3 and 4 —String inefficiency (0 to 1). Independent measures include the three treatments (with the Baseline as the base); the binary measure of time spent ($Time_A$ > 20 seconds); and their interaction terms. Additional controls include the three treatments, age, gender, tertiary education, and subject performance in terms of sets generated.

	Set inefficiency		String inefficiency	
	(1) OLS	(2) OLS	(3) OLS	(4) OL
Components worked	0.02***	0.02***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.03	-0.09***	0.10***	0.09***
	(0.02)	(0.03)	(0.01)	(0.01)
Additional controls	No	Yes	No	Yes
Observations	466	462	466	462
N clusters				
R-squared	0.06	0.16	0.02	0.02

Table C3: Ordinary least squares: inefficiency rates across components worked

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variables: Columns 1 and 2 —Set inefficiency (0 to 1); Columns 3 and 4 —String inefficiency (0 to 1). Independent measure includes the number of components worked. Additional controls include the three treatments, age, gender, tertiary education, and subject performance in terms of sets generated.

Ex-post measures

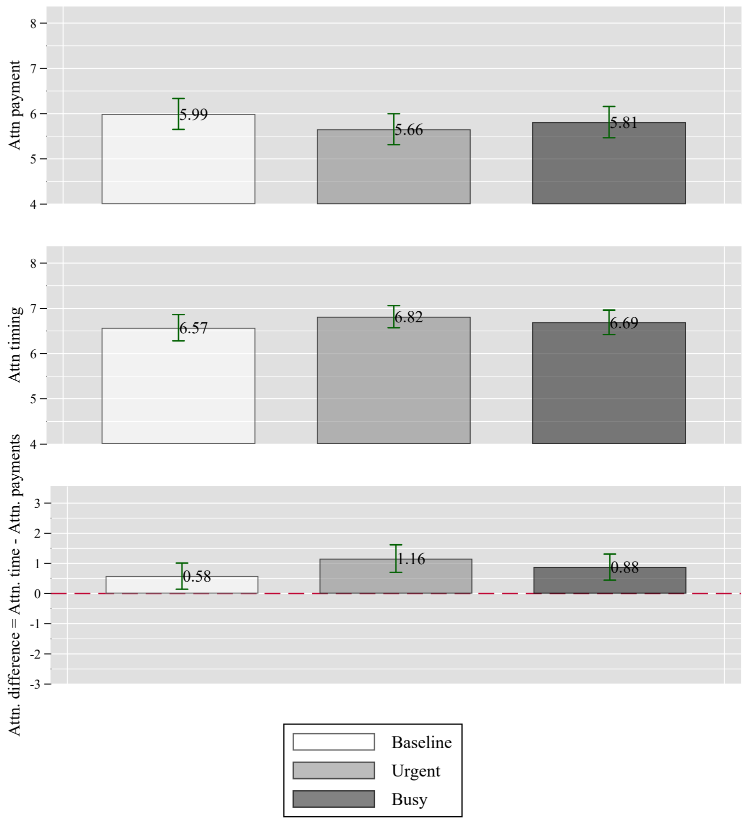


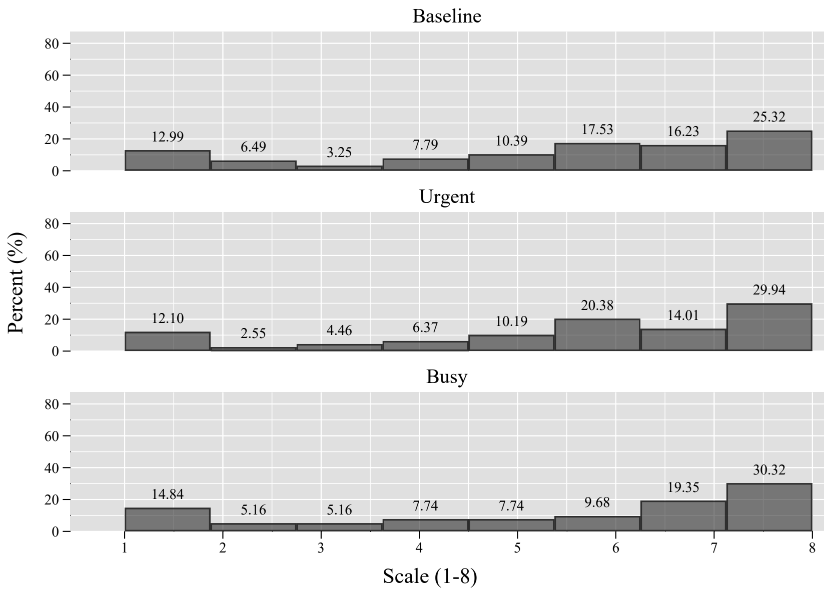
Figure C16: Relative attentional differences (aggregated)

	Attn. payment	Attn. timing	Attn. diffeence	$Time_A > 20$ seconds		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
TREATMENTS						
Base. Baseline						
Urgent	-0.36 (0.23)	0.25 (0.20)	0.14 (0.09)			
Busy	-0.03 (0.24)	0.12 (0.21)	0.04 (0.09)			
ATTENTION						
Attention to Payment				-0.08*** (0.01)		
Attention to Time					0.01 (0.01)	
Attentional difference						0.19*** (0.03)
Constant	5.07*** (0.66)	6.42*** (0.58)	8.31*** (0.26)	1.29*** (0.16)	0.83*** (0.18)	-0.71** (0.29)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
N observations	462	462	462	462	462	462
R-squared	0.14	0.01	0.08	0.16	0.05	0.13

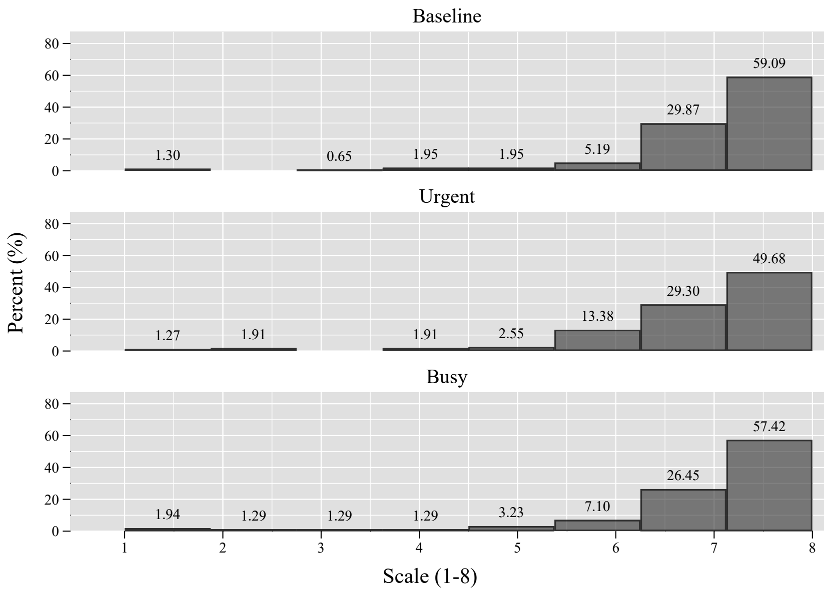
Table C4: Ordinary least squares: attentional focus

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.

Notes. Dependent variables: Column 1 —Attention to payment (1 to 8); Column 2 —Attention to timing (1 to 8); Column 3 —Attentional difference (−7 to +7); and Columns 4, 5 and 6 —the binary measure of time spent on activity A ($Time_A > 20$ seconds). Independent measures include the studies grouped into (i) three treatments (Columns 1, 2 and 3): Baseline (the base), Urgent; and Busy; (ii) attentional focus metrics (Columns 4, 5 and 6); attention to payment (1 to 8); attention to timing (1 to 8); and attentional difference (−7 to +7) Additional controls include age, gender, tertiary education, subject performance in terms of total strings correctly reversed, activity order, and expectations about performance.



(a) Perceived urgency



(b) Perceived busyness

Figure C17: Distribution of manipulation checks

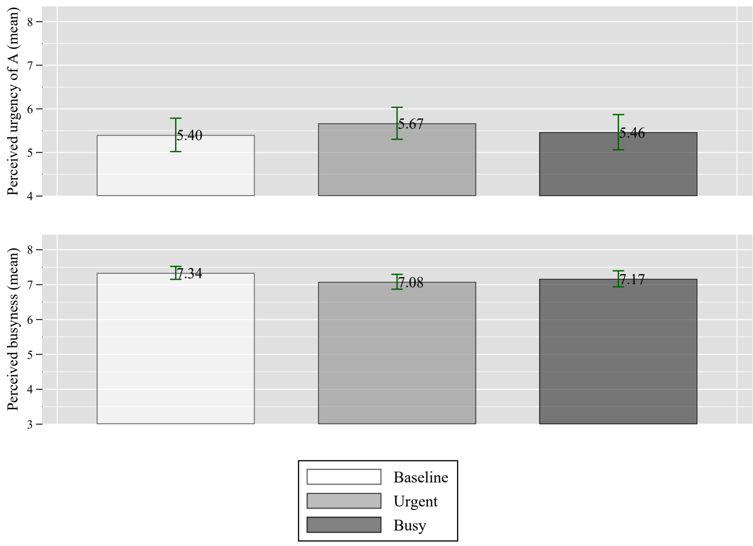


Figure C18: Perception of urgency and busyness (aggregated)

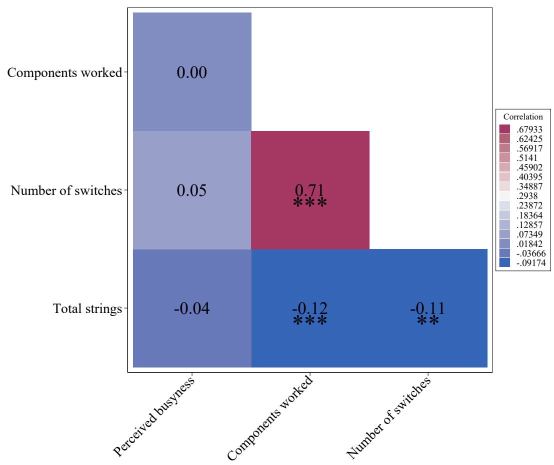


Figure C19: Ex-post measure correlations

* p < 0.1, ** p < 0.05, *** p < 0.01.
Notes. Spearman's rank correlation coefficients (ρ). The blue and red shaded boxes represent negative and positive correlations, respectively with intensity of shaded region representing relative strength of correlation (as indicated by legend).

	Full sample			High comprehension		
	(1) ME	(2) ME	(3) ME	(4) ME	(5) ME	(6) ME
TREATMENTS						
Base. Baseline						
Urgent	-0.02	0.00	0.03	0.01	0.02	0.05
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Busy	-0.08	-0.05	-0.06	-0.08	-0.05	-0.06
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
PERCEPTIONS						
Base. Equal attention						
More on timing	0.29***			0.33***		
	(0.05)			(0.06)		
More on payment	-0.05			0.00		
	(0.06)			(0.07)		
Perceived urgency of A (6-8)		0.33***			0.32***	
		(0.03)			(0.04)	
Perceived busyness (6-8)			0.20**			0.26**
			(0.08)			(0.11)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	465	465	463	324	324	322
Psuedo R^2	0.14	0.16	0.07	0.14	0.15	0.08
Log-lik	-273.78	-266.85	-294.30	-191.59	-189.42	-204.71
χ^2	80.62	88.37	37.53	61.71	63.53	30.50

Table C5: Probit model: time on activity A (binary)

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.
Notes. Dependent variable: A binary measure taking on a value of 1 if $Time_A > 20$ seconds. Independent measures include the: three treatments (with the Baseline as the base); a binary measure of attentional differences with equal attention on both payment and timing as the base; and binary measure of high perceived urgency and busyness (6-8). Additional controls include age, gender, tertiary education, and subject performance in terms of sets generated.

C.5.3 Data from Chapters 2 and 3 pooled

Correlations across variables

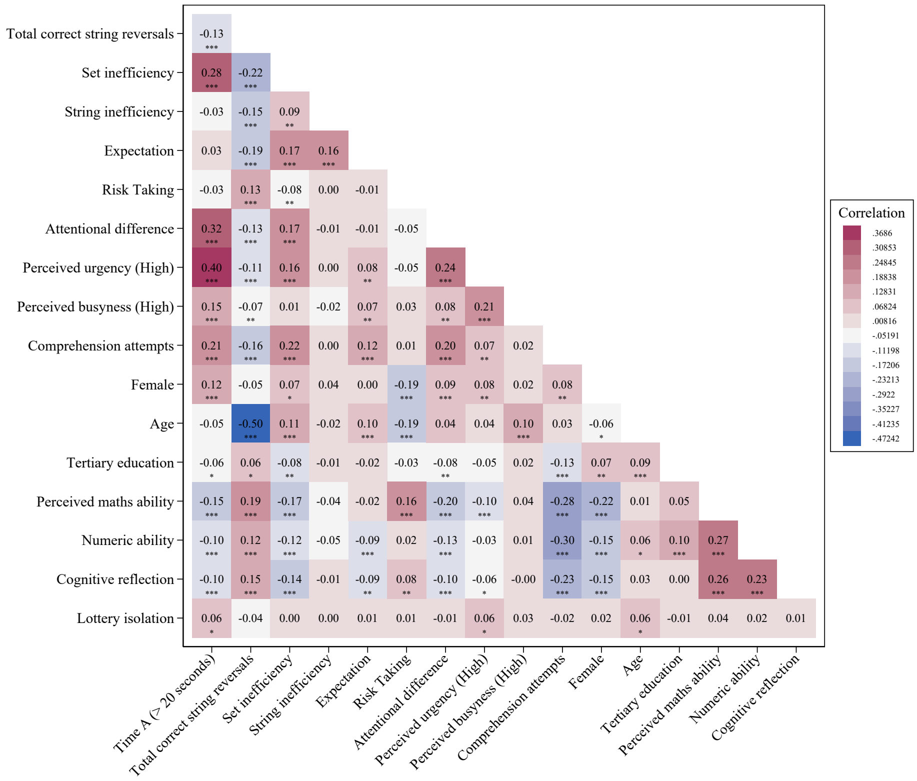


Figure C20: Correlations across variables

* p < 0.1, ** p < 0.05, *** p < 0.01.

Notes. Spearman's rank correlation coefficients (ρ). This data includes the experiments from Chapter 3 as well as studies 1, 2 and 3 from Chapter 2. The blue and red shaded boxes represent negative and positive correlations, respectively with intensity of shaded region representing relative strength of correlation (as indicated by legend).

Simulated experiment

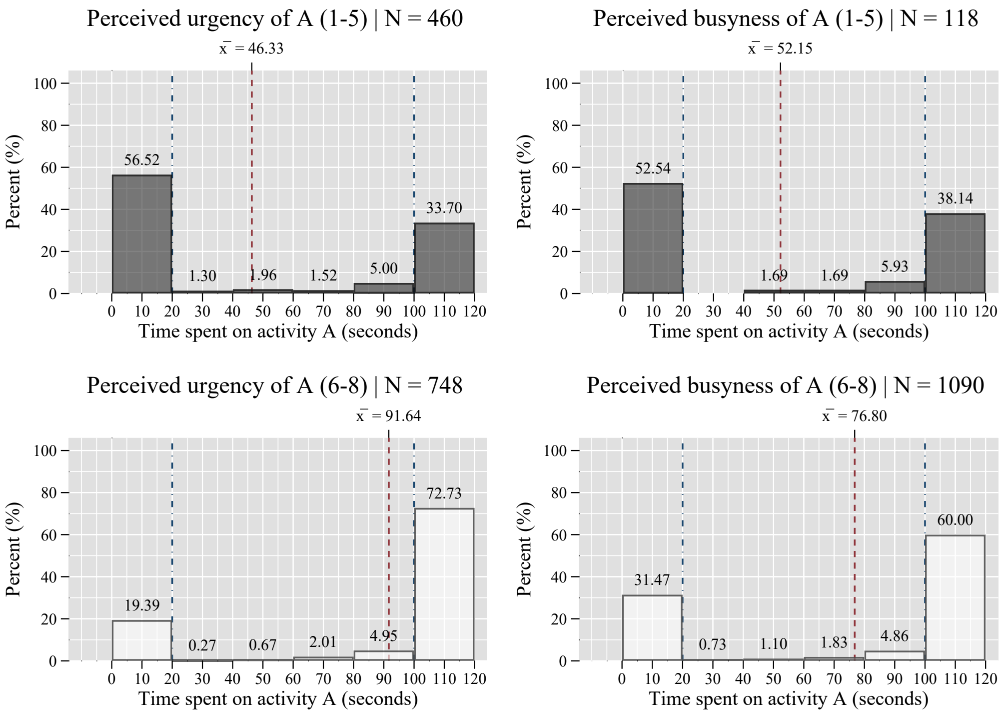


Figure C21: Results for dichotomised manipulation checks

Notes. $\bar{x}$ represents the mean. The bin size has been set to 6, meaning that each bin represents 20-second intervals. This data includes the experiment from Chapter 3 as well as studies 1, 2 and 3 from Chapter 2 ($\alpha = 9, 7, 5, 1.5, 0.5$, excluding $\mu = 1$ since the x-axis for $\mu = 1$ runs to 360 seconds).