

# Arbitraging Narrow Bracketers \*

Francesco Fallucchi

Marc Kaufmann

University of Bergamo<sup>†</sup>

Central European University (CEU)<sup>‡</sup>

April 2, 2025

## Abstract

Many important economic outcomes result from the combined effects of several choices, so the best option is not determined from each choice in isolation, but depends on how each choice alters total outcomes. We formally show that narrow bracketing — treating choices in isolation — can be distinguished from broad bracketing — combining the choices — if and only if there exist price variation across context: there is some bundle for which a person is willing to pay more in one choice than in another. In this case, a narrow bracketer can be arbitrated, buying the bundle when it is expensive and selling when it is cheap in simultaneous choices. We design and run two experiments to identify bracketing from price variation. In a between-subjects design where we vary the amount of work to generate price variation, we reject broad bracketing and fail to reject narrow bracketing. In a within-subject design we directly test bracketing by attempting to arbitrage our participants. For price variation coming from varying amounts of money, 50.3% of subjects are classified as narrow bracketers, and only 14.6% as broad bracketers, the remainder being inconsistent with both. This changes for price variation coming from violations of expected utility — 13.6% narrow vs 46.3% broad — and of the weak axiom of revealed preference — 26.3% narrow vs 38.1% broad.

**JEL Classifications:** C91, D91, J01

**Keywords:** Choice bracketing, Arbitraging, Experiment, Individual decision-making

---

\*A previous version of this paper was circulated under the title "Narrow Bracketing in Work Choices". The studies were pre-registered in the AEA RCT Registry with ID 0003412 (see <https://doi.org/10.1257/rct.3412>) and OSF (see <https://osf.io/5q7wn/>). The data and analysis are hosted on GitHub at <https://github.com/MarcKaufmann/narrow-bracketing-in-work-choices>.

<sup>†</sup> francesco.fallucchi@unibg.it, Senior Assistant Professor, Department of Economics, Via dei Caniana 2, 24127 Bergamo, Italy.

<sup>‡</sup>Corresponding author: kaufmannm@ceu.edu, Associate Professor, Department of Economics and Business, Central European University, Quellenstrasse 51-55, 1100 Vienna, Austria.

# 1 Introduction

Many of the most important economic outcomes — such as savings, health, careers, and relationships — result from countless interdependent decisions. In the work domain, such cumulative decision making is bound to become even more important as remote, flexible, and online work becomes more popular (Farrell et al., 2018, 2019). In order to obtain the best outcomes in these circumstances, a person must, in every decision, at least take into account cumulative outcomes so far. Whenever a person fails to do so, she is engaging in narrow bracketing (Read et al. (1999)). The prior literature focuses on nonwork choices, such as monetary and lottery choices (Tversky and Kahneman (1981), Riedelmeier and Tversky (1992), Rabin and Weizsäcker (2009), Koch and Nafziger (2019), Vorjohann (2020), and Dekel et al. (2021)); social choices Exley and Kessler (2018); or both of these and induced preferences (Ellis and Freeman (2024)). Broad bracketing in these situations is either complicated (when combining lotteries), or not necessarily the relevant target (when making social choices). In contrast, we investigate narrow bracketing in economically essential work choices, where accounting for these outcomes is simple and relevant.

We contribute to the literature on bracketing by formalizing a measurable notion of price variation across choices that is key to identifying narrow bracketing and apply it to identify narrow bracketing across two experiments. First, we show that every narrow bracketer can be arbitrated across choices with different buying and selling prices: they purchase a good when it is expensive in one choice only to sell it when it is cheap in another. We characterize the set of preferences under which such arbitrage opportunities do not arise due to equal buying and selling prices, in which case narrow bracketers behave optimally and are observationally equivalent to broad bracketers. Second, building on this theoretical foundation, we design and conduct two experiments to identify price variations arising from different sources — from nonlinear preferences, violations of expected utility, or inconsistencies with the weak axiom of revealed preference (WARP) — and use these variations to classify individuals as narrow bracketers, broad bracketers, or inconsistent with both.

To illustrate buying and selling prices, consider a person making choices over money and hours of work. Suppose that this person has a marginal disutility of \$5 per hour of work for the first hour of work, and a marginal disutility of \$10 per hour for every additional hour. Then this person chooses 1 hour of work for \$6 over no work for no pay — or in other words, they sell 1 hour of work for \$6. This person also chooses 1 hour of work for no pay over 2 hours of work for \$9 — i.e., they refuse to sell 1 *extra* hour of work (the difference between the two options) for \$9, which we refer to as *buying* one hour of work for \$9. Defining buying and selling of 1 extra hour of work in this way, we find that the highest buying price for 1 extra hour of work is \$10, and the lowest selling price is \$5. Now consider a person who is given both of these choices simultaneously and who will receive their combined outcome. Narrow bracketing requires that the person optimizes each choice separately, so a narrow bracketer would choose the same way as if each choice was the only choice. Therefore, they would sell 1 hour of work for \$6 in the first choice, and buy 1 hour of

work for \$9 in the second choice, even though switching both choices by buying 1 hour for \$6 in selling 1 hour of work for \$9, would have earned them \$3 more for the same total amount of 2 hours of work. A broad bracketer would optimize the sum of the choices instead, and thus choose between the following total outcomes: 3 hours of work for \$15, 2 hours for \$9, 2 hours for \$6, or 1 hour for no money. Therefore, they would never choose 2 hours for \$6 over 2 hours for \$9 and so cannot be arbitrated by offering simultaneous choices.

This example highlights our first result in Section 2, which shows that whenever there exists some bundle for which the highest buying price exceeds the lowest selling price, then a narrow bracketer can be arbitrated by buying high in one choice and selling low in another. More surprisingly, if for every bundle  $X$  the highest buying price equals the lowest selling price, denoted  $P(X)$ , then a narrow bracketer behaves identically to a broad bracketer, and thus we cannot identify bracketing. We show that this occurs if and only if the choices can be rationalized by preferences — so that they satisfy the weak axiom of revealed preference (WARP) — and that  $P(\cdot)$  provides a utility representation for these preferences, satisfying additivity:  $P(X + Y) = P(X) + P(Y)$ . Additivity of  $P(\cdot)$  implies that from each choice set, the person chooses the bundle with the highest price, which implies in particular that the utility over amounts is linear in quantities and utility over lotteries is linear in probabilities, and so satisfies expected utility. For people with such preferences, narrow bracketing is fundamentally unidentified, and so in order to identify bracketing, we must find some bundle for which the highest buying price exceeds the lowest selling price. This is guaranteed to be the case if we observe a violation of WARP, of expected utility, or of linear utility.

In our first main experiment described in Section 4, a between-subjects design conducted on Amazon Mechanical Turk recruiting 716 subjects, we generate price variation by varying the amount of tasks participants have to do by default, in order to test a particular form of narrow bracketing often called endowment bracketing. Concretely, in our first scenario, some of the participants are asked for the smallest price — their reservation wage — for which they are willing to do 15 tasks rather than 0 tasks, while others are asked for their reservation wage for doing 30 rather than 15 tasks. We find that these reservation wages are significantly different, \$2.88 vs. \$2.31 ( $p\text{-value} < 0.001$ ) respectively, which provides us with the difference in prices needed to identify population-level bracketing. In a third treatment, participants are told on the same page that they have to do a baseline of 15 tasks and then asked for their reservation wage for doing 15 additional tasks. Under this treatment, a broad bracketer would recognize that the decision is equivalent to choosing between 30 tasks instead of 15, whereas a narrow bracketer would perceive it as a choice between 15 tasks and none. We find that the reservation wage elicited is \$2.07 — 28% significantly lower than the reservation wage of \$2.88 expected for broad bracketers, but not significantly different from the reservation wage of \$2.31 expected for narrow bracketers, suggesting that work decisions are consistent with narrow bracketing. In a second scenario, where we increase the number of baseline tasks in each treatment by 15, we do not generate enough price variation and, therefore, fail to reject both narrow and broad bracketing.

In Section 5, we describe the design of our second study, where we use arbitrage in a within-study design to test

for narrow bracketing beyond endowment bracketing. Following closely the structure of our example described above, we first search for pairs of choices that reveal potential for arbitrage: participants are willing to buy work for a higher price in one choice than they are willing to sell in the other. We do so by offering participants choices from three domains, each varying a different feature to identify a potential source that would generate a price difference. First, we look for non-linear utility by varying the amount of tasks; second, we look for violations of WARP by adding extra (unchosen) options to choice sets; and finally, we look for violations of expected utility by varying the probability with which a fixed amount of tasks has to be done. If we identify such pairs of choices, we let participants make this pair of choices simultaneously, as well as make a single choice equal to the sum of both choices, so that a narrow bracketer will throw away money when making the simultaneous choices, but not when making the combined choice.<sup>1</sup> A broad bracketer, in contrast, will exhibit the same behavior across both scenarios, allowing us to classify an individual as aligning with narrow but not broad bracketing if they discard money in the simultaneous choice; as aligning with broad but not narrow bracketing if their choices match in both the simultaneous and the combined scenarios; or as neither in other cases.

Our results from this experiment are described in Section 6. We conducted it on Prolific, recruiting 572 subjects to make 35 decisions of work linked to the three domains described above. We find that around 60% of the participants who complete the experiment are identified under our arbitrage condition in at least one of the three settings, and on average we have an identification of 30% in each setting. Among them, we find that roughly 50% are identified as narrow bracketers in the linear setting, while less than 15% are broad bracketers; in the other two settings, the share of broad bracketers is higher than the share of narrow bracketers: 38% vs. 26% for WARP violations and 46% vs. 14% for violations of expected utility. The remaining subjects cannot be classified as inconsistent with both narrow and broad bracketing. While our focus is on individuals for whom we can find pairs of choices that ensure price variation, we also offer simultaneous and combined choices to the remaining participants and categorize them, since these extra choices might themselves provide identifying variation. As is to be expected, the majority of these individuals remain unidentified (on average 70% in each domain, with 39% of participants unidentified in all domains), as their choices are consistent with WARP, expected utility, and linear utility, and thus with both narrow and broad bracketing. For the remaining subjects, we have findings similar to those among the identified ones.

Our contribution to the literature on narrow bracketing is threefold. First, we provide a characterization for when bracketing is identifiable — that narrow bracketing can be distinguished from broad bracketing — and show that when it is the case, we can identify it via an arbitrage test. Such arbitrage was already present in Tversky and Kahneman (1981) and explicitly highlighted in Rabin and Weizsäcker (2009) for the specific case of lotteries over money when absolute risk aversion is varying and in Andreoni et al. (2018) for the case of arbitraging payments for work over time. Our result establishes that such arbitrage arguments can be applied much more broadly and are an identifying feature

---

<sup>1</sup> Assuming they prefer more money to less, which might not hold in actual data for instance if participants are inattentive.

of narrow bracketing. An advantage of arbitrage tests for narrow bracketing is that they make clear that the behavior is a mistake, since the person throws away money, whereas more indirect tests — such as in our first experiment — cannot rule out that the person made a mistake in the combined choice. We relate our characterization result to the literature at the end of Section 2, where we briefly discuss where the identifying variation in prices comes from in earlier tests of bracketing.

Second, our framework provides direct tests for both narrow and broad bracketing, as well as a test for whether bracketing is identified in a given dataset. Some papers only test (and usually reject) broad bracketing (Rabin and Weizsäcker (2009), Abeler and Marklein (2017), Koch and Nafziger (2019)) and thus cannot establish that what people do when they are not bracketing broadly is actually narrow bracketing, other papers do effectively test for both (Kahneman and Tversky (1979), problem 10), but none provide a general framework for testing both. One notable exception is Ellis and Freeman (2024) who apply a revealed preference approach to test narrow and broad bracketing. In their framework, conceptually narrow and broad bracketing are only defined for choices consistent with transitive preferences, and so practically one cannot talk about or test bracketing in any choice dataset violating WARP. Our framework defines bracketing rather as a property of choices and thus extends to situations where people are subject to decoy, range, or reference effects. Moreover, our finding that price variation is essential provides practical guidance to researchers seeking identification.

Finally, our experimental results using the arbitrage test show the value of testing for narrow bracketing across different variations of the choices. Our results for price variation driven by non-linear utility — most similar to previous studies — strongly support narrow bracketing. The evidence in favor of narrow bracketing is weaker for price variation driven by violations of expected utility or of WARP. This attenuation may result from our within-subjects design, which could make direct arbitrage between choices more salient, or it may indicate that different variations encourage more deliberate integration of choices, shedding light on the mechanisms underlying narrow bracketing.

We conclude in Section 7, discussing identification strategies that don't rely on direct (and thus overly salient) arbitrage, and how the identification condition will change when agents make choices from choices with additional structure — such as choosing how much to buy of one good at once, but making separate choices for each good.

## 2 Conceptual Framework of Narrow Bracketing

In this section, we consider an agent with a given value function  $v$  that is monotonic in money and ask when a person can incur a cost from narrow bracketing when facing two simultaneous choices. We show that narrow bracketing leads to different choices and thus to a cost if and only if there is variation across choice contexts in what we define the buying and selling prices for a bundle of goods. Under such variation, a narrow bracketer can be arbitrated by providing them with two choices, one in which they choose the bundle despite giving up a large amount of money

(the price is high), one in which they choose a small amount of money over the bundle (the price is low) — in other words, they buy high and sell low. When these prices are constant across all choice contexts, then a narrow bracketer behaves like a broad bracketer, thus incurs no costs, and it is impossible to identify broad from narrow bracketing. Such constant prices imply that the value function  $v$  is consistent with transitive preferences that are additive: they satisfy expected utility, have constant absolute risk aversion, and are linear in each of the goods for two or more goods.

## 2.1 Setup and Definitions

**Domain of Choice.** An agent faces  $i \geq 1$  simultaneous choices from choice sets  $S_i$ . We will assume that each  $S_i$  has at most  $N$  options for some fixed but finite  $N$ . Each option  $x \in S_i$  is a bounded, random amount of  $n + 1$  goods, the first  $n \geq 0$  goods are consumption goods and the  $n + 1$ st good is money.<sup>2</sup> So each  $x$  is a bounded random variable over  $\mathbb{R}^{n+1}$ , which we denote by  $\mathbb{X}$ .

**Value function.** Let  $S$  be the set of possible combined outcomes, i.e.,  $S = \sum_i S_i$ . When choosing directly over final outcomes, the agent's choices are derived by maximizing the value function — or decision utility —  $v(x|S)$  over final outcomes  $x \in S$ . Thus under the assumption that the agent makes no mistakes in single choices, this assumes that an agent's utility depends only on the final outcome  $x$  and on the set of possible alternatives  $S$ , but not on whether this was achieved in a single or multiple choices. We assume that  $v(\cdot|S)$  is continuous in  $x$  and  $S$ .<sup>3</sup> With some abuse of notation, we write  $m$  (for *money*) to denote bundles consisting only of sure amounts of money with  $m \in \mathbb{R}$ . We further assume monotonicity in money-only choices and non-satiation in money. Monotonicity in money-only choices implies that when  $S = \{m_1, \dots, m_n\}$  is a choice between  $n$  monetary amounts, then  $m_i > m_j \iff v(m_i|S) > v(m_j|S)$ . Non-satiation in money holds if adding or subtracting sufficiently large sums of money from an option makes it eventually the only chosen option, or not chosen. Formally, for any  $S$  and any  $X \in S$ , let  $S_m \equiv S \setminus \{X\} \cup \{X - m\}$  be the set where  $X$  is replaced by  $X - m$ . Then there is some  $m_B$  such that only  $X - m$  is chosen from  $S_m$  for all  $m < m_B$  and some  $m_S$  such that  $X - m$  is never chosen for any  $m > m_S$ .<sup>4</sup>

**Narrow and Broad Bracketing.** Consider a person who makes two simultaneous choices from  $S_1$  and  $S_2$ .<sup>5</sup> We say that a person *brackets narrowly* if they maximize  $v$  in  $S_1$  and  $S_2$  separately. So they choose  $x_i^N$  from  $S_i$  satisfying  $x_i^N \in \arg \max_{x \in S_i} v(x|S_i)$ , and thus end up with a combined bundle  $z^N = x_1^N + x_2^N$ . Similarly, we say that a person *brackets broadly* if they maximize  $v$  over the combined choice set  $S = S_1 + S_2$ : they pick  $z^B$  satisfying  $z^B \in \arg \max_{z \in S} v(z|S)$ .

<sup>2</sup>Any good that satisfies the conditions we state in the next paragraph can serve as money.

<sup>3</sup>Formally, let  $S_0 = S \setminus \{x\}$ , then  $v(x|S_0 \cup \{x\})$  is continuous in  $x$  and  $S_0$ .

<sup>4</sup>Note that this allows for local non-monotonicity in money, say, due to strong decoy effects and thus allows for a wide range of effects driven by context or comparison set effects.

<sup>5</sup>Our definition straightforwardly generalizes to any  $i > 1$ .

## 2.2 Arbitraging Narrow Bracketers

**Buying and Selling Prices** Consider some bundle  $X$ . We will now define buying and selling prices for  $X$  in different choice contexts. Take some  $Y$  and  $X$  and some  $S$  not containing  $Y$ . Consider a person who chooses from the choice set  $S_m = S \cup \{Y, Y + X - m\}$ . Then if the person chooses option  $Y + X - m$  even though option  $Y$  was available, it means that they were willing to give up an amount of money  $m$  in order to change their outcome by  $X$ . Thus the person is willing to pay the price  $m$  to buy  $X$  on top of  $Y$ . We then define the highest such price  $m$  as the *buying price*  $P_B(X|Y) \equiv \sup\{m : \exists S \text{ with } \{Y\} \cap S = \emptyset \text{ s.t. } v(Y + X - m|S_m) \geq \max_{Z \in S \cup \{Y\}} v(Z|S_m)\}$ . Similarly, if the person chooses option  $Y$  even though option  $Y + X - m$  was available, it means that they were willing to give up getting  $X$  on top of  $Y$  in order to receive the price  $m$  — to receive an extra amount of money  $m$ . In other words, they are willing to sell  $X$  on top of  $Y$  for the price  $m$ . We define the lowest such price  $m$  as the *selling price*  $P_S(X|Y) \equiv \inf\{m : \exists S \text{ with } \{Y\} \cap S = \emptyset \text{ s.t. } v(Y|S_m) \geq \max_{Z \in S_m} v(Z|S_m)\}$ . Note that the set of buying and selling prices for  $X$  on top of  $Y$  is never empty: when choosing only between  $X$  and  $Y$ , by non-satiation the person always chooses  $X - m$  for sufficiently low  $m$ , and never chooses  $X - m$  — and hence chooses  $Y$  — for sufficiently large  $m$ .

Let us now define  $\Delta$ , the largest price differential between the buying and the selling price for any good:  $\Delta \equiv \sup_{X, Y, Y'} P_B(X|Y) - P_S(X|Y')$ .<sup>6</sup> Our first result shows that when the buying price of a good ever strictly exceeds its selling price, then a narrow bracketer can be offered choices such that they throw away an amount of money arbitrarily close to price difference. All proofs are in the appendix.

**Proposition 1.** *If  $\Delta > 0$ , then for every  $\delta \in [0, \Delta)$  there exist choice sets  $S_1$  and  $S_2$  s.t. a narrow bracketer ends up with a final outcome  $A - \delta$  even though  $A \in S_1 + S_2$ .*

The intuition for the result is straightforward: if the person is willing to buy  $X$  for a strictly larger price on top of  $Y$  than she is willing to sell  $X$  for when it is on top of  $Y'$ , then a narrow bracketer facing these two simultaneous choices will buy  $X$  for the high price in the first choice, sell it for the low price in the second choice, and thus make a loss equal their difference.

So when  $\Delta > 0$ , there are choices where a narrow bracketer makes a dominated choice throwing away an amount of money up to size  $\Delta$ . Assuming monotonicity in money, a broad bracketer would never choose that way, hence offering a person these choices allows us to identify whether the person is a narrow or a broad bracketer.

The next proposition shows that when  $\Delta = 0$ , so that every bundle  $X$  has a unique price, then this price function is an additive utility representation for the choices of this person.

**Proposition 2.** *If  $\Delta = 0$ , then choices satisfy WARP and there is a utility representation  $P(X)$  for the choices with  $P(X) = P_B(X|0)$ , and this utility representation is additive, satisfying  $P(X + Y) = P(X) + P(Y)$ .*

---

<sup>6</sup> $\Delta$  could be infinite.

Intuitively, if across all possible choices, the person is willing to buy  $X$  for  $m$  or less, and sells for  $m$  or more, then the value of  $X$  across all choice settings is equal to  $m$ . In particular, this is the price of  $X$  when the alternative is 0. Moreover, since the value of  $X$  is in some sense independent of the choice where it is bought or sold, the price of  $X + Y$  is the price of  $X$  plus the price of  $Y$ .

It is easy to see that a person who maximizes an additive utility function such as  $P(\cdot)$  will choose the same way whether they are bracketing narrowly or broadly, since  $\max_{X,Y} P(X + Y) = \max_{X,Y} P(X) + P(Y) = \max_X P(X) + \max_Y P(Y)$ . Therefore in this case we cannot identify whether the person is a narrow or a broad bracketer. Hence, when  $\Delta = 0$ , we cannot distinguish between narrow and broad bracketing, while Proposition 1 shows that when  $\Delta > 0$  we can, since a narrow bracketer can be arbitrated. This proves the following result:

**Proposition 3.** *We can distinguish between narrow and broad bracketing — narrow and broad bracketing are identified — if and only if  $\bar{\Delta} > 0$ .*

By Proposition 2, we know that  $\Delta > 0$  requires that choices violate either WARP or the additivity of the utility representation  $P(\cdot)$ . We now illustrate how to look for potential price differentials, since these guarantee that we can identify bracketing.

## 2.3 Discussion

Propositions 1, 2, and 3 together show that the central error committed by a narrow bracketer is that they neglect price differentials, and thus arbitrage opportunities, across choices. These prices for any bundle  $X$  are measured in the amount of some numeraire good — “money” — they are willing to give up to obtain the additional bundle  $X$ . In other words, it is the rate of substitution between  $X$  and the numeraire good. Thus, as long as the rate of substitution is constant for any pair of goods across all choices, there are no price differentials and hence the narrow bracketer behaves like a broad bracketer (Proposition 2). When it varies, then a narrow bracketer fails to arbitrage across choices, and makes choices such that they leave money on the table (Proposition 1), which, under the assumption that people never leave money on the table in single choices, implies that we can distinguish between narrow and broad bracketing (Proposition 3).

One consequence of this is that when a person’s observed choices are consistent with an additive utility representation as in Proposition 2, then we cannot identify from those choices whether the person is bracketing narrowly or broadly. Thus when we find that one group of people is less consistent with broad bracketing than another, we cannot infer that this group is more likely to narrowly bracket - an alternative explanation is that their choices are consistent with both.

Another point worth highlighting is that “money” is defined by its properties in Subsection 2.1. This means that any good that satisfies the necessary monotonicity properties can serve as “money”. For example, when considering



repeated choices about how to spend one's time, an increase in the time spent with others rather than alone as well as the probability of achieving one's life goal may both serve as numeraire. Whether this is the case or not may differ from one person to another, since whether something is a good satisfying the monotonicity properties we require depends on preferences (or the assumptions on preferences we are willing to make). With such numeraires, narrow bracketing might lead people to end up in situations where they are strictly worse off, spending more time alone or lowering their likelihood of achieving a life goal more than was possible.

**Three sources of price differentials** As we show in Proposition 4 in Appendix A.1, additivity of  $P(\cdot)$  implies (among other things) that the choices satisfy expected utility and are linear in the amount of work done. Here we illustrate how violating one of these or WARP leads to price differentials. In our experiments, we use these price differentials as a basis to identify whether people bracket these price differentials narrowly or broadly.

First, consider a person who has a disutility from work equal to  $d(e) = 10 \cdot e^2/2$  and whose total utility from working  $e$  hours and earning  $\$m$  is  $m - d(e)$ . Then this person chooses to do 1 hour of work over 0 hours of work if they are paid at least  $\$5$  more, and this person chooses to do 2 hours of work over 1 hour of work if they are paid at least  $\$15$  more. Thus, the price differential across these two situations is  $15 - 5 = 10$ .

Now consider a person who faces choices over lotteries: they either have to do no work at all, or they have to work for 2 hours. Suppose that this person chooses to receive  $\$10$  for sure for doing 2 hours when a die rolls a 1 or a 2 rather than doing no work for no money, and chooses no money for sure and doing 2 hours when a die rolls 3 or 4 over  $\$5$  for sure for doing 2 hours of work when a die rolls 3, 4, 5, or 6. Thus they choose to take on a  $1/3$  increase in probability of doing 2 hours of work for  $\$10$  in the first case, but reject it for  $\$5$  in the second, and thus the price differential is  $\$5$ .

Finally, consider a person who violates WARP as follows: when choosing between doing 2 hours of work for  $\$19$  and doing no work, they choose no work, as would be in line with the disutility above. However, when choosing between 2 hours of work for  $\$19$ , doing no work, and doing 2 hours of work for  $\$10$ , they choose to do 2 hours of work for  $\$19$ . Moreover, suppose they continue choosing 2 hours of work for as little as  $\$15$ . Then the price differential is at least  $\$4 = 19 - 15$ .

Hence, in each of these cases, we know by Proposition 1 that we can use these price differentials to construct choices to test bracketing via arbitrage.

**Source of price differentials in the Literature** We now review a small subset of the literature that tests bracketing and for each identify potential price differentials underlying their identification.

Tversky and Kahneman (1981) identify choices in which people's choices are consistent with risk-aversion in gains and risk-seeking in losses, so that they reject a given lottery when it changes the gains, yet they accept it in the loss domain. In our terminology, they sell the lottery when it is offered in the gains domain and they buy it in the loss

domain. As Rabin and Weizsäcker (2009) point out, it is sufficient if there is a *change* in the degree of risk aversion, which leads to a price differential because this changes their willingness to buy the lottery.

Vorjohann (2020) defines narrow bracketing as correlation neglect. To give a stark example, consider a person who may not like risky assets, but faces two assets, one of which always goes down when the other goes up. In such a situation, a person who can only buy a single asset might be willing to pay only a low price due to the potential of losses, yet when they can buy both and hedge against losses, their willingness to pay rises. Thus, whenever correlation affects prices, narrow bracketing implies correlation neglect, although it is not equivalent to it in more general settings.

Ellis and Freeman (2024) study choices across multiple domains, including portfolio choices with similar sources of price differentials as Tversky and Kahneman (1981) and Rabin and Weizsäcker (2009). Additionally, they also test bracketing over consumer problems with induced values that are non-linear; as well as over divisions of money between other (anonymous) subjects. In the latter case, if a person is inequity averse, then they may choose an equal split of (5, 5) over a split of (7, 4), as well as over a split of (4.5, 6). Therefore, the price of \$1 to person 2 equals \$2 from person 1 in the first choice, and \$0.5 in the second, leading to price differentials. Note that a person who displays efficiency motives, such as maximizing the total amount of money that the two subjects receive, would maximize an additive function. With such preferences, we would not be able to test whether this person is bracketing narrowly or broadly.

**Other sources of price variation** Our result that bracketing is unidentified if and only if there exists an additive utility representation denoted  $P(\cdot)$  (for *price*) generalizes to richer choice settings. For instance, suppose that a person chooses how much time to spend on each of two activities, and is (exogenously) required to choose the total amount for activity 1 and separately the total amount for activity 2. If we assume that their utility from working  $w_1$  hours on the first and  $w_2$  hours on the second activity is  $u(w_1, w_2) - (w_1 + w_2)$ ,<sup>7</sup> then we can show that bracketing is unidentified if there is some utility  $P(\cdot)$  that is additive *over the set of choices we could possibly observe*. This means that  $P(X + Y) = P(X) + P(Y)$  has to hold for all  $X$  which affect only time spent on activity 1 but not activity 2, and all  $Y$  that affect activity 2 but not 1. It is clear that any additively separable utility function  $u(w_1, w_2) = u_1(w_1) + u_2(w_2)$  satisfies this condition, where the  $u_i$  need not be linear; and that if it is not additively separable, then a narrow bracketer behaves differently in some choices.<sup>8</sup> This highlights both the key role of additivity, as well as the necessary variation in order to be able to identify bracketing.

<sup>7</sup>For simplicity we assume in this example that all remaining time can be used in a linear way for leisure. This can be generalized.

<sup>8</sup>Given that we do not allow more than 1 choice per activity, we cannot arbitrage, since this would require one choice where we sell and one choice where we buy the same good.

### 3 Study 1: Testing for bracketing of work choices

In the following sections, we describe the first pre-registered online experiments, written in Lioness (Giamattei et al., 2020). Study 1 (conducted between December 2019 and January 2020) tests the effects of narrow and broad bracketing on the willingness to work in a real effort choice setting.<sup>9</sup> The study consists of four parts. We start by describing the real effort task and the main parts and treatments of Study 1 (detailed instructions are provided in Appendix B).

#### 3.1 Study 1: Design

**Part 1 Tutorial** The study starts with a tutorial phase where participants begin by familiarizing themselves with the real-effort task through practice trials, continuing until they correctly complete three tasks. Each task involves decoding a sequence of twelve letters into their corresponding numerical values, as illustrated in Figure 1.

**Encryption task**

Decode sequences into numbers until you got 10 right.

Correctly decoded text sequences: 0

**Decoding table**

letter	u	t	f	c	g	n	v	q	a	l
corresponds to number	0	1	2	3	4	5	6	7	8	9

Text sequence: tvqqnqvfgug

Your answer (enter only numbers):

167757642404

Submit

Figure 1: Decoding task

For each sequence, participants receive a newly generated table that randomly assigns ten letters to the digits 0 through 9. This mapping is re-created after each attempt, regardless of whether the previous response was correct. By

<sup>9</sup>See <https://doi.org/10.1257/rct.3412-4.4999999999999999>, in particular the 'December Design' under *Supporting Documents and Materials*.

continuously changing letter-to-number assignments, we introduce a greater learning challenge compared to conventional encryption or typing tasks (e.g., Erkal et al., 2011; De Quidt et al., 2017; De Quidt, 2018). This design increases the convexity of the effort costs, a key mechanism for satisfying our identifiability assumptions.

**Part 2: Elicit Tediousness** After completing the tutorial, participants rate the perceived tediousness of the task on a scale from 1 ("not tedious at all") to 10 ("extremely tedious"). This provides a control variable that is consistent across all treatment conditions.

**Part 3: Elicit Reservation Wages** We elicit participants' reservation wage for a high-work option that requires completing 15 additional tasks compared to the low-work option, with the specific conditions varying by treatment. The four treatments (detailed below) differ in how participants' initial endowment of tasks and money is structured and presented.

Participants make decisions in two sequential choice scenarios, Scenario 1 and Scenario 2. In each scenario, we elicit a reservation wage using an incentivized price list, where participants decide whether to accept or reject the additional workload for a series of incremental *extra* wages ranging from \$0.25 to \$4.00 in increments of \$0.25. These extra wages are offered on top of the \$4.00 baseline payment for the alternative workload.<sup>10</sup>

The framing of payments varies by treatment (see description of Main Treatments). Participants are informed that one binding decision from either scenario will be randomly selected and implemented.

**Part 4: Complete Tasks** We randomly determine the binding choice and inform subjects about the total payment and number of sequences to decode. Subjects can complete the tasks without time constraints. We then ask for a short demographic questionnaire and display a summary of total earnings.

## 3.2 Treatments

In our treatments, participants made a single active choice in each scenario. The choice combined an endowment of work and/or money. This structure, often referred to as *endowment bracketing*,<sup>11</sup> is most studied in the context of background risk, where individuals fail to integrate it with other risky choices. In this regard, Barberis et al. (2006) and Mu et al. (2021a) highlight that ignoring background risk helps explain risk aversion over small stakes.<sup>12</sup>

---

<sup>10</sup>Details on individual choices for each treatment are provided in Appendix B.

<sup>11</sup>See (Koch and Nafziger (2019), Ellis and Freeman (2024)).

<sup>12</sup>Failure to fully incorporate endowments can lead individuals to treat money and goods as less fungible than they are, linking endowment bracketing to mental budgeting (Heath and Soll, 1996; Hastings and Shapiro, 2013; Abeler and Marklein, 2017). Furthermore, as shown in Imas (2016), the behavior of individuals differs between real and paper losses, suggesting that factors beyond bracketing influence the role of endowments in decision-making.

We report three between-subjects treatments:<sup>13</sup> NONE, where there is no endowment; MONEY/LOW, where the endowment consists of \$2.00; and BOTH, where the endowment consists of \$2.00 and 15 sequences to decode. NONE and BOTH have identical total choice sets: the total amount of money and work possible is the same for these treatments. The choice of additional work and money is identical in BOTH and MONEY/LOW, but since MONEY/LOW has no endowment of work, each option in MONEY/LOW leads to exactly 15 fewer sequences than in BOTH – hence MONEY/LOW is the only treatment with different total outcomes.

We show below the text displayed to participants for a single choice of the price list for Scenario 1 for each treatment, while we refer to Appendix B for the screens of the choice list items. Scenario-2 choices require exactly 15 sequences more than the Scenario-1 choices of the same treatment. Participants have to choose between a list of Option A and Option B choices, with payments for Option B (in addition to the completion fee) going from \$4.25 to \$8 in steps of \$0.25.

---

<sup>13</sup>We have several additional treatments that were relevant for our earlier paper titled "Narrow Bracketing in Work Choices" and which we deemphasize in the interest of space after adding study 2. These treatments are MONEY, BEFORE, and AFTER. MONEY is described in more detail in Appendix C.10; BEFORE and AFTER are described in Appendix E.1.

**NONE: No Endowment**

**By completing the HIT you will receive a total payment (which includes the \$2.00 completion fee) depending on your choices.**

	OPTION A	OPTION B
5) 15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/> 30 sequences for a total payment of \$7.25

**BOTH: Both Money and Work Endowment**

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

	OPTION A	OPTION B
5) 0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences for an extra \$5.25

**MONEY/LOW: Money but no Work Endowment**

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

	OPTION A	OPTION B
5) 0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/> 15 sequences for an extra of \$5.25

### 3.3 Bracketing Hypotheses

Note that the treatments NONE and BOTH have *identical* total outcomes in the same scenario, but with 15 sequences and \$2.00 shifted to the endowment for BOTH. Hence broad bracketing predicts identical choices over *total* outcomes across them.<sup>14</sup> MONEY/LOW on the other hand has identical Options in the choice list as BOTH, even though it has no work endowment and hence require 15 fewer tasks. So the choice options *ignoring* endowments are identical for MONEY/LOW and BOTH, so that narrow bracketing predicts identical choices of options – but not of total outcomes. Denoting by  $m_T$  the average reservation price elicited in treatment  $T$ , we get the following hypotheses:<sup>15</sup>

**Hypothesis 1** (Broad Bracketing). *Behavior is consistent with broad bracketing if  $m_{NONE} = m_{BOTH}$  in every Scenario.*

**Hypothesis 2** (Narrow Bracketing). *Behavior is consistent with narrow bracketing if  $m_{BOTH} = m_{MONEY/LOW}$  in every Scenario.*

As we discussed in Section 2, we cannot always identify bracketing (that is, distinguish narrow from broad bracketing). In the case of our experiment, we need that  $m_{MONEY/LOW} \neq m_{NONE}$ : if  $m_{MONEY/LOW} = m_{NONE}$ , then choices are consistent with linear preferences, so that we will either fail to reject both narrow and broad bracketing or reject both simultaneously. This will happen if participants are as willing to do 15 additional sequences on top of 0 sequences, as they are willing to do them on top of 15 sequences.

**Assumption 1** (Identification Assumption). *We can identify narrow vs broad bracketing if and only if  $m_{NONE} \neq m_{MONEY/LOW}$ .*

Since we have two Scenarios, we have two tests for broad and two tests for narrow bracketing. For tests with 5% significance, we would therefore apply the Bonferroni correction of rejecting the null hypothesis in a given Scenario only if  $m_{NONE}$  differs from  $m_{BOTH}$  ( $H_0$  : broad bracketing) or from  $m_{MONEY/LOW}$  ( $H_0$  : narrow bracketing) at the 2.5%-level, to avoid overrejection based on having two tests.

### 3.4 Study 1: Summary Statistics

We recruited in total 716 subjects on Amazon Mechanical Turk between the end of December 2019 and the beginning of February 2020. Table 8 in Appendix C shows how many participants and in which treatment we recruited by session. In Table 1 we report summary statistics across treatments. Of the subjects recruited, 127 did not complete the experiment (17.7% attrition rate). Although differential attrition across treatments could itself hint at the consequences

<sup>14</sup>In a similar design in online auctions on eBay, Hossain and Morgan (2006) find higher revenues and number of bidders when the starting price is reduced by the same amount that the shipping costs are increased. This is consistent with participants ignoring shipping costs (partially), similar to ignoring endowments in our setting.

<sup>15</sup>One can formally relate these hypotheses to our theoretical framework by noting that situations of a single choice with a given endowment correspond to situations where the second choice set is a singleton:  $\mathcal{Y} = \{Y\}$ .

of narrow bracketing, we see no evidence for it (see Table 9 in Appendix C). Across all treatments, between half and two-thirds of participants failing to complete the HIT drop before completing the practice tasks, while the remaining others drop out after finding out how many tasks they have to do in total. Treatments are similar in terms of gender composition ( $\chi^2$  test p-value: 0.84), while participants are slightly older in the NONE treatment compared to other treatments (37.8 years vs 35.0-35.1,  $\chi^2$  test p-value: 0.29). Finally, individuals rate the task on average as 7.33-7.54 out of 10 in tediousness, which does not significantly vary across treatments ( $\chi^2$  test p-value: 0.86). Roughly 18% of the choices made within a choice-list are inconsistent: in a few cases subjects make only one inconsistent choice, while in other cases choices are inconsistent throughout the list of wages offered (such as when they switch repeatedly between the options, even though we monotonically increase payment). In our main analysis we drop a scenario if individuals make more than one inconsistent choices in it. To detect an effect size of 0.40 at a 5% level of significance with 90% power, we would need 174 observations in BOTH and 116 in MONEY/LOW and NONE treatments. The number of observations collected with consistent choices are above these thresholds and therefore considered sufficient for our treatments comparisons, although as the discussion on identification makes clear, the effect size decreases as preferences become more linear. Participants earned \$7.30 on average, for an average working time of 35 minutes.<sup>16</sup>

Table 1: Summary statistics for main treatments.

	NONE	BOTH	MONEY/LOW	p-value
Participants	200	320	196	
Attrition	18%	20.3%	13.3%	0.13
Final Participants	164	255	170	
Share Female	0.4	0.38	0.4	0.84
Age	37.8	35	35.1	0.29
Tediousness	7.54	7.45	7.33	0.86
<b>Inconsistent Choices</b>				
Scenario 1	17%	18.8%	20.4%	0.69
Scenario 2	15%	18.4%	18.9%	0.52

## 4 Study 1: Results

We now analyse the reservation wages across treatments: the smallest extra wage for which subjects prefer OPTION B over OPTION A, where OPTION B always requires decoding correctly 15 sequences more than OPTION A. The

<sup>16</sup>Subjects' feedback rated on average this payment as generous. For details, see <https://turkerview.com/requesters/A3TEY5GKYRHXXWG>.



extra payments start at \$0.25, if a subject always accepts the extra work, and increases in \$0.25 increments to \$4.00. If a subject never accepts the extra work, we code the reservation wage as \$4.25.

## 4.1 Main Results

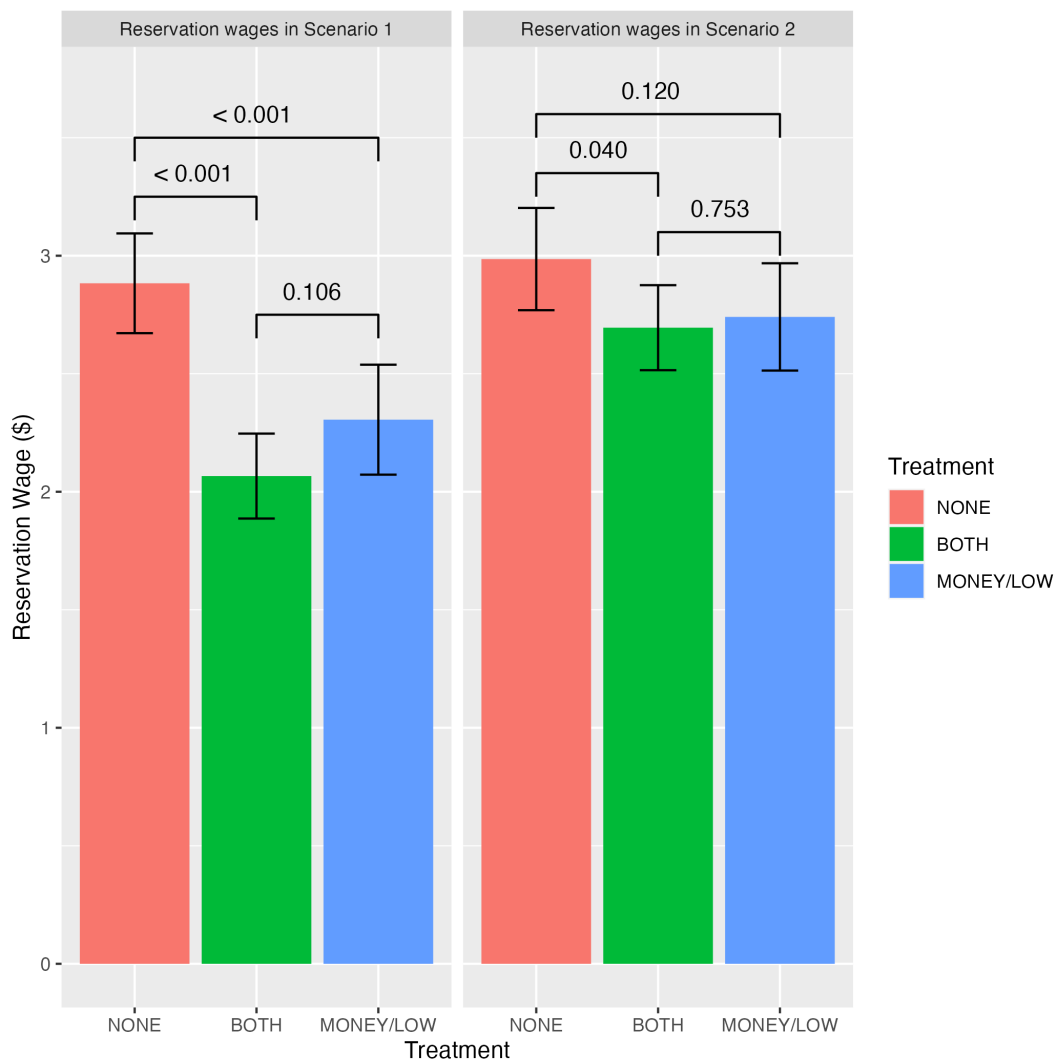


Figure 2: The reservation wages by treatment for each of the two scenarios, along with confidence intervals (2 standard errors above and below the estimate). The p-values compare average reservation wages between two treatments via two-sided t-tests.

**Result 1.** *We reject Hypothesis 1 that individuals bracket decisions broadly.*

Based on figure 2, we reject broad bracketing as per Hypothesis 1, since we reject broad bracketing in one of the scenarios (Scenario 1) at the Bonferroni-corrected p-value of less than 2.5%-significance (see our discussion in 5.2).<sup>17</sup>

Concretely, the average extra reservation wage in treatment NONE in Scenario 1 is \$2.88 compared to \$2.07 in BOTH,

<sup>17</sup>See C.7 in Appendix C for bar plots and kernel density plots of the raw reservation wage data by treatment and scenario.

and this difference is significant at the 0.001-level. Note that identifying assumption 1 holds in Scenario 1, since the reservation wage in NONE and in MONEY/LOW are not equal.<sup>18</sup>

Notice that the Scenario-2 reservation wage of \$2.74 in MONEY/LOW and of \$2.88 in NONE are not statistically significantly different (p-value of 0.120). Thus identifying assumption 1 fails in this Scenario, and consequently we fail to reject broad bracketing in this scenario. We also fail to reject narrow bracketing in this scenario: the reservation wage of \$2.70 for BOTH is not statistically significantly different from MONEY/LOW. Scenario 2 lacks power to identify bracketing, because we cannot rule out constant marginal disutility at 15 and 30 baseline sequences.

Since the difference in Scenario 1 between the reservation wage of \$2.31 in MONEY/LOW and of \$2.07 in BOTH is not significant (p-value: 0.106), we cannot reject narrow bracketing in either Scenario and hence do not reject it overall. Since the identification assumption holds in Scenario 1, the identification assumption holds, so the failure to reject it does not stem from a failure of the identification assumption to hold.

**Result 2.** *We fail to reject Hypothesis 2 that individuals bracket decisions narrowly.*

We perform similar comparisons between the treatments via Wilcoxon rank-sum tests as well as when restricting to the balanced sessions only, both of which lead to identical conclusions – see Appendix C for details.<sup>19</sup>

Overall, our results can be summarized by saying that participants average behavior is consistent with a disutility function  $d(\cdot)$  satisfying  $d(15) \approx 2.3$  (based on Scenario 1, MONEY/LOW), with  $d(30) - d(15) \approx 2.8$  (based on Scenario 1, NONE; and Scenario 2, MONEY/LOW) and with  $d(45) - d(30) \approx 3.00$  (based on Scenario 2, NONE); and that participants narrowly bracket these preferences. The disutility of effort is thus convex, and growing less convex with additional tasks.<sup>20</sup>

## 5 Study 2: Testing for choice bracketing across domains

Following our theory, Study 2 aims to test for narrow and broad bracketing by generating variation in prices in three different domains in a within-subjects design.<sup>21</sup>

<sup>18</sup>Since we keep the choice sets exactly equal across treatments, we ensure that differences in behavior are not because of (non-bracketing) interactions with work they do outside of our experiment. For example, if participants could earn less in some treatments, they might decide to spend less time on our experiment and work more on other tasks on MTurk instead. If our participants did that, it would by itself be a sign that they failed to realize that their overall outcomes are actually the same.

<sup>19</sup>In addition to using the Wilcoxon rank-sum tests, we also perform tests for different sessions of BOTH. Specifically, in the initial sessions, we mistakenly displayed the endowments on the page right before the first choice page. We fixed this, displaying it on the first choice page only, which is why we collected more data for treatment BOTH. See E.2 and C.5 for the same results when we restrict treatment BOTH to when we display the endowments on the first choice page only, or when we display it right before the first choice page. See C.4 for when we restrict the treatments to those sessions in which data collection was balanced – since some sessions were not balanced. In all these cases, we reject broad bracketing in Scenario 1 at the 2.5%-level and hence overall, and do not reject narrow bracketing in either Scenario 1 or Scenario 2.

<sup>20</sup>We report in Appendix E the description and results of a follow-up study where we attempt to debias subjects. However, the results confirm the effect of narrow bracketing in work choices.

<sup>21</sup>The design of Study 2 was preregistered in OSF: <https://osf.io/5q7wn/>.

## 5.1 Design

Study 2 follows the structure of Study 1, with the main differences occurring in Part 3 where we elicit willingness to work. After assessing perceived tediousness, subjects received instructions on the choices to be made and answered two comprehension multiple choice questions.<sup>22</sup> Only those who correctly answered advanced to the decision stage, where they responded to 35 scenarios that involved different work options at varying pay levels.<sup>23</sup> Following our theoretical framework, these scenarios were aimed at identifying participants in three domains that we label 'Linear', 'WARP', and 'Probability', where we attempted to find price variation by finding violations of linear utility, of WARP, and of expected utility respectively.

We now describe how we attempted for each domain attempt to find a pair of choices with a buying price in one choice that is higher than the selling price in another choice, so as to use these choices to arbitrage the participant.

**Linear** Identification in the linear domain was performed in two steps. First, we elicited the maximum number of sequences participants were willing to decode (up to 25) at five different piece rates, ranging from £0.02 to £0.12 per sequence. The participants then responded to follow-up questions adapted from prior responses, presenting choices between their previous selection and an alternative involving five more (or fewer) sequences at the same piece rate. Example:

**Scenario 1: Choose your preferred option:**

- Option A: Decode 10 sequences for £1.20.
- Option B: Decode 15 sequences for £1.80.

**Scenario 2: Choose your preferred option:**

- Option C: Decode 10 sequences for £0.80.
- Option D: Decode 5 sequences for £0.40.

Participants are identified in this domain if we find two scenarios such that they refuse to decode more sequences at a higher price (e.g., choosing Option A in Scenario 1 and thus refuse 5 tasks for £0.60) and accept to decode more sequences at a lower price (e.g., choosing Option C in Scenario 2 and thus accept 5 tasks for £0.40).

**WARP** WARP violations were assessed by presenting choices between tasks with different combinations of effort and payment. In some of the choices (as in Scenario 4) we included a third option (a decoy) to influence the attractiveness of the other item.<sup>24</sup> Example:

<sup>22</sup>The questions are reported in Appendix D. These questions were crucial to understanding the experimental incentives.

<sup>23</sup>See Appendix D for the instructions of the main task as well as the individual Scenarios in study 2.

<sup>24</sup>Following Kaptein et al. (2016) we tested for different decoys as well as choices presented in different orders.

**Scenario 3: Choose your preferred option:**

- Option A: Decode 5 sequences for £0.45.
- Option B: Decode 12 sequences for £0.80.

**Scenario 4: Choose your preferred option:**

- Option C: Decode 5 sequences for £0.45.
- Option D: Decode 12 sequences for £0.75.
- Option E: Decode 12 sequences for £0.70.

Participants were identified in this setting if they selected Option A in Scenario 3 but chose Option D in Scenario 4. These choices are consistent with a WARP violation.

**Probability** In the probability domain, we identify subjects presenting choices involving probabilistic work obligations and sure payments. Example:

**Scenario 5: Choose your preferred option:**

- Option A: Receive £0.50 for sure. If the six-sided die rolls 5 or 6, decode 20 sequences; otherwise, decode 0 sequences.
- Option B: Decode 0 sequences for £0.25.

**Scenario 6: Choose your preferred option:**

- Option C: Receive £0.70 for sure. If the six-sided die rolls 1, 2, 3, or 4, decode 20 sequences; otherwise, decode 0 sequences.
- Option D: Receive £0.50 for sure. If the six-sided die rolls 3 or 4, decode 20 sequences; otherwise, decode 0 sequences.

Participants are identified in the probability domain if they choose Option B in Scenario 5 and thus reject receiving an extra doing 20 sequences if a 5 or a 6 is rolled and Option C in Scenario 6, that is, rejecting an increase in the probability of decoding 20 sequences for £0.25, but accepting the same increase in probability for £0.20.

**Arbitrage** After the identification phase, subjects completed scenarios that tested for narrow and broad bracketing, following our arbitrage approach. In *Simultaneous* scenarios we displayed the two scenarios on a single screen. In the instructions we specify that when two decisions are shown on the same Scenario page, they are both relevant for the amount of work and payment.<sup>25</sup> We show below the *Simultaneous* version for Scenarios 1 and 2.<sup>26</sup>

<sup>25</sup>To ensure that participants understand this instructions, and in particular the way the amount of work and bonuses are defined, we ask them to answer two comprehension questions before proceeding to the decision part of the experiment. Subjects who fail to correctly answer the two questions are excluded from the participation in the experiment. The questions are shown in Appendix D.5.

<sup>26</sup>See Appendix D.7 for a screenshot of the Scenario implemented.

### Scenario 7 [SIMULTANEOUS LINEAR]

In this Scenario, you have to make two decisions.

#### Decision 1: Choose your preferred option:

- Option A: Decode 10 sequences for £1.20.
- Option B: Decode 15 sequences for £1.80.

Before answering, read the next decision

#### Decision 2: Choose your preferred option:

- Option A: Decode 10 sequences for £0.80.
- Option B: Decode 15 sequences for £0.40.

*Combined* scenarios combined them into a single choice. We show below an example of the combined linear case (equivalent to the simultaneous Scenarios 1 and 2), combined probability (equivalent to the simultaneous Scenarios 5 and 6) and combined WARP (equivalent to the simultaneous Scenarios 3 and 4):

### Scenario 8 [COMBINED LINEAR]:

#### Choose your preferred option:

- Option A: Decode 25 sequences for £2.60.
- Option B: Decode 20 sequences for £2.20.
- Option C: Decode 20 sequences for £2.00.
- Option D: Decode 15 sequences for £1.60.

### Scenario 9 [COMBINED PROBABILITY]: Choose your preferred option:

- Option A: Receive £1.20 for sure. If the six-sided die rolls 1, 2, 3, 4, 5 or 6, decode 20 sequences; otherwise, decode 0 sequences.
- Option B: Receive £1.00 for sure. If the six-sided die rolls 3, 4, 5 or 6 decode 20 sequences; otherwise, decode 0 sequences.
- Option C: Receive £0.95 for sure. If the six-sided die rolls 1, 2, 3 or 4 decode 20 sequences; otherwise, decode 0 sequences.
- Option D: Receive £0.75 for sure. If the six-sided die rolls 3 or 4 decode 20 sequences; otherwise, decode 0 sequences.

### Scenario 10 [COMBINED WARP]: Choose your preferred option:

- Option A: Decode 10 sequences for £0.90.
- Option B: Decode 17 sequences for £1.20.
- Option C: Decode 17 sequences for £1.15.
- Option D: Decode 17 sequences for £1.25.
- Option E: Decode 24 sequences for £1.55.
- Option F: Decode 24 sequences for £1.50.

The remaining part of study 2 is similar to study 1, with one exception. We have included two questions throughout

the study to check for attention.<sup>27</sup> We use the answers to these questions as a control of attentiveness.

## 5.2 Categorization

We do not anticipate that every participant will be identified in one or more domains. To discern which subjects can be identified within each domain, we adhere to Proposition 3. Conditional on subjects being identified, they will be classified into one of the following three mutually exclusive categories:

**Narrow Bracketers.** Behavior is consistent with narrow bracketing if the decisions in the 'Narrow' scenario are equivalent to the two equivalent choices when they are presented individually.

**Broad Bracketers.** Behavior is consistent with broad bracketing if the overall amount of work and payment in the 'Narrow' Scenario are equal to those in the 'Broad' Scenario.

**Unknown.** Behavior is not consistent with either narrow and broad bracketing if the decisions in the simultaneous (but separate) Scenario differ from the two choices when made individually (rejecting narrow), and when the decisions in the simultaneous Scenario differ from the decision in the combined-choice Scenario (rejecting broad).

While our primary interest is on subjects that are ex-ante identified, those who are unidentified will also answer a set of questions with 'simultaneous' and 'combined' Scenarios. In principle, these subjects can also be classified under the three aforementioned categories. However, unlike the identified subjects, it is likely that many of them will be unidentifiable, and hence be categorized both as narrow and broad bracketers at the same time.

## 5.3 Study 2: Summary statistics

We recruited 572 subjects on Academic Prolific in January 2025. Of these, 59 failed to answer the comprehension questions and were excluded from participation. The remaining 513 subjects comprised 52.05% females and 46.59% males, while 1.36% are non-binary or have not disclosed their gender. Most participants were aged 18-34, with less than 17% over 45. On average, participants rated the task as 6.59 out of 10 in tediousness,<sup>28</sup> earned on average £3.79, and decoded on average 11.09 sequences, with an average study time reported in Prolific of approximately 28 minutes.<sup>29</sup>

---

<sup>27</sup>(see, e.g., Fallucchi and Nosenzo, 2022)

<sup>28</sup>The software failed to record the self-reported tediousness of two subjects.

<sup>29</sup>The payment aligned with the platform standards, exceeding £6 per hour.

## 6 Study 2: Results

We begin by examining the number of participants identified in our experiment as shown in Table 2. Out of 513 subjects, and in each of the three domains approximately 30% are identified. Examining the distribution of the identified subjects, we observe that 39% of the sample is not identified in any domain, 34.52% are identified in a single domain, 22.82% in two domains, and merely 3.57% in all three domains.<sup>30</sup>

Table 2: Summary statistics for study 2.<sup>31</sup>

Participants	572		
Failed Comprehension	59		
Final Participants	513		
	LINEAR	WARP	PROBABILITY
Identified	149	152	147
Share female (%)	54.36	47.37	56.46
Tediousness	6.97	5.92	6.18
Share of inattentive (%)	8.05	9.87	12.24
Unidentified	355	361	366
Share female (%)	51.27	54.02	50.27
Tediousness	6.41	6.87	6.76
Share of inattentive (%)	13.80	13.30	12.30

**Result 3.** *The majority (50.30%) of identified subjects are classified as narrow bracketers in LINEAR. The share of broad bracketers is higher than the share of narrow bracketers in WARP (38.16% vs 26.32%) and PROBABILITY (46.26% vs 13.61%).*

Table 3 shows the classification of individuals according to their decision-making patterns. A participant is categorized as a narrow bracketer when their two working choices under the 'Simultaneous' choice are consistent with those when considered in isolation. This is observed in 50.30% of cases for Linear, 26.32% for WARP, and 13.61% for Probability among identified participants. In contrast, a Broad bracketer is someone whose decisions is the same between the 'Simultaneous' and 'Combined' choices, indicating that the total selected workload and payment are unchanged when the two 'Narrow' choices are aggregated. Broad bracketing occurs more frequently in WARP (38.16%) and Probability (46.26%) compared to Linear (14.57%). Overall, of the 305 individuals identified in at least one domain, 43.28% are classified as narrow bracketers at least once.

We also observe a group of subjects inconsistent with both Narrow and Broad bracketing, which we categorize as 'Unknown' above. These individuals account for 35.29% in WARP, 33.77% in Linear, and 40.14% in Probability.

<sup>30</sup>We examined correlations across domains and found the highest correlation between WARP and Probability at 17%.

Table 3: Share of narrow and broad bracketers among identified subjects.

	LINEAR	WARP	PROBABILITY
<b>Identified</b>	149	152	147
Consistent with Narrow Bracketing	76 (50.33%)	40 (26.32%)	20 (13.61%)
Consistent with Broad Bracketing	22 (14.57%)	58 (38.16%)	68 (46.26%)
Unknown	51 (33.77%)	54 (35.29%)	59 (40.14%)
<b>Unidentified</b>	355	361	366
<b>Total</b>	<b>504</b>	<b>513</b>	<b>513</b>

**Exploratory Analysis** Next, we explore the correlates of Narrow bracketing in the three domains. We investigate potential gender differences in the likelihood of being identified as a narrow bracketer, as well as controlling for the perceived tediousness of the task and whether subjects failed the attention checks during the study. The results of these analyses are presented in Table 4.

Table 4: Likelihood of being a narrow bracketer (robust standard errors in parentheses).

Dependent: narrow bracketer	LINEAR	WARP	PROBABILITY
<b>Female</b>	-0.165* (0.083)	0.046 (0.072)	0.147** (0.057)
<b>Tediousness</b>	0.006 (0.017)	-0.026 (0.014)	0.013 (0.010)
<b>Inattentive</b>	-0.069 (0.153)	0.064 (0.132)	0.042 (0.093)
<b>Constant</b>	0.560*** (0.142)	0.390*** (0.103)	-0.034 (0.071)
<b>Total</b>	<b>149</b>	<b>151</b>	<b>146</b>

Note: Standard errors in parentheses. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ .; \*\*\* $p \leq 0.001$ .

As shown by the regression when we check for gender differences in narrow bracketing, we find contrasting results: in the linear domain, women are significantly less likely to be narrow bracketers than others; in WARP and probability we find an opposite sign, and in probability the likelihood of women to be identified as narrow bracketer is significantly higher.<sup>32</sup> In short, there is no robust evidence for systematic gender differences. We also note that

<sup>32</sup> $\chi^2$  tests confirm similar results:  $\chi^2(1) = 4.318$   $p$ -value = 0.038 in linear,  $\chi^2(1) = 0.449$   $p$ -value = 0.573 in WARP and



perceived tediousness and our measure of attentiveness do not affect the likelihood of being a narrow bracketer in any domain.

Table 5: Narrow and Broad Bracketing among unidentified.

	LINEAR	WARP	PROBABILITY
<b>Total</b>	355	361	366
Consistent with Narrow and Broad Bracketing	233 (66.01%)	181 (50.14%)	199 (54.37%)
Conditional on not being consistent with Narrow and Broad Bracketing			
Consistent with Narrow Bracketing	64 (53.33%)	61 (38.89%)	38 (22.75%)
Consistent with Broad Bracketing	16 (13.33%)	25 (13.89%)	56 (33.33%)
Unknown	40 (33.33%)	94 (52.22%)	73 (43.45%)

Unlike most previous studies, our analysis focuses on the subjects identified following our arbitrage condition. We provide the categorization of subjects that are unidentified in table 5, which shows that the share of subjects that can be categorized as narrow bracketers is much higher, but so is the share of broad bracketers. Notably, at least half of the subjects participating in the study make choices that are consistent with both narrow and broad bracketing in all domains (66% in Linear, 50% in WARP and 54% in Probability). Conditional on not making consistent choices with both narrow and broad bracketing, also for the unidentified subjects we find in Linear the higher share of narrow bracketers; similarly, data confirm the higher share of broad bracketers in Probability; in WARP results differ between identified and unidentified, with a higher share of narrow bracketers in the latter.

## 7 Conclusion

In this paper, we show that narrow bracketing leads to different behavior from broad bracketing if and only if a person values some bundle differently in different situations, in which case they can be arbitrated across two choices buying the bundle for more in one choice than they sell it for in another.

We use this type of arbitrage in our main set of experiments to identify bracketing when the source for the change in payment is due to (i) non-linear disutility from work, (ii) violations of expected utility, or (iii) violations of the weak axiom of revealed preference (WARP). We find similar proportions of narrow bracketers as the previous literature for the first source, both in our between-subjects design and our within-subjects design, but the majority of people either

---

$\chi^2(1) = 4.463$   $p - value = 0.035$  in probability.

brackets broadly or is inconsistent with either narrow or broad bracketing for the latter two sources. One (ex post) hypothesis for this difference is that these types of choices are more easily perceived as connected: in the case of probability choices, we highlight that the outcome of both choices depend on the outcome of the same roll of a die; thus one has to think of this same dice roll when making each choice; in the case of choices with decoys that affect choices, it may be that the mere act of seeing the decoy shifts perceptions for both choices that are displayed and made simultaneously. Varying the source in other ways while testing narrow bracketing may provide a promising avenue for identifying the mechanism behind narrow bracketing, helping us to identify factors of choices that increase or reduce the amount of narrow bracketing. Our paper and experiments are focused on measuring when narrow bracketing occurs, and do not provide direct evidence on the mechanisms.

Finally, the fact that about one third of identified participants are not consistent with either narrow or broad bracketing raises further questions. Several natural reasons present themselves: it may be that people are simply confused or inattentive, although the fact that the proportion is similar for choices over simple work-money trade-offs as well as for lotteries suggests otherwise. Alternatively, people may randomize their choices as we know they do even for identical but repeated choices (Agranov and Ortoleva (2017)), a possibility that we do not account for. Finally, participants may genuinely change their decisions in one choice when facing other choices unlike narrow bracketers, but in a way that differs from how a broad bracketer would do it. Modeling such partial bracketing and disentangling alternative theories requires a different design from our experiments, but will hopefully benefit from the insights and approach outlined in this paper.

## 8 Acknowledgements

We thank David Freeman, Gergő Hajdu, Holger Herz, Mats Köster, Botond Köszegi, Tomy Lee, Robert Lieli, Sergey Lychagin, Ariada Muço, Daniele Nosenzo, Ádám Szeidl, Georg Weizsäcker, seminar participants at CEU, University of Portsmouth, University of Trento, BEEN 2020 (University of Verona), ESA 2020, ASFEE 2021 (Dijon), YEM 2022 (Brno), ASFEE 2022 (Lyon), and FUR 2022 (Ghent) conferences for helpful comments.

## References

- Johannes Abeler and Felix Marklein. Fungibility, labels, and consumption. *Journal of the European Economic Association*, 15(1):99–127, 2017.
- Marina Agranov and Pietro Ortoleva. Stochastic Choice and Preferences for Randomization. *Journal of Political Economy*, 125(1):40–68, February 2017. ISSN 0022-3808. doi: 10.1086/689774. URL <https://www.journals.uchicago.edu/doi/full/10.1086/689774>. Publisher: The University of Chicago Press.
- James Andreoni, Christina Gravert, Michael Kuhn, Silvia Saccardo, and Yang Yang. Arbitrage Or Narrow Bracketing? On Using Money to Measure Intertemporal Preferences. Technical Report w25232, National Bureau of Economic Research, Cambridge, MA, November 2018. URL <http://www.nber.org/papers/w25232.pdf>.
- Nicholas Barberis, Ming Huang, and Richard H Thaler. Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American economic review*, 96(4):1069–1090, 2006.
- Jonathan De Quidt. Your loss is my gain: a recruitment experiment with framed incentives. *Journal of the European Economic Association*, 16(2):522–559, 2018.
- Jonathan De Quidt, Francesco Fallucchi, Felix Kölle, Daniele Nosenzo, and Simone Quercia. Bonus versus penalty: How robust are the effects of contract framing? *Journal of the Economic Science Association*, 3(2):174–182, 2017.
- Shir Dekel, Micah Goldwater, Dan Lovallo, and Bruce Burns. Effect of choice bracketing on risk aggregation in repeated-play gambles with no feedback. 2021.
- Andrew Ellis and David J Freeman. Revealing choice bracketing. *American Economic Review*, 114(9):2668–2700, 2024.
- Nisvan Erkal, Lata Gangadharan, and Nikos Nikiforakis. Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–48, 2011.
- Christine L Exley and Judd B Kessler. Equity concerns are narrowly framed. Technical report, National Bureau of Economic Research, 2018.
- Francesco Fallucchi and Daniele Nosenzo. The coordinating power of social norms. *Experimental Economics*, 25(1): 1–25, 2022.
- Diana Farrell, Fiona Greig, and Amar Hamoudi. The online platform economy in 2018: Drivers, workers, sellers, and lessors. *JPMorgan Chase Institute*, 2018.

- Diana Farrell, Fiona Greig, and Amar Hamoudi. The evolution of the online platform economy: Evidence from five years of banking data. In *AEA Papers and Proceedings*, volume 109, pages 362–66, 2019.
- Marcus Giamattei, Kyanoush Seyed Yahosseini, Simon Gächter, and Lucas Molleman. Lioness lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, pages 1–17, 2020.
- Justine S Hastings and Jesse M Shapiro. Fungibility and consumer choice: Evidence from commodity price shocks. *The quarterly journal of economics*, 128(4):1449–1498, 2013.
- Chip Heath and Jack B Soll. Mental budgeting and consumer decisions. *Journal of consumer research*, 23(1):40–52, 1996.
- Tanjim Hossain and John Morgan. ... plus shipping and handling: Revenue (non) equivalence in field experiments on ebay. *Advances in Economic Analysis & Policy*, 5(2), 2006.
- Alex Imas. The realization effect: Risk-taking after realized versus paper losses. *American Economic Review*, 106(8): 2086–2109, August 2016. doi: 10.1257/aer.20140386. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20140386>.
- Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2): 263–291, March 1979. ISSN 0012-9682. doi: 10.2307/1914185. URL <http://www.jstor.org/stable/1914185>.
- Maurits C Kaptein, Robin Van Emden, and Davide Iannuzzi. Tracking the decoy: maximizing the decoy effect through sequential experimentation. *Palgrave Communications*, 2(1):1–9, 2016.
- Alexander K Koch and Julia Nafziger. Correlates of narrow bracketing. *The Scandinavian Journal of Economics*, 121(4):1441–1472, 2019.
- Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz. Background risk and small-stakes risk aversion, 2021a.
- Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz. Monotone additive statistics, 2021b.
- Matthew Rabin and Georg Weizsäcker. Narrow bracketing and dominated choices. *American Economic Review*, 99(4):1508–43, 2009.
- Daniel Read, George Loewenstein, Matthew Rabin, Gideon Keren, and David Laibson. Choice bracketing. In *Elicitation of preferences*, pages 171–202. Springer, 1999.

Donald A Redelmeier and Amos Tversky. On the framing of multiple prospects. *Psychological Science*, 3(3):191–193, 1992.

Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *science*, 211(4481): 453–458, 1981.

Pauline Vorjohann. Reference-dependent choice bracketing. 2020.

## A Appendix: Proofs

### Proof of Proposition 1.

*Proof.* Since  $\Delta > 0$ , for every  $\delta \in [0, \Delta)$ , there exists some  $X, Y$ , and  $Y'$  s.t.  $P_B(X|Y) - P_S(X|Y') > \delta$ , hence there exist choice sets  $S$  and  $S'$  such that  $X + Y - m$  is chosen from  $S_1 \equiv \{X + Y - m, Y\} \cup S$  and  $Y'$  is chosen from  $S_2 \equiv \{X + Y' - m', Y'\} \cup S'$  with  $P_B(X|Y) - P_S(X|Y') \geq m - m' > \delta$ .

Then a narrow bracketer facing simultaneous choices  $S_1$  and  $S_2$  will pick  $X + Y - m$  from  $S_1$  and  $Y'$  from  $S_2$  for a total outcome of  $X + Y + Y' - m$ . However, they could have chosen  $Y$  from  $S_1$  and  $X + Y' - m'$  from  $S_2$  for a total of  $X + Y + Y' - m' = X + Y + Y' - m + (m - m')$ , thus leaving an amount of  $m - m' > \delta$  on the table, and the claim follows.  $\square$

### Proof of Proposition 2.

*Proof.* Let  $\bar{\Delta} \equiv \sup_{X,Y,Y'} P_B(X|Y) - P_S(X|Y')$  and suppose that  $\bar{\Delta} = 0$ . Then  $P_B(X|Y) = P_S(X|Y')$  for all  $X$  and  $Y$ , and hence  $P_B(X|Y) = P_B(X|Y') \equiv P_B(X)$  for all  $X, Y, Y'$ . Moreover,  $P_S(X|Y') = P_B(X)$  for all  $Y'$ , hence  $P_S(X|Y') = P_S(X) = P_B(X)$ . Hence we see that  $P_B(X) = P_S(X) = P(X)$ : the person always buys when the price of  $X$  on top of the highest-value alternative is less than  $P(X)$  and always sells when the price on top of the highest-value alternative is more than  $P(X)$ , and will do both when the price is  $X$ . Thus there is a unique price  $P(X)$  for each  $X$ .

Now first we show that if weak axiom of revealed preference (WARP) is violated, then it is violated by at least an amount  $\varepsilon > 0$ , meaning that there exist  $X, Y, S$ , and  $S'$  s.t.  $\{X, Y\} \subset S$ ,  $\{X + \varepsilon, Y\} \subset S'$ , with  $X \in C(S)$  and  $Y \in C(S')$ ,  $X + \varepsilon \notin C(S')$ , where  $C$  is the choice correspondence. Note that for  $\varepsilon = 0$ , we have the usual definition of WARP.

Now suppose that we have a WARP violation. Let  $X, Y, S$ , and  $S'$  be a WARP violation. We have that  $X, Y \in S \cap S'$ , with  $X \in C(S)$  and  $Y \in C(S')$  but  $X \notin C(S')$ . Note that if  $X \in C(\{X, Y\})$ , then there is also a WARP violation such that  $S$  equals  $\{X, Y\}$ . Similarly, if  $X \notin C(\{X, Y\})$ , then we can set  $S'$  equal  $\{X, Y\}$ .

**Case 1:**  $X \in C(\{X, Y\})$  and  $S = \{X, Y\}$ .

First notice that if  $C(\{X, Y\}) = \{X\}$ , then there is some  $\varepsilon > 0$  s.t.  $C(\{X - \varepsilon, Y\}) = \{X - \varepsilon\}$  by continuity: if  $Y$  was chosen for all sufficiently small  $\varepsilon$ , then it also would be chosen from  $\{X, Y\}$ . Thus in this case we have immediately a WARP violation of size  $\varepsilon$ .

If  $C(S') = \{Y\}$ , then by continuity, we have that  $C(S'_\varepsilon) = \{Y\}$  where  $S'_\varepsilon = S \setminus \{X\} \cup \{X + \varepsilon\}$  for sufficiently small  $\varepsilon$ : if this was not the case, say  $Z \neq Y, X + \varepsilon$  and  $Z \in C(S'_\varepsilon)$ , then by continuity  $Z \in C(S')$  — and a similar argument rules out that  $X - \varepsilon$  is chosen. Thus in this case, we have now found a WARP violation of size  $\varepsilon$ .

If  $C(S') \neq \{Y\}$ , then consider  $C(S'_\varepsilon)$ : if this contains  $Y$ , then we immediately have a WARP violation of size  $\varepsilon$ . So suppose that  $Y \notin C(S'_\varepsilon)$  for sufficiently small  $\varepsilon$ . Then there is some  $Z \in C(S'_\varepsilon)$  that is not  $X + \varepsilon$  (by continuity it can't be) nor  $Y$ . Now consider the set  $\{Y, Z\}$ . If  $C(\{Y, Z\}) = \{Y\}$ , then we can show as we did in the previous paragraph that there is some  $\varepsilon > 0$  s.t.  $C(\{Y - \varepsilon, Z\}) = \{Y - \varepsilon\}$ . Therefore, since  $Z \in C(S')$ , we have found a WARP violation of size  $\varepsilon$ , since  $Y - \varepsilon$  is chosen over  $Z$  in one choice, yet  $Z$  is chosen over  $Y$  in the other. The same argument applies if  $C(\{Y, Z\}) = \{Z\}$ , since the situation is symmetric.

The remaining possibility is thus that  $C(\{Y, Z\}) = \{Y, Z\}$ . We also know that  $Z$  is chosen from  $S'_\varepsilon$ , while  $Y$  and  $X + \varepsilon$  are not. Thus this is an example of a WARP violation, since  $Y$  is chosen from  $\{Y, Z\}$ , but only  $Z$  is from  $S'_\varepsilon$ . Moreover, our new example has choice  $S$  of size 2, and choice set  $S'_\varepsilon$  of same size as choice set  $S'$ . However, we can pick  $\varepsilon$  s.t. the size of the choice set,  $|C(S'_\varepsilon)|$ , is at least one less than  $|C(S')|$ : by continuity, every option chosen from  $S'_\varepsilon$  will also be chosen in the limit as  $\varepsilon \rightarrow 0$ , moreover, there are a finite number of options that don't change with  $\varepsilon$ , other than  $X - \varepsilon$ , which we know is not chosen. Thus the size can at most be equal, but given that we know that  $Y$  is not chosen for  $\varepsilon > 0$ , yet chosen for  $\varepsilon = 0$ , the size is at least one less.

But then we can repeat the previous argument for this newly found WARP violation and either we find a WARP violation of size  $\varepsilon$  or a new violation where the size of  $C(S')$  decreases again by at least one. Since this can never decrease to 0, this process must eventually stop, which it only does once we find a WARP violation of size  $\varepsilon$ .

**Case 2:** Suppose  $C(\{X, Y\}) = \{Y\}$

In this case, the same argument as we made previously immediately applies that  $C(\{X + \varepsilon, Y\}) = \{X + \varepsilon\}$  for sufficiently small  $\varepsilon$  and hence we have a WARP violation of size  $\varepsilon$ .

Together these cases prove that if there is any WARP violation, we must have a WARP violation of size  $\varepsilon$ . Since a WARP violation of size  $\varepsilon$  implies a price differential of size  $\varepsilon$ , this proves that  $\Delta = 0$  implies that WARP is satisfied.

Hence we know that the choices can be represented by a transitive preference,  $\succsim$ . Consider the choice set where the person chooses between  $Y$  and  $Y + X - m$ . Since  $P(X|Y) = P(X)$ , the person always chooses  $Y$  over  $X + Y - m$  for  $m > P(X)$ , and chooses only  $X + Y - m$  over  $Y$  for  $m < P(X)$ . Thus  $X + Y - m \succ Y$  if and only if  $m < P(X)$ ,  $Y \succ X + Y - m$  if and only if  $m > P(X)$ , and thus by continuity  $X + Y - P(X) \sim Y$ . Letting  $Y = Z + P(X)$ , we see that  $X + Z + P(X) - P(X) = X + Z \sim Z + P(X)$  for all  $Z$  and  $X$ , hence setting  $Z = 0$  we find that  $X \sim P(X)$ .

Since the person always chooses the strictly larger amount of money, when  $P(X) > P(Y)$ ,  $P(X)$  is strictly preferred to the amount  $P(Y)$ , so  $P(X) \succ P(Y)$ , and therefore  $X \sim P(X) \succ P(Y) \sim Y$ . Thus  $X$  is strictly preferred to  $Y$  if and only if  $P(X) > P(Y)$ . Thus  $P(\cdot)$  is in fact a utility representation for the preference relation.

Moreover, since, as we showed above,  $X + Z \sim P(X) + Z$ , we also have that  $P(X) + Z = Z + P(X) \sim P(Z) + P(X) = P(X) + P(Z)$ . Further, since  $X \sim P(X)$ , we also have that  $X + Z \sim P(X + Z)$ , hence  $P(X + Z) \sim X + Z \sim P(X) + Z \sim P(X) + P(Z)$ . Finally, indifference between these two amounts of money

implies that they are equal, i.e., if  $P(X + Z) = P(X) + P(Z)$ .

This completes the proof. □

## A.1 Implications of additive utility representations

We showed in the main text that  $P(\cdot)$  is additive when  $\Delta = 0$ . The next Proposition formalizes what additivity implies across different choice domains.

**Proposition 4.** *Suppose  $P(\cdot) : \mathbb{X} \rightarrow \mathbb{R}$  be an additive and continuous function. Then the following hold:*

1. *Let  $\mathbb{X}$  be a space of random variables rich enough so that if  $X, Y \in \mathbb{X}$ , then there is some random variable  $A$  distributed uniformly on  $[0, 1]$  that is independent of  $X$  and  $Y$  with  $\mathbb{1}(A \in (p, q)) \cdot X + \mathbb{1}(A \notin (p, q)) \cdot Y \in \mathbb{X}$  for any  $p, q \in [0, 1]$ . Assuming that  $P(X) = P(Y)$  whenever  $X$  and  $Y$  have the same distribution,  $P(\cdot)$  is the certainty equivalent for an expected-utility agent*
2. *if  $\mathbb{X}$  is a space of bounded real random variables, then  $P(\cdot)$  is the certainty equivalent for a CARA agent*
3. *if  $\mathbb{X}$  is a space of bounded and independent real random variables, then  $P(\cdot)$  is a weighted average of certainty equivalents of CARA agents*
4. *if  $\mathbb{X} = \mathbb{R}_{\geq 0}^n$  or  $\mathbb{X} = \mathbb{R}^n$ , then  $P(x) = x \cdot \lambda$  for some  $\lambda \in \mathbb{R}^n$  and for all  $x \in \mathbb{X}$*

Except for the first result on expected utility, results 2 through 4 have been noted in separate papers as conditions under which bracketing unidentified: Rabin and Weizsäcker (2009) assume expected utility and show that narrow bracketing incurs no cost (which is another way of stating that it is unidentified) if and only if agents have CARA preferences; Ellis and Freeman (2024) show that narrow bracketing is unidentified in multi-good choices when the utility is linear; and Mu et al. (2021b) consider additive certainty equivalents like  $P(\cdot)$  and prove statement 3 above. Our contribution is to show how additivity of  $P(\cdot)$  is the unifying feature behind all these results, to highlight that it implies other conditions (such as on expected utility, result 1) and that it will generalize to all other settings. We will now prove the special cases of Proposition 4.

*Proof. Case 1:*  $\mathbb{X}$  is a (sufficiently rich) space of random variables

Denote by  $X, Y$ , and  $Z$  three random variables. Let  $A$  be distributed uniformly on  $[0, 1]$ , independently of  $X, Y, Z$  and  $A(p, q)$  be the event that  $A \in (p, q)$ . Then  $\tilde{X} = \mathbb{1}(A(0, p)) \cdot X + \mathbb{1}(A(p, 1)) \cdot Z$  and  $\tilde{Y} = \mathbb{1}(A(0, p)) \cdot Y + \mathbb{1}(A(p, 1)Z)$  are the random variables yielding the value of  $X$  respectively  $Y$  with probability  $p$  and the value of  $Z$



with probability  $1 - p$ .

$$\begin{aligned}
P(\mathbb{1}(A(0, p+q)) \cdot X) &= P(\mathbb{1}(A(0, p)) \cdot X + \mathbb{1}(A(p, p+q)) \cdot X) \\
&= P(\mathbb{1}(A(0, p)) \cdot X) + P(\mathbb{1}(A(p, p+q)) \cdot X), \text{ by additivity} \\
&= P(\mathbb{1}(A(0, p)) \cdot X) + P(\mathbb{1}(A(0, q)) \cdot X), \text{ by equal distributions}
\end{aligned}$$

where we used the fact that  $\mathbb{1}(A(0, q)) \cdot X$  and  $\mathbb{1}(A(p, p+q)) \cdot X$  have the same distribution, hence also the same utility. Writing  $f_X(p) = P(\mathbb{1}(A(0, p)) \cdot X)$ , we have that  $f_X$  is additive, i.e. it satisfies  $f_X(p+q) = f_X(p) + f_X(q)$ . We assumed that it is continuous in  $p$ , hence we know that  $f_X$  is linear, i.e.  $f_X(p) = \lambda p$  for some  $\lambda$ . Since  $f_X(1) = P(X)$ , we have  $\lambda = P(X)$ , which shows that  $P(\mathbb{1}(A(0, p)) \cdot X) = pP(X)$ .

Let  $\tilde{X} = \mathbb{1}(A(0, p)) \cdot X + \mathbb{1}(A(p, 1)) \cdot Z$  and similarly for  $\tilde{Y} = \mathbb{1}(A(0, p)) \cdot Y + \mathbb{1}(A(p, 1)) \cdot Z$ . Suppose that  $X \sim Y$ , so that  $P(X) \geq P(Y)$ . Then we have  $P(\tilde{X}) = P(\mathbb{1}(A(0, p)) \cdot X) + P(\mathbb{1}(A(p, 1)) \cdot Z) = pP(X) + (1-p)P(Z) \geq pP(Y) + (1-p)P(Z) = P(\mathbb{1}(A(0, p)) \cdot Y) + P(\mathbb{1}(A(p, 1)) \cdot Z) = P(\tilde{Y})$ . Hence  $p \cdot X + (1-p) \cdot Z \sim p \cdot Y + (1-p) \cdot Z$  (the same argument applies to any event  $B$  with probability  $p$ ) showing that the independence axiom holds for  $\sim$ . Together with continuity, this implies expected utility.

**Case 2:**  $\mathbb{X}$  is a space of bounded real random variables

From case 1 we know that we have expected utility preferences.

Now let  $X$  be any random variable and  $w$  and  $w' \in \mathbb{R}$ . Then  $X + w$  denotes the random variable yielding  $w$  more than  $X$ . Then  $P(X + w) = P(X) + P(w) \implies P(X + w) - P(w) = P(X) = P(X + w') - P(w')$ . This is the certainty equivalent of  $X$ , once on top of  $w$ , once on top of  $w'$ , which has to be constant for all  $w$  and  $w'$ . Hence we must have constant absolute risk aversion.

**Case 3:** This result follows directly from Theorem 1 in Mu et al. (2021b), since  $P(\cdot)$  in this context is a monotone additive statistic over bounded real-valued random variables.

**Case 4:**  $\mathbb{X} = \mathbb{R}^n$

Letting  $e_i$  be the unit vector in dimension  $i$ , i.e. it is the bundle that provides one unit of good  $i$  and nothing else, then we can define  $f_i(x) := P(x \cdot e_i)$  for  $x \in \mathbb{R}$ .  $f_i(x+y) = P((x+y) \cdot e_i) = P(x \cdot e_i + y \cdot e_i) = P(x \cdot e_i) + P(y \cdot e_i) = f_i(x) + f_i(y)$ , so  $f_i$  is additive and continuous. It is well-known that additivity plus continuity for a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  implies that  $f(\cdot)$  is linear.<sup>33</sup> Thus  $f_i(x) = \lambda_i \cdot x$  for some  $\lambda_i \in \mathbb{R}$ . By additivity, we have that  $P(\mathbf{x}) = P(\sum_i x_i \cdot e_i) = \sum_i P(x_i \cdot e_i) = \sum_i f_i(x_i) = \sum_i \lambda_i x_i = \mathbf{x} \cdot \boldsymbol{\lambda}$ .  $\square$

<sup>33</sup>Additivity defining a function is also called Cauchy's functional equation and the stated result dates back to Cauchy.

## **B Appendix: Instructions**

### **B.1 Welcome Screen**

Welcome

Thank you for accepting our HIT.

During the HIT, please do not close this window or leave the HIT's web pages in any other way.

If you do close your browser or leave the HIT, you will not be able to re-enter and we will not be able to pay you!

You will receive a baseline payment of \$2.00 once you complete the HIT. Additionally, you can earn an extra bonus that will depend on your choices.

You will receive a code to enter into MTurk to collect your payment once you have finished.

Please read all instructions carefully.

### **B.2 Instructions**

Thank you for accepting to participate in this HIT. On top of the guaranteed payment of \$2.00 you will have the chance to earn an extra bonus, as explained later.

The task In this HIT you will decode several sequences of random letters into numbers with the given decoding table. For each letters sequence, the decoding table changes. The main part of the HIT will require you to decode several of these tasks.

To gain familiarity with the task you will now have to correctly decode 3 sequences. Note that each letter must be decoded correctly. After entering the decoded sequence, hit the submit button. Subsequently, irrespective of whether the text sequence was decoded correctly or not, a new sequence and decoding table will appear. Once you decode 3 sequences correctly, we describe the main part of the HIT.

## Encryption task

Decode sequences into numbers until you got 10 right.

Correctly decoded text sequences: 0

Decoding table

letter	u	t	f	c	g	n	v	q	a	l
corresponds to number	0	1	2	3	4	5	6	7	8	9

Text sequence: tvqqnqvfgug

Your answer (enter only numbers):

167757642404

Figure 3: SCREENSHOT

In the example you see the text sequence *tvqqnqvfgug*. The decoding table tells you that u=0, t=1,... This means that you have to decode *tvqqnqvfgug* into *167757642404* and enter this numeric value into the answer field.

### B.3 Task

### B.4 Main Task instructions

#### THE TASK

**[BOTH — MONEY/LOW — MONEY]** By completing the HIT you will receive \$2.00. To do so you are required to decode some sequences correctly for a bonus.

We will give you two pages of choices, with 16 choices on each page. Each choice is between a low number and a high number of **[additional - only in BOTH]** sequences to decode before the required sequences for different bonuses.

Example Choice (DOES NOT COUNT):

- 10 [**additional - only in BOTH**] sequences for an extra \$4.00
- 20 [**additional - only in BOTH**] sequences for an extra \$5.00

After you made your choice the computer will select randomly one of the 16 choices from one of the 2 pages. That option will be implemented. Thus you should select your preferred option for each choice.

[**NONE**] To complete the HIT you will be asked to decode a certain number of sequences correctly. The number of sequences you will be required to decode will depend on your choices.

We will give you two pages of choices, with 16 choices on each page. Each choice is between a low number and a high number of sequences to decode for different amounts (which includes the \$2.00 completion fee of the HIT).

- 10 sequences for a total payment of \$6.00
- 20 sequences for a total payment of \$7.00

After you made your choice the computer will select randomly one of the 16 choices from one of the 2 pages. That option will be implemented. Thus you should select your preferred option for each choice.

## B.5 Main Task - Scenario 1

[NONE] Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive a total payment (which includes the \$2.00 completion fee) depending on your choices.**

		OPTION A	OPTION B	
1)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$6.25
2)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$6.50
3)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$6.75
4)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$7.00
5)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$7.25
6)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$7.50
7)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$7.75
8)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$8.00
9)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$8.25
10)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$8.50
11)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$8.75
12)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$9.00
13)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$9.25
14)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$9.50
15)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$9.75
16)	15 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	30 sequences for a total payment of \$10.00

**[MONEY] Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

		OPTION A	OPTION B	
1)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$4.25
2)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$4.50
3)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$4.75
4)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$5.00
5)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$5.25
6)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$5.50
7)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$5.75
8)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$6.00
9)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$6.25
10)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$6.50
11)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$6.75
12)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$7.00
13)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$7.25
14)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$7.50
15)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$7.75
16)	15 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra of \$8.00

[BOTH]

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

		OPTION A	OPTION B	
1)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$4.25
2)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$4.50
3)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$4.75
4)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$5.00
5)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$5.25
6)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$5.50
7)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$5.75
8)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$6.00
9)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$6.25
10)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$6.50
11)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$6.75
12)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$7.00
13)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$7.25
14)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$7.50
15)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$7.75
16)	0 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	15 additional sequences for an extra \$8.00

[MONEY/LOW]

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

		OPTION A	OPTION B	
1)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$4.25
2)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$4.50
3)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$4.75
4)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$5.00
5)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$5.25
6)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$5.50
7)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$5.75
8)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$6.00
9)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$6.25
10)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$6.50
11)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$6.75
12)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$7.00
13)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$7.25
14)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$7.50
15)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$7.75
16)	0 sequences for an extra of \$4.00	<input type="radio"/>	<input type="radio"/>	15 sequences for an extra of \$8.00



## B.6 Main Task - Scenario 2

[NONE]

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive a total payment (which includes the \$2.00 completion fee) depending on your choices.**

		OPTION A	OPTION B	
1)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$6.25
2)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$6.50
3)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$6.75
4)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$7.00
5)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$7.25
6)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$7.50
7)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$7.75
8)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$8.00
9)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$8.25
10)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$8.50
11)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$8.75
12)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$9.00
13)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$9.25
14)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$9.50
15)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$9.75
16)	30 sequences for a total payment of \$6.00	<input type="radio"/>	<input type="radio"/>	45 sequences for a total payment of \$10.00

[BOTH]

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

		OPTION A	OPTION B	
1)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$4.25
2)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$4.50
3)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$4.75
4)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$5.00
5)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$5.25
6)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$5.50
7)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$5.75
8)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$6.00
9)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$6.25
10)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$6.50
11)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$6.75
12)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$7.00
13)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$7.25
14)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$7.50
15)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$7.75
16)	15 additional sequences for an extra \$4.00	<input type="radio"/>	<input type="radio"/>	30 additional sequences for an extra \$8.00

**[MONEY/LOW]**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

		OPTION A	OPTION B	
1)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$4.25
2)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$4.50
3)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$4.75
4)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$5.00
5)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$5.25
6)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$5.50
7)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$5.75
8)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$6.00
9)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$6.25
115)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$6.50
11)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$6.75
12)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$7.00
13)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$7.25
14)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$7.50
15)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$7.75
16)	15 sequences for an extra bonus of \$4.00	<input type="radio"/>	<input type="radio"/>	30 sequences for an extra bonus of \$8.00

## **B.7 Results**

### **[BOTH] SUMMARY OF THE TASK**

The computer randomly selected the Choice # from Scenario #.

For this option you selected that you are (are not) willing to decode # additional sequences for \$X.XX.

In total you will decode # sequences to receive the HIT payment and the bonus.

### **[NONE & MONEY/LOW] SUMMARY OF THE TASK**

The computer randomly selected the Choice # from Scenario #.

For this option you selected that you are (are not) willing to decode # sequences in total for \$X.XX.

In total you will decode # sequences to receive the HIT payment and the bonus.

## **B.8 MAIN TASK**

## **B.9 PAYMENT PAGE**

Your earnings

In today's HIT you have earned a bonus of \$.

Your guaranteed participation fee is: \$2.00.

So, in total, you have earned \$.

To receive your earnings, please enter this code into MTurk

After you have done that, you can close this window. We thank you for participating in our study.

## C Appendix: Additional Results

### C.1 Main Results: Wilcoxon tests, tests by gender

Here we report results comparing the treatments by Wilcoxon and t-tests, running them for all participants, as well as for female and male participants separately.

Table 6: Between-treatment p-values for main treatments based on two-sided Wilcoxon rank-sum tests, treating each individual in each scenario as a single independent observation. The first two columns are for pooled data, the next two for data restricted to female participants, the final two for data restricted to male participants.

	Pooled		Female		Male	
Treatments	NONE	BOTH	NONE/F	BOTH/F	NONE/M	BOTH/M
<b>Scenario 1</b>						
BOTH	< 0.001		< 0.001		0.002	
MONEY/LOW	< 0.001	0.117	0.001	0.445	0.052	0.300
<b>Scenario 2</b>						
BOTH	0.023		0.277		0.038	
MONEY/LOW	0.108	0.704	0.357	0.984	0.121	0.809

Table 7: Between-treatment p-values for main treatments based on two-sided t-tests by scenario, treating each individual in each scenario as a single independent observation. The first two columns are for pooled data, the next two for data restricted to female participants, the final two for data restricted to male participants.

	Pooled		Female		Male	
Treatments	NONE	BOTH	NONE/F	BOTH/F	NONE/M	BOTH/M
<b>Scenario 1</b>						
BOTH	< 0.001		< 0.001		0.001	
MONEY/LOW	< 0.001	0.106	< 0.001	0.342	0.055	0.241
<b>Scenario 2</b>						
BOTH	0.040		0.264		0.072	
MONEY/LOW	0.120	0.753	0.383	0.909	0.133	0.886

## C.2 Implementation Details

We collected more participants in the BOTH treatment than in other treatments, because our initial version of the BOTH treatment informed participants of their endowment on the page *before* the first choice. In later versions we informed participants of their endowment only on the first choice page to (fast) reference effects not present in other treatments, where participants saw the information only on the first choice page. For this reason we continued collecting observations in this version of BOTH until we had enough data to compare it to the other treatments, which is why we ended up with more observations in BOTH (pooling all versions) than for the other treatments. As we show in Appendix C.5, our results are robust to this change.

Other minor details that changed between our pre-registration and our actual implementation are that, fixing total outcomes, we only used two levels for the endowment (0 and 15 tasks) instead of three (0, 8, and 16 tasks); and that we limited ourselves to 2 scenarios per person rather than 5.

We added the treatment MONEY later to disentangle the contribution of bracketing of money and work dimensions separately. Specifically, the MONEY treatment combines all the tasks, hence differences between NONE and MONEY cannot be driven by a failure to combine work, while differences between MONEY and BOTH are direct evidence of narrow bracketing of money. To distinguish fully between the two, we should also have included a treatment that combines money but not work. This version of the MONEY treatment has the benefit that it can directly be compared to MONEY/LOW, providing additional tests of narrow bracketing.

The following table 8 shows how many participants we recruited in which treatments during sessions on different days.

Table 8: Participant numbers by sessions and treatments

Session ID	Session Date	NONE	BOTH	MONEY/LOW	MONEY
1	2019-12-18	9	9	8	0
2	2019-12-19	11	12	13	0
3	2019-12-19	39	38	40	0
4	2019-12-19	18	20	19	0
5	2019-12-20	68	0	0	0
6	2019-12-21	0	68	0	0
7	2019-12-23	0	0	61	0
8	2019-12-30	23	24	23	0
9	2020-01-21	15	13	15	43
10	2020-01-22	5	4	5	14
11	2020-01-28	12	13	12	36
12	2020-01-30	0	0	0	68
13	2020-02-04	0	0	0	52
14	2020-02-05	0	38	0	0
15	2020-02-06	0	81	0	0
		200	320	196	213

### C.3 Attrition

Table 9 displays at what stage participants dropped out of the study.

Table 9: Attrition in % by a given stage

Treatments	Practice	Choice 1	Answer 1	Learn Tasks	End
<b>Main</b>					
NONE	11%	11%	11%	12%	18%
BOTH	11%	11%	11%	12%	20%
MONEY/LOW	9%	10%	10%	11%	13%
MONEY	8%	8%	8%	8%	16%
<b>Follow Up</b>					
BEFORE	7%	7%	7%	7%	18%
AFTER	7%	7%	7%	7%	14%

Table 10: Between-treatment p-values for NONE, BOTH, and MONEY/LOW treatments based on two-sided t-test, treating each individual in each scenario as a single independent observation. Restricted to sessions in which these three treatments were balanced.

Treatments	NONE	BOTH
<b>Scenario 1</b>		
BOTH	< 0.001	
MONEY/LOW	0.003	0.316
<b>Scenario 2</b>		
BOTH	0.245	
MONEY/LOW	0.222	0.936

#### C.4 Main treatments balanced data only

Since our experimental sessions were not always balanced, one possible concern might be that we get different results due to different populations across sessions. To alleviate this concern with respect to the main treatments (NONE, BOTH, and MONEY/LOW), we report here the differences in t-tests and Wilcoxon tests when restricting ourselves to the data that was collected in a balanced session – that is, data where membership was randomized and equal within each session. As the results show, we still reject broad bracketing, and fail to reject narrow bracketing.



Table 11: Between-treatment p-values for NONE, BOTH, and MONEY/LOW treatments based on two-sided Wilcoxon rank-sum tests, treating each individual in each scenario as a single independent observation. Restricted to sessions in which these three treatments were balanced.

Treatments	NONE	BOTH
<b>Scenario 1</b>		
BOTH	< 0.001	
MONEY/LOW	0.004	0.369
<b>Scenario 2</b>		
BOTH	0.270	
MONEY/LOW	0.284	0.941

Table 12: Between-treatment p-values for main treatments based on two-sided t-test, treating each individual in each scenario as a single independent observation. Restricted to those sessions of BOTH where baseline is revealed right before the first choice page.

Treatments	NONE	BOTH
<b>Scenario 1</b>		
BOTH	< 0.001	
MONEY/LOW	< 0.001	0.243
<b>Scenario 2</b>		
BOTH	0.400	
MONEY/LOW	0.120	0.490

## C.5 Baseline Tasks revealed right before the choice page only

Next we report results from the initial BOTH treatments where the information was displayed on the page right before. The results are essentially the same, although there is no longer a statistically significant difference in scenario 2, since BOTH lies between MONEY/LOW and NONE and is not significantly different from either, reflecting the lower power due to closer to 'linear' preferences in Scenario 2 (the difference between MONEY/LOW and NONE is lower).

We compare the means of BOTH treatments with message displayed before the first choice page and on the first choice page by scenario directly in Table 13. This shows that for scenario 2, these two versions are significantly (and sizeably) different, reflecting also that in one case this leads to rejection of broad bracketing in scenario 2 and once it doesn't. No matter which is the accurate treatment, both reject broad bracketing, and neither rejects narrow bracketing.

There are two possible reasons for the difference: either it is due to the display of information, in which case the later data with information display on the page is the appropriate test, rejecting broad in both scenarios. In this case, the treatments BOTH and NONE are not balanced within sessions, since we had completed collection of data on NONE (mostly at least, we have a small overlap between the treatments). Or it is due to changes in the population due to sampling at different times. In this case the earlier data is the appropriate test, and balances observations against the NONE treatment – i.e. the rejection of broad bracketing cannot be due (or more correctly, is statistically unlikely to

Table 13: Between-treatment p-values for BOTH when information on baseline is presented for the first time right before the first choice or exactly on the first choice page. Based on two-sided Wilcoxon rank-sum tests, treating each individual in each scenario as a single independent observation.

Scenarios	t-test
Scenario 1	0.551
Scenario 2	0.001

be due) to different preferences.

Table 15: Frequencies (in %) of individuals who switch up, switch down, or stay at the same reservation wage from scenario 1 to scenario 2. The final column reports how many more people switch up rather than down.

Treatment	Down	Stay	Up	Drop out	Up - Down
NONE	16	44	34	6	18
BOTH	6	27	60	8	54
MONEY/LOW	10	38	49	4	39
MONEY	30	42	26	2	-4

## C.6 Individual-level changes in reservation wage between scenarios

Frequencies of individuals who switched up, down, or stayed, and size of jumps by switching up or down.

Table 14: Average individual-level change in reservation wage, conditional on whether the jump was up or down.

Treatment	Down	Up
NONE	-1.10	0.83
BOTH	-0.83	1.07
MONEY/LOW	-0.69	0.94
MONEY	-1.10	0.93

## C.7 Plots of reservation wages by treatment

Figure 4: A bar plot of the raw reservation wages by treatment and scenario

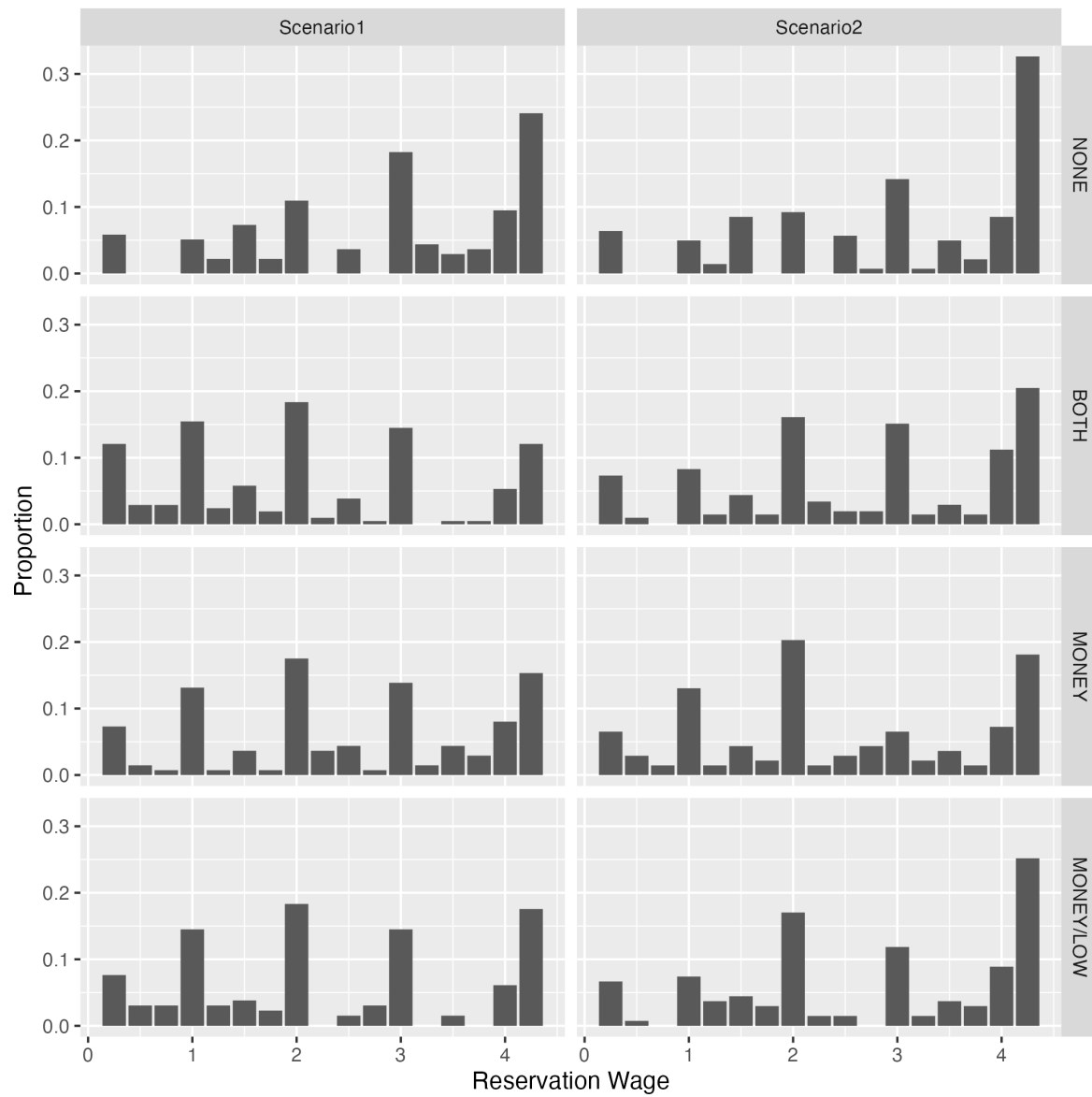
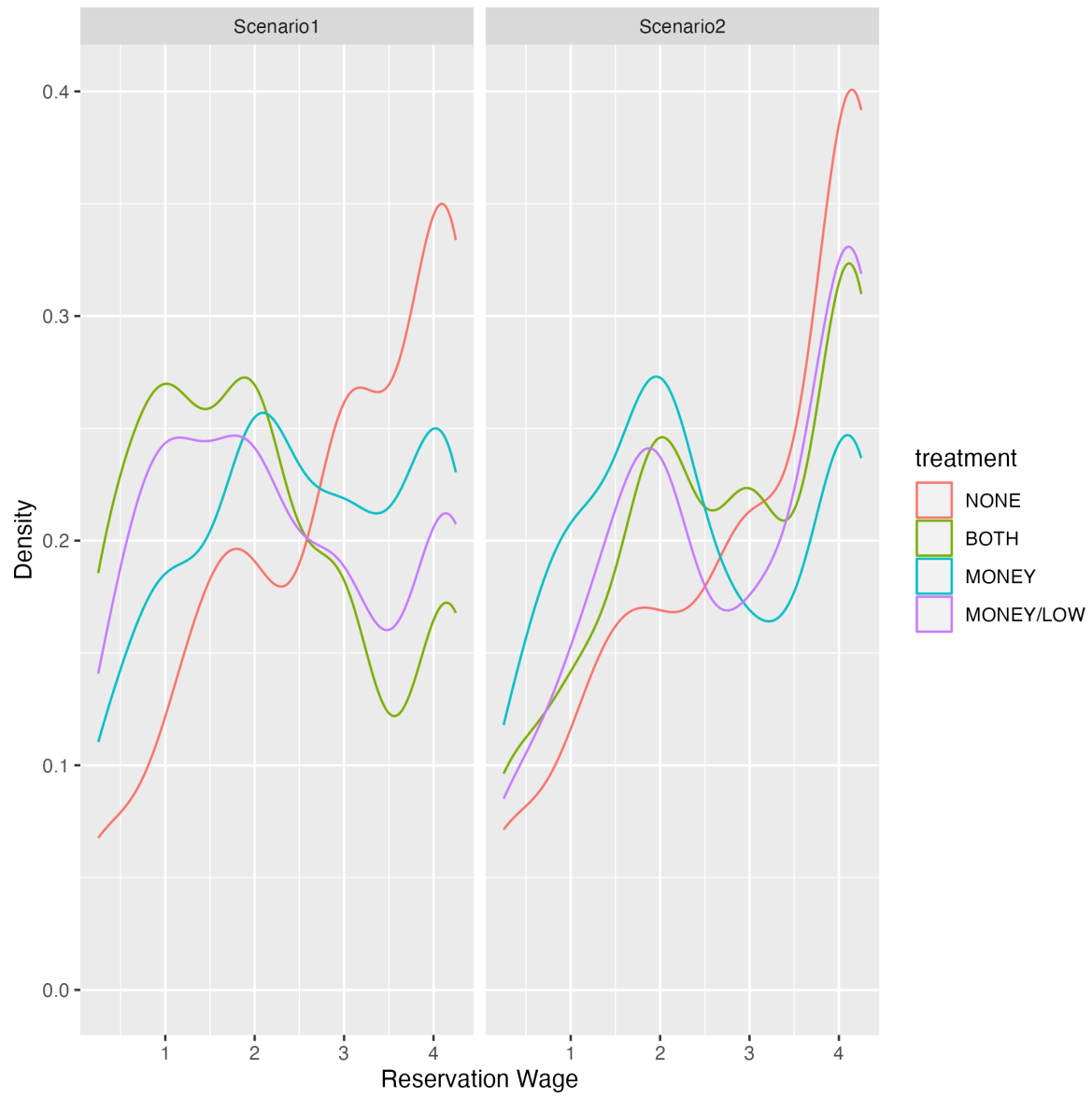


Figure 5: A kernel density plot of the raw reservation wages by treatment and scenario



## C.8 Linear Regression Results

In our pre-registration, we stated that we would use a linear regression that averages across the scenarios to test for bracketing, whether broad or narrow. This is however not the appropriate test, since bracketing should apply for each Scenario separately, not just for the average scenario, which is why those are the tests we use instead. This is best illustrated via an actual example we have in our data. We report the linear regression results in 16, which rejects broad bracketing overall; fails to reject narrow bracketing; and rejects broad bracketing of money alone. It does not however reject broad bracketing of work alone, since the treatment effects for MONEY and BOTH are  $-0.47$  and  $-0.55$  respectively with standard errors of around 0.11, so they are not different.

Table 16: Linear Regressions of reservation wages by treatment averaged across scenarios. With and without clustered standard errors by participant. The default treatment is NONE, so that broad bracketing predicts a null estimate for the fixed effect for BOTH (rejected), and narrow bracketing predicts equal fixed effects for BOTH and MONEY/FALSE (not rejected). Broad bracketing of money (but not necessarily work) predicts a null estimate for MONEY (rejected); broad bracketing of work (but not necessarily money) predicts equal fixed effects for BOTH and MONEY (not rejected). The last-mentioned test is not rejected because the difference is positive in Scenario 1 and negative in Scenario 2, which is average in this regression. Since bracketing makes predictions that should hold for every scenario, the test in our main text is the better test (accounting for double-testing).

	<i>Dependent variable:</i>	
	Reservation wage	
	No clustering	Clustering by participant
	(1)	(2)
scenarioScenario2	0.31 (0.07)	0.31 (0.04)
treatmentBOTH	$-0.55$ (0.10)	$-0.55$ (0.13)
treatmentMONEY	$-0.47$ (0.11)	$-0.47$ (0.14)
treatmentMONEY/LOW	$-0.41$ (0.11)	$-0.41$ (0.15)
Constant	2.78 (0.09)	2.78 (0.10)
Observations	1,231	1,231
R <sup>2</sup>	0.04	0.04
F Statistic (df = 4; 1226)	12.62	12.62

*Note:*

Standard errors in parentheses.

But, as the results from 17 shows, this is because the treatment effect in Scenario 1 is substantially smaller for BOTH than for MONEY (2.07 vs 2.50); yet in Scenario 2, it is substantially larger for BOTH than for MONEY

Table 17: Mean comparisons between NONE, MONEY, and BOTH

Scenario	Treatment	Mean	N	Std. Err.
Scenario1	BOTH	2.07	208	0.09
Scenario2	BOTH	2.70	206	0.09
Scenario1	NONE	2.88	137	0.11
Scenario2	NONE	2.99	141	0.11
Scenario1	MONEY	2.50	137	0.11
Scenario2	MONEY	2.43	138	0.11

(2.70 vs 2.43) which averages out to a statistically insignificant difference of 0.08 when averaged across scenarios. Since narrow and broad bracketing make predictions for every choice decision, their predictions hold for each scenario individually, thus the linear regression is not the appropriate test and overly conservative.

The fact that the reservation wage in MONEY stays the same across the two scenarios is consistent with participants bracketing all their endowments narrowly: the reservation wage for NONE is also almost constant, 2.88 in Scenario 1 vs 2.99 in Scenario 2 (see 18). Our data suggests that doing 15 more tasks is equally unpleasant when one has to do 15 tasks or 30 tasks, both in treatments NONE and MONEY. The lower reservation wage in MONEY reflects the fact that MONEY has an endowment of money. If they ignore it and have concave utility from money, then for the same experienced disutility of work, they should be more willing to work for money in MONEY than in NONE. In the latter, they take into account that their wealth is higher.

Table 18: Means of main treatments by scenario

Treatment	Res. Wage	Std Err	% upper bound	N
<b>Scenario 1</b>				
NONE	2.88	0.11	24%	137
BOTH	2.07	0.09	12%	207
MONEY/LOW	2.31	0.12	18%	131
<b>Scenario 2</b>				
NONE	2.99	0.11	33%	141
BOTH	2.70	0.09	20%	205
MONEY/LOW	2.74	0.11	25%	135

### C.9 Size of reservation wage changes due to bracketing

Going beyond our pre-registration, we estimate how much the reservation wage of participants changes due to bracketing. We report these changes and their standard errors in Table 19. As we can see, the reservation wages change by as much as \$0.82. We also provide the equivalent of this change in reservation wages in terms of tasks and time spent

Table 19: The table reports  $\Delta$ , the change in reservation wages between NONE and BOTH. The highest average reservation wage for 15 more tasks is 2.99 across all treatments and scenarios, so that  $2.99/15 \approx 0.20$  is an upper bound for the average cost per task. Using this, we can convert  $\Delta$  into task-equivalents by  $\Delta/0.20$ , and the cost in time-equivalents (in seconds) by  $(\Delta/0.20)/46$ , since the average time taken for a task is 46 seconds.  $\hat{\Delta}$  stands for  $m_{NONE} - m_{MONEY/LOW}$ : the change in the marginal disutility of doing 15 extra tasks on top of a low vs on top of a high baseline. Under full narrow bracketing,  $\hat{\Delta}$  and  $\Delta$  should be equal.

Scenario	Gender	$\Delta$	Std.Err.	$\hat{\Delta}$	Task equivalent	Time equivalent (in secs)
<b>Pooled</b>						
Scenario1	Pooled	0.82	0.14	0.58	4.10	188
Scenario2	Pooled	0.29	0.14	0.25	1.46	67
<b>Female</b>						
Scenario1	Female	1.14	0.21	0.91	5.71	262
Scenario2	Female	0.26	0.23	0.23	1.28	59
<b>Male</b>						
Scenario1	Male	0.60	0.18	0.38	3.01	138
Scenario2	Male	0.32	0.18	0.30	1.61	74

on tasks (in seconds). To do so, we use the fact that across all treatments and scenarios, the average reservation wage for 15 additional tasks is never higher than \$2.99, meaning that (on average) people are willing to do an additional task for  $\$2.99/15 \approx \$0.20$ , so that \$0.82 is equivalent to about 4 tasks. This in turn is equivalent to 3 minutes (188 seconds), given that the average time spent per task is 46 seconds.

We also report the change in reservation wages by gender based on Koch and Nafziger (2019), who state that “[w]omen seem to be more prone to narrow bracketing than men”. The estimated changes range from \$0.14 (Scenario 2) to \$0.41 (Scenario 1) in the pooled data, from \$0.13 to \$0.57 for female and from \$0.16 to \$0.30 for male participants. Thus we also find that women have larger changes due to bracketing. Note however that under full narrow bracketing, since the reservation wage in BOTH should equal that in MONEY/LOW, we would expect that  $\Delta := m_{NONE} - m_{BOTH} = m_{NONE} - m_{MONEY/LOW}$ . Since  $m_{NONE}$  measures the marginal disutility when the baseline is 15 tasks higher than for  $m_{MONEY/LOW}$ , this is a measure of the convexity of disutility, and  $\Delta$  is predicted to be larger the more convex the preferences are. We therefore report  $\hat{\Delta} := m_{NONE} - m_{MONEY/LOW}$  in Table 19.

This shows that the results are in line with both men and women bracketing narrowly, yet women having larger changes in reservation wages due to more convex disutility from work. For women, the marginal disutility of doing 15 sequences on top of 0 sequences (elicited in Scenario 1 of MONEY/LOW) is \$0.91 higher than their marginal disutility of doing 15 sequences on top of 15 sequences (as elicited in Scenario 1 of BOTH). For men, this figure stands only at \$0.38. For both genders, we can reject broad bracketing, but not narrow bracketing (see Appendix C.1 for details). This suggests that some of the gender differences in Koch and Nafziger (2019) may be due to gender differences in preferences.



## C.10 Bracketing in Work and Money Dimensions Separately

Our results show that people do not bracket broadly. It is still possible that this is only due to a failure to bracket broadly their endowment of money, even though they broadly bracket their work endowment. We therefore ran an additional treatment, MONEY, in which there was only an endowment of MONEY, but no endowment of work. Thus, if participants in treatment BOTH bracketed work broadly, they should behave identically to participants in MONEY in both scenarios. We did not preregister for this treatment and we ran it after having started our main treatments. For this reason, the participants in MONEY are not balanced by session against the other treatments. See Appendix C.2 for details.

Formally, we have:

**Hypothesis 3** (Broad Bracketing of Work). *Behavior is consistent with broadly bracketing work if  $m_{MONEY} = m_{BOTH}$  in every Scenario.*

**Hypothesis 4** (Broad Bracketing of Money). *Behavior is consistent with broadly bracketing money if  $m_{MONEY} = m_{NONE}$  in every Scenario.*

Table 20 reports the differences in means between treatments MONEY and BOTH. The results show a sizeable and statistically significant difference in scenario 1 (p-value: 0.002), and a smaller and statistically insignificant difference for Scenario 2, similar to the main results. We thus reject that participants broadly bracket work.<sup>34</sup>

**Result 4.** *We reject Hypothesis 3 that individuals bracket the work dimension broadly.*

Regarding gender differences, we see that men's choices are consistent with broad bracketing of work, while women's choices are not. However, men's choices are also consistent with full narrow bracketing, because men's choices are consistent with linear costs, while women's choices aren't. Thus the identification assumption fails for men and holds for women. Thus while it might be justified to claim that women incur larger costs than men from bracketing work, it would be wrong to claim that they are bracketing work more narrowly.

Table 21 similarly shows that people do not broadly bracket money, and that both men and women bracket it narrowly.

**Result 5.** *We reject the Hypothesis 4 that individuals bracket the money dimension broadly.*

Figure 6 provides a bar plot of all the means by scenario.

---

<sup>34</sup>This is the only result that is not statistically significant when we instead use a linear regression that averages differences in reservation wages across scenarios – see Appendix C.8. The reason is clear from 20: in Scenario 1, the difference between BOTH and MONEY is  $2.07 - 2.50 = -0.43$ , while it is  $2.70 - 2.43 = 0.27$  in Scenario 2, which averages out to a difference of  $(-0.43 + 0.27)/2 = 0.08$  that is indistinguishable from 0. Since bracketing predicts identical behavior for each Scenario, averaging across scenarios as the linear regression does is not the appropriate test and overly conservative – both for testing broad and narrow bracketing.

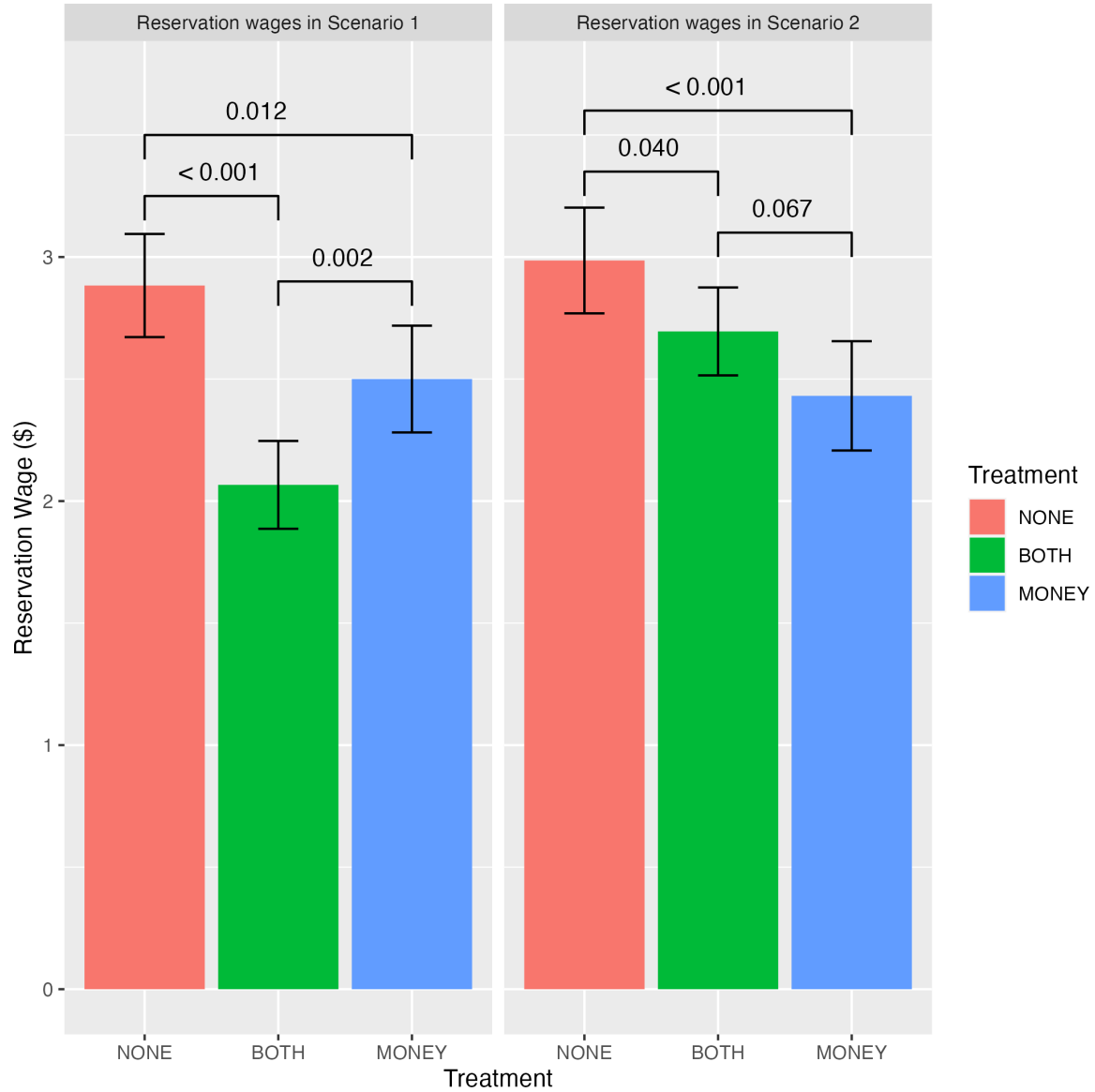
Table 20: The table reports  $\Delta$ , the change in reservation wages between MONEY and BOTH. Under broad bracketing of work,  $\Delta$  should be 0, see Hypothesis 3.  $\hat{\Delta}$  stands for  $m_{MONEY} - m_{MONEY/LOW}$ . When  $\hat{\Delta} = 0$ , we cannot identify full narrow bracketing from broad bracketing of work.

Scenario	Gender	$\Delta$	Std. Err.	$\hat{\Delta}$
<b>Pooled</b>				
Scenario1	Pooled	0.43	0.14	0.19
Scenario2	Pooled	-0.26	0.14	-0.31
<b>Female</b>				
Scenario1	Female	0.65	0.22	0.42
Scenario2	Female	-0.26	0.24	-0.29
<b>Male</b>				
Scenario1	Male	0.28	0.19	0.06
Scenario2	Male	-0.28	0.18	-0.31

Table 21: The table reports  $\Delta$ , the change in reservation wages between MONEY and NONE. Broad bracketing of money requires  $\Delta = 0$ .  $\hat{\Delta}$  stands for  $m_{BOTH} - m_{MONEY/LOW}$ . Full narrow bracketing requires  $\hat{\Delta} = 0$ .

Scenario	Gender	$\Delta$	Std. Err.	$\hat{\Delta}$
<b>Pooled</b>				
Scenario1	Pooled	-0.38	0.15	-0.24
Scenario2	Pooled	-0.55	0.16	-0.05
<b>Female</b>				
Scenario1	Female	-0.49	0.24	-0.23
Scenario2	Female	-0.51	0.26	-0.03
<b>Male</b>				
Scenario1	Male	-0.32	0.20	-0.22
Scenario2	Male	-0.60	0.20	-0.03

Figure 6: A bar plot of the reservation wages for treatments NONE, BOTH, and MONEY by scenario.



## D Appendix: Study 2 Instructions

### D.1 Main Task Instructions

#### THE MAIN STUDY

We now describe the main study.

**Guaranteed Participation Bonus:**

- For completing the study, you will receive a guaranteed payment of £3.00.

### **Additional Bonus**

- You have the chance to earn an additional bonus based on your decisions and performance.
- Specifically, you can earn this bonus by successfully decoding a certain number of sequences.

### **Study Structure:**

- During the study, you will encounter 35 choice pages. Each page presents a different scenario with various sets of sequences for you to decode.
- Your choice is to select your preferred option from these scenarios on each choice page.

### **Bonus Determination:**

- After you make your decisions, the computer will randomly select one of the 35 scenarios as the basis for your bonus payment and workload.
- This means that the number of sequences you need to decode and any additional bonus you earn will depend on the scenario selected by the computer.

### **Further Instructions:**

- Detailed explanations of how your decisions influence the sequences and bonuses will be provided on the next page.

## **D.2 Study Description - Part 1**

In each decision screen, you will be presented with one of three different types of SCENARIOS.

### **Type 1 SCENARIOS**

The first type is a choice between two or more alternatives, each involving different bonuses and sequences to decode.

Example (DOES NOT COUNT):

*Choose your preferred option:*

- Option A: Decode 5 sequences for £0.30
- Option B: Decode 1 sequence for £0.00
- Option C: Decode 2 sequences for £0.10

If, for example, you select option C, you will decode 2 sequences and receive £0.10.

### **Type 2 SCENARIOS**

The second type is a choice of how many sequences you are willing to decode if they are paid a certain sum each.

Example (DOES NOT COUNT):

*You can decode up to 10 sequences, and you will earn £0.05 per sequence decoded correctly.*

**How many sequences would you like to decode?**

If, for example, you choose that you are willing to decode 5 sequences for £0.05, that means that you are not willing to decode more than 5 sequences for such payment.

### **Type 3 SCENARIOS**

In these scenarios, the options will involve a certain probability of decoding sequences.

Example (DOES NOT COUNT):

*Choose your preferred option:*

- Option A: Receive £1.00 for sure. And if the six-sided virtual die rolls 1 decode 5 sequences; otherwise, decode 0 sequences
- Option B: Decode 1 sequence for £0.00

Probability-based selection process:

If, for example, you select Option A, you will receive £1.00 for sure. Additionally, you will either decode 5 sequences (if the virtual die rolls a 1) or 0 sequences (otherwise).

## **D.3 Study Description - Part 2**

Some Scenarios will have two choices like the one presented earlier to be taken at the same time. In that case, the output of bonuses and sequences to decode of both choices will matter.

Example (DOES NOT COUNT):

In this Scenario you have to make two decisions.

### **Decision 1**

*Choose your preferred option:*

- Option A: Decode 5 sequences for £0.30
- Option B: Decode 1 sequence for £0.00
- Option C: Decode 2 sequence for £0.10

Before answering, read the next decision

## Decision 2

*Choose your preferred option:*

- Option D: Decode 10 sequences for £1.00
- Option E: Decode 1 sequence for £0.00

Suppose that in the example above you choose Option A in Choice 1 and Option D in Choice 2.

If the computer selects this Scenario, then you will be asked to decode 15 (5+10) sequences and you will receive an additional bonus of £1.30 (£0.30+£1.00).

**Selection Process:** The computer will randomly select one scenario at the end of all the choices.

Based on the choices you made in this randomly selected scenario, you will be asked to decode the sequences and receive the payment associated with your chosen option.

Suppose that in the example above you choose Option A in Choice 1 and Option D in Choice 2.

If the computer selects this Scenario, then you will be asked to decode 15 (5+10) sequences and you will receive an additional £1.30 (£0.30+£1.00).

## D.4 Study Description - Summary

### TO SUMMARIZE

- You will make decisions in 35 Scenarios in total.
- The computer will randomly select one of these Scenarios.
- For this selected Scenario you will be asked to decode the sequences and receive a bonus depending on the choice(s) you made in that Scenario.

## D.5 Comprehension Questions

### QUIZ

**Before proceeding you have to answer to the following questions. If you fail to answer twice the study will finish and you will not receive any payment.**

**Which of the following is true?**

- Only the decisions made in one of the Scenarios will be relevant for the payment of the bonus.
- The bonus sums up the decisions made in all Scenarios.

**Which of the following is true?**

When more decisions are shown under the same Scenario page..

- ..only one of them will be relevant for the payment of the bonus.
- ..all the decisions will be relevant for the payment of the bonus.

## D.6 Main Task - list of Scenarios presented individually

SCENARIO 1 Choose your preferred option:

- OPTION A: Decode 0 sequences for £0.25
- OPTION B: Decode 10 sequences for £0.45
- OPTION C: Decode 10 sequences for £0.40

SCENARIO 2 Choose your preferred option:

- OPTION A: Decode 5 sequences for £0.45
- OPTION B: Decode 12 sequences for £0.80

SCENARIO 3 Choose your preferred option:

- OPTION A: Decode 7 sequences for £0.40

- OPTION B: Decode 13 sequences for £0.45
- OPTION C: Decode 15 sequences for £0.70

SCENARIO 4 Choose your preferred option:

- OPTION A: Decode 4 sequences for £0.40
- OPTION B: Decode 14 sequences for £0.79

SCENARIO 5 Choose your preferred option:

- OPTION A: Decode 5 sequences for £0.45
- OPTION B: Decode 12 sequences for £0.75
- OPTION C: Decode 12 sequences for £0.70

SCENARIO 6 Choose your preferred option:

- OPTION A: Decode 5 sequences for £0.35
- OPTION B: Decode 16 sequences for £0.90

SCENARIO 7 Choose your preferred option:

- OPTION A: Decode 5 sequences for £0.40
- OPTION B: Decode 12 sequences for £0.75
- OPTION C: Decode 10 sequences for £0.50

SCENARIO 8 Choose your preferred option:

- OPTION A: Decode 10 sequences for £0.45
- OPTION B: Decode 14 sequences for £0.75
- OPTION C: Decode 4 sequences for £0.40

SCENARIO 9 Choose your preferred option:



- OPTION A: Decode 15 sequences for £0.75
- OPTION B: Decode 7 sequences for £0.40

SCENARIO 10 Choose your preferred option:

- OPTION A: Decode 7 sequences for £0.40
- OPTION B: Decode 9 sequences for £0.45
- OPTION C: Decode 15 sequences for £0.70

SCENARIO 11 Choose your preferred option:

- OPTION A: Decode 0 sequences for £0.25
- OPTION B: Decode 10 sequences for £0.50

SCENARIO 12 Choose your preferred option:

- OPTION A: Decode 5 sequences for £0.35
- OPTION B: Decode 16 sequences for £0.85
- OPTION C: Decode 14 sequences for £0.45

SCENARIO 13 Description

You can decode up to 25 sequences, and you will earn £0.12 per sequence decoded correctly.

How many sequences would you like to decode?

SCENARIO 14 Description

You can decode up to 25 sequences, and you will earn £0.08 per sequence decoded correctly.

How many sequences would you like to decode?

SCENARIO 15 Description

You can decode up to 25 sequences, and you will earn £0.05 per sequence decoded correctly.

How many sequences would you like to decode?

SCENARIO 16 Description

You can decode up to 25 sequences, and you will earn £0.03 per sequence decoded correctly.

How many sequences would you like to decode?

#### SCENARIO 17 Description

You can decode up to 25 sequences, and you will earn £0.02 per sequence decoded correctly.

How many sequences would you like to decode?

#### SCENARIO 18 Choose your preferred option:

- OPTION A: Receive £0.30 for sure. And if the virtual six-sided die rolls 6, decode 20 sequences; otherwise, decode 0 sequences
- OPTION B: Decode 0 sequences for £0.15

#### SCENARIO 19 Choose your preferred option:

- OPTION A: Receive £0.40 for sure. And if the virtual six-sided die rolls 4 or 5, decode 20 sequences; otherwise, decode 0 sequences
- OPTION B: Receive £0.30 for sure. And if the virtual six-sided die rolls 5 decode 20 sequences; otherwise, decode 0 sequences

#### SCENARIO 20 Choose your preferred option

- OPTION A: Receive £0.50 for sure. And if the virtual six-sided die rolls 5 or 6, decode 20 sequences; otherwise, decode 0 sequences
- OPTION B: Decode 0 sequences for £0.25

#### SCENARIO 21 Choose your preferred option:

- OPTION A: Receive £0.70 for sure. And if the virtual six-sided die rolls 1, 2, 3 or 4, decode 20 sequences; otherwise, decode 0 sequences
- OPTION B: Receive £0.50 for sure. And if the virtual six-sided die rolls 3 or 4 decode 20 sequences; otherwise, decode 0 sequences

## D.7 Main task - Example of Simultaneous choices

### SCENARIO 27

In this Scenario, you have to make two decisions.

#### DECISION 1

Choose your preferred option:

**OPTION A:** Decode 10 sequences for £1.20

**OPTION B:** Decode 15 sequences for £1.80

☐ OPTION A☐ OPTION B

Before answering, read the next decision

#### DECISION 2

Choose your preferred option:

**OPTION C:** Decode 10 sequences for £0.80

**OPTION D:** Decode 5 sequences for £0.40

☐ OPTION C☐ OPTION D

Proceed to the next Scenario

## **E Follow-up to Study 1: an Attempt to Debias**

In a follow-up study we explored if we can reduce narrow bracketing by making the increasing costs of the additional tasks more salient through a different presentation of choices. We design two new treatments, BEFORE and AFTER, that are identical to the BOTH treatment, except for describing additional sequences as "additional sequences before" or "additional sequences after" the 15 required tasks.

[BEFORE - Scenario 1]

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

	A	B
1) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$4.25
2) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$4.50
3) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$4.75
4) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$5.00
5) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$5.25
6) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$5.50
7) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$5.75
8) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$6.00
9) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$6.25
10) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$6.50
11) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$6.75
12) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$7.00
13) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$7.25
14) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$7.50
15) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$7.75
16) 0 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences before the 15 required for an extra \$8.00

[BEFORE - Scenario 2]

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

	A	B
1) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$4.25
2) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$4.50
3) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$4.75
4) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$5.00
5) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$5.25
6) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$5.50
7) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$5.75
8) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$6.00
9) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$6.25
10) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$6.50
11) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$6.75
12) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$7.00
13) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$7.25
14) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$7.50
15) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$7.75
16) 15 additional sequences before the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences before the 15 required for an extra \$8.00

[AFTER - Scenario 1]

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

	A	B
1) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$4.25
2) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$4.50
3) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$4.75
4) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$5.00
5) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$5.25
6) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$5.50
7) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$5.75
8) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$6.00
9) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$6.25
10) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$6.50
11) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$6.75
12) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$7.00
13) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$7.25
14) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$7.50
15) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$7.75
16) 0 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 15 additional sequences after the 15 required for an extra \$8.00

[AFTER - Scenario 2]

**Note: you are required to decode 15 sequences correctly, in addition to the sequences based on your choices.**

**Choices to make now:** for each choice in this Scenario, choose the preferred option.

**By completing the HIT you will receive \$2.00 plus a bonus depending on your choices.**

	A	B
1) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$4.25
2) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$4.50
3) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$4.75
4) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$5.00
5) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$5.25
6) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$5.50
7) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$5.75
8) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$6.00
9) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$6.25
10) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$6.50
11) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$6.75
12) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$7.00
13) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$7.25
14) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$7.50
15) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$7.75
16) 15 additional sequences after the 15 required for an extra \$4.00	<input type="radio"/>	<input type="radio"/> 30 additional sequences after the 15 required for an extra \$8.00

Figure 7 shows the means by Scenario and by treatment. Treatments BEFORE and AFTER differ from treatment BOTH only by highlighting the number of tasks to do and labeling them as “before” or “after” the baseline tasks. In both BEFORE and AFTER the extra reservation wage is higher than in BOTH, but in both cases this difference is not statistically significant ( $p - values > 0.097$ ). In Appendix E.2, we show, however, that the AFTER treatment is statistically significantly different from BOTH when we limit ourselves to those observations in BOTH that received their information about baseline on the first choice page only, which may indicate a partial success of debiasing.

**Result 6.** *We reject the hypothesis that drawing attention to the later tasks reduces narrow bracketing.*

**Result 7.** *We find tentative support for hypothesis that drawing attention to convexity by highlighting earlier tasks reduces narrow bracketing.*



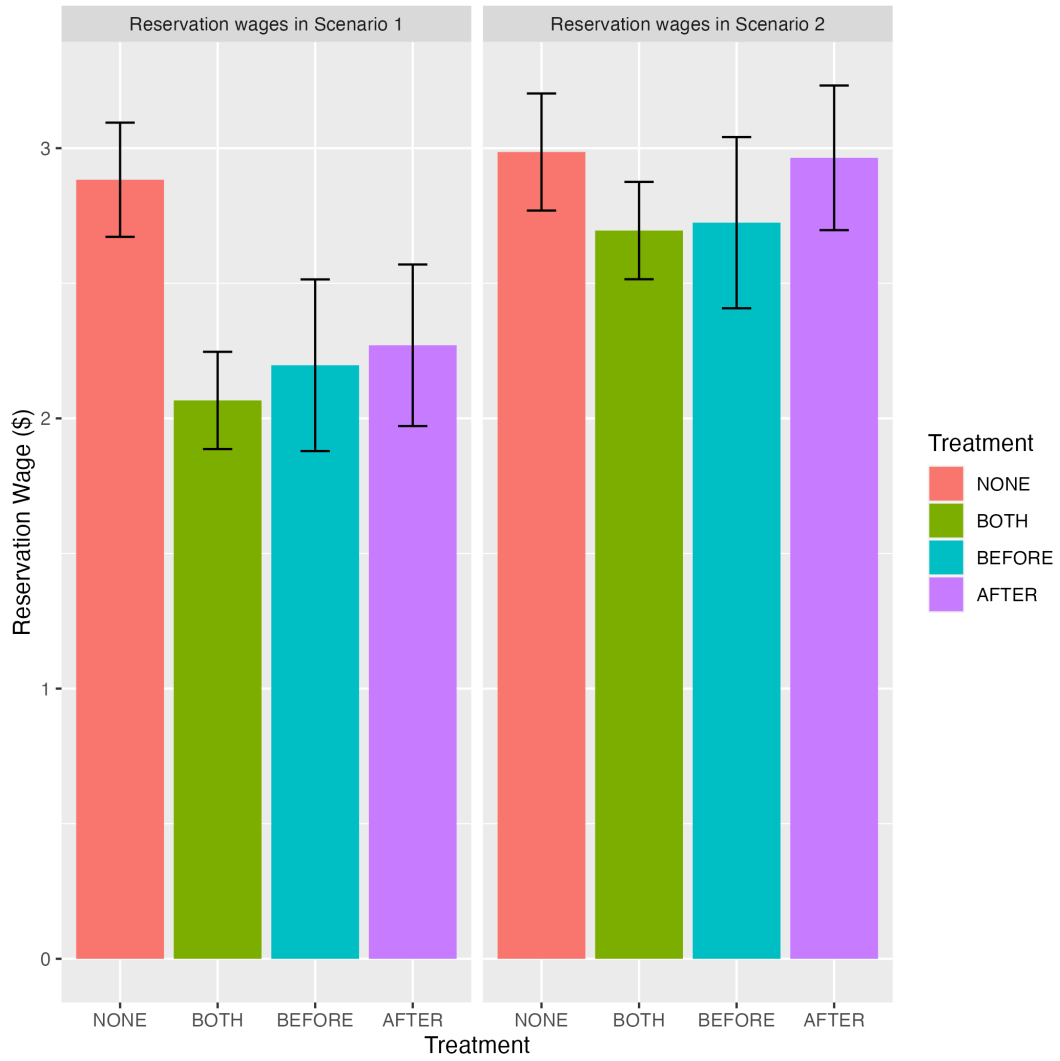


Figure 7: The reservation wages for NONE, BOTH, BEFORE, and AFTER by treatment for each of the two scenarios, along with confidence interval (2 standard errors above and below the estimate).

## E.1 BEFORE and AFTER summary statistics

Table 22: Summary statistics for follow-up treatments

	<b>BOTH</b>	<b>BEFORE</b>	<b>AFTER</b>	<b>p-value</b>
Participants	320	150	152	
Attrition	20.3%	18%	13.8%	0.23
Final Participants	255	123	131	
Share Female	0.38	0.36	0.39	0.68
Age	35	35.1	35.9	0.47
Tediousness	7.45	7.48	7.56	0.93
<b>Inconsistent Choices</b>				
Scenario 1	18.8%	37.3%	33.6%	0
Scenario 2	18.4%	35.3%	34.9%	0

## E.2 Baseline Tasks revealed on the choice page only

Here we report the results when we restrict the data from treatment BOTH to those sessions where the baseline endowment is only revealed on the first choice page, rather than on the page right before, as was inadvertently the case for early sessions.

Table 24: Between-treatment p-values for BOTH, BEFORE, AFTER, and NONE treatments based on two-sided t-tests, treating each individual in each scenario as a single independent observation. Restricted to sessions of BOTH when the endowment is mentioned on choice page first.

Treatments	BOTH	BEFORE	AFTER
<b>Scenario 1</b>			
BEFORE	0.373		
AFTER	0.202	0.735	
NONE	< 0.001	< 0.001	0.001
<b>Scenario 2</b>			
BEFORE	0.263		
AFTER	0.015	0.249	
NONE	0.006	0.175	0.901

Table 23: Between-treatment p-values for main treatments based on two-sided t-test, treating each individual in each scenario as a single independent observation. Restricted to those sessions of BOTH where baseline is revealed only on first choice page.

Treatments	NONE	BOTH
<b>Scenario 1</b>		
BOTH	< 0.001	
MONEY/LOW	< 0.001	0.106
<b>Scenario 2</b>		
BOTH	0.006	
MONEY/LOW	0.120	0.162

Broad bracketing is rejected as before, narrow bracketing is still not rejected.

Now BOTH and AFTER are statistically significantly different in this case, as indicated by Table 24. However, the issues around the different sample population for BEFORE/AFTER remain, given that we collected most of the data after COVID-19 induced lockdowns.

## JEEA Manuscript # 13964: Response to Editor

Dear Vasiliki,

We thank you very much for giving us the opportunity to resubmit our paper for the Journal of the European Economics Association. You and the referees outlined a number of important and major challenges for the resubmission. We have further developed and simplified the theory after the helpful comments we received, and designed and run a new set of experiments tightly linked to this theory, and hope that with these major changes to cover the main concerns that led to the decision to reject and resubmit. Given the new and clear focus of the paper, we have retitled it to "Arbitraging Narrow Brackets".

You stated several substantive points for why the previous version of the paper needed to improve. Concretely, you wrote:

*On one hand I very much like the topic you study and I think the project has great potential; on the other hand, there are important issues raised by the referees that make the paper as is clearly unsuitable for JEEA. I summarize these points (in no particular order):*

- 1. the paper is poorly written, unclear even to specialists –especially the introduction*
- 2. the theoretical contribution is disconnected from the experiment; it is abstract and hard to parse for non-theorists while at the same time it misses some definitions and explanations so it is even hard to appreciate even for theorists*
- 3. it is not clear your experiment tests this theory*
- 4. it is not clear your experiment is simpler and better at testing narrow bracketing compared to the Kahneman-Tversky experiments.*

Based on this, you enumerated the steps we should take to publish this paper at JEEA. We list them here, as well as the changes we have made to address each step.

- 1. Clarify the theory you are testing. Provide in the main text a version that is closer to the experiment – R2 has a nice suggestion on how to do that.*

**Response:** See our response to the next point.

- 2. I personally like that you are working with a choice correspondence that allows for violation of WARP and you do not use preferences as a primitive. In the end however I was confused about the role of Propositions 1 and 2 in your paper – you do not even discuss them in your introduction. I do think there is something so please*

*clarify, explain and include the abstract theoretical setting and proposition in the appendix (and explain them better). The paper should have the version connection closest to the experiment.*

**Response:** Your and the referees comments spurred us on to simplify the theory, which led to a major change in our framework. In the spirit of the suggestion of referee 2, we now define bracketing in terms of maximizing value functions, although one's that allow for context effects and thus for violations of WARP — effectively we are still working with choice correspondences, but without the notational overhead. More importantly however, we realized that our primary result is about buying and selling prices. Concretely, if we observe a person choose  $Y + X - m$  when  $Y$  is available, we say that they buy  $X$  for the price  $m$ ; and similarly, if they pick  $Y$  instead, we say that the person *sells*  $X$  for the price  $m$ . We can then define the highest buying price and the lowest selling price, both of which are in principle observable quantities and are also defined when people violate WARP (under mild assumptions on the value of money). With this new framework, it turns out to be even easier to show when bracketing is unidentified and when it is; to show that it isn't identified when the highest buying price equals the lowest selling price and this price is an additive utility representation of the choices; and that when it is identified, the highest buying price exceeds the lowest selling price, and that a narrow bracketer can be arbitrated by offering them choices such that they buy high and sell low. We highlight in our next response how this leads to a closer connection with our past and new experiments.

3. *Strengthen the experiments along the lines suggested by the referees. This requires running further experiments.*

**Response:** Since we ran additional experiments, we had to streamline the exposition of our earlier experiments, which we have done in line with our updated theoretical framework, highlighting in particular the changes in prices needed to identify bracketing.

In our revised paper, we emphasize the important role played by the price differential in our main study from our experiment, connecting it with our theoretical framework. We have removed some of the treatments that might be interesting in their own right, but not directly related to the theory, such as our analysis of gender differences and our test of whether stating that the extra work had to be done "before" or "after" the baseline work would reduce narrow bracketing.

We further run new experiments that are essentially a direct application of finding price differentials and their potential to arbitrage narrow bracketers. We now also highlight that violations of WARP imply that such price differentials exist, as do violations of expected utility, both of which follow straightforwardly from our results. We use both types of violations to identify bracketing, showing thus practically — not just theoretically — how this can be done, the potential for which we only hinted at (inchoately) in our previous version.

4. *Fully rewrite the paper with a general audience in mind: Picture an intelligent reader that is neither a theorist*

*nor an experimentalist. To that reader you want to quickly explain what narrow bracketing is (they may not know the term), explain why it matters (i.e. why it is a relevant friction—mention pricing, financial choices etc). Explain how it connects to rational inattention and other closely related biases and proceed with your work. Writing a concise yet accessible and nice to read paper is not easy, and for most of us is very time consuming but you have to do this if you want to eventually publish your paper at JEEA.*

**Response:** It is hard for us to be good judges, but we feel that our additional theoretical insights and the focus on prices and arbitrage across choices have helped overcome much of our unclear prose. Given that we only completed the data collection early 2025, we are sure that we can still improve along this dimension, but have put particular emphasis on simplifying the theory, its intuition, and relation with the experimental results.

5. *Thoroughly address all the comments by these excellent referees.*

**Response:** We have addressed all the comments of the referees to the best of our ability, hoping that they agree with the necessary cuts we had to make.

Finally, you highlighted that “[...] we all share a common goal: we want your paper to succeed, and by that I mean being read, cited, used “on the field” by empirical researchers and as stepping stone to other studies. Please keep these thoughts in mind.” We agree with you, and thank you and all the referees for the genuinely thoughtful comments which have enormously improved our paper, and we hope you will agree.

We look forward to hearing from you and the referees.

Best wishes,

Francesco and Marc

## JEEA Manuscript # 13964: Response to Referee 1

(Report beginning with “Narrow bracketing suggests that decisions are considered in isolation rather than together. ...”)

We are extremely grateful for your comments on our submission and your careful reading thereof. Your criticism of the shortcomings of the theoretical framework have spurred us to work harder to distill the key insights, and we thought carefully about the weaknesses that you highlighted in our original design when designing our new experiment, and both are the better for it. In fact, the changes have been so substantial that we also changed the title of the paper to “Arbitraging Narrow Bracketers”.

Below we include all your comments in full as requested by the editor and reply or comment on each.

*Narrow bracketing is an important and understudied topic. The paper makes two main contributions to our understanding thereof. First, its main contribution is to provide a between subjects test of narrow bracketing and broad bracketing. It does so in a real effort setting, another novelty. While they reject broad bracketing, as usual, the setting is rich enough that they can test, and fail to reject (barely), narrow bracketing. This is an interesting finding as very few papers test it; Ellis & Freeman (2020) is the only one of which I am aware. However, I have some concerns with the experiment’s execution and analysis. Second, a theoretical analysis of the bracketing. It does a much better job on the former than the latter which could probably be removed from the paper. Given the high bar at a journal like the JEEA, the paper seems a better fit at a good field journal rather than a top general interest.*

We hope that you will find our updated theoretical analysis simpler and more insightful and that it should not be cut — but we agree especially from the vantage point of our new results that the previous iteration was much more opaque.

*The experiment tests narrow bracketing and broad bracketing in a real effort task. Previous work is generally in the context of risk (TK) or social preferences (e.g. Exley and Kessler 2018). Since behavior often differs with real effort versus monetary tasks (the evidence for non-geometric discounting is stronger with real effort), this contributes to a growing body of knowledge about bracketing. However, it’s also not surprising given the existing evidence that documents failures of broad bracketing. One finds non-broad-bracketing in most settings where you look for it and a lot of behavioral phenomena only make sense with at least some degree of narrow bracketing.*

We share your view that many behavioral phenomena only make sense with some narrow bracketing, and to some extent that narrow bracketing is not too surprising in the context of work choices with varying endowments. Given

this point of view, you will maybe find our results that narrow bracketing is less pronounced for violations of expected utility and of WARP more interesting. We repeat here our Result 3 (page 23):

The majority (50.30%) of identified subjects are classified as narrow bracketers in LINEAR. The share of broad bracketers is higher than the share of narrow bracketers in WARP (38.16% vs 26.32%) and PROBABILITY (46.26% vs 13.61%).

As these results highlight, we find more evidence than usual in favor of broad bracketing, and we speculate in the conclusion why that might be — but since we didn't run any treatments to directly test potential explanations, we cannot do more than speculate at this point.

*The test is simple but straightforward: broad implies that the reservation wage should be the same regardless of whether the endowment of tasks/money is integrated into the choices, while narrow implies that the reservation wage should be the same regardless of the endowment of tasks/money. The authors make these tests, and find that on average broad bracketing is reject and narrow bracketing is (barely) not. I am not sure that failing to reject with a p-value of .106 should be considered strong evidence in favor of narrow bracketing.*

We share your view that we would probably have rejected narrow bracketing in our experiment with sufficiently large data. However, the behavior is much better approximated by narrow than by broad bracketing not just in terms of statistical significance, but in terms of effect size (the difference in means).

*It's unclear that MTurk is the right platform for such an experiment. Most subjects in the mTurk pool try to get through tasks as quickly as possible. So I'm not sure that the experiment is measuring tradeoffs internally consistently, since they are implicitly trading off more work in this experiment with completing another task as soon as they finish.*

We are sympathetic that there are issues with different subject pools. Note however that MTurkers in all our treatments should be equally willing to get through tasks as quickly as possible. Therefore any two treatments that are identical after combining endowments with additional tasks still leads to exactly the same trade-offs, thus the fact that we find treatment differences is evidence that subjects are not naturally bracketing broadly. One can view this as a downside of our study, but we think that such behavior may be more representative of naturally occurring choices, precisely because people in day-to-day situations face similar tradeoffs to the MTurkers.

In our new experiment, we have switched to Prolific, which is perceived by many to be higher-quality than MTurk is now. Moreover, since our new design relies on arbitraging participants in a within-study design, we categorize participants as narrow bracketers when they throw away money compared to what was available. This is clearly a



mistake, no matter how these subjects might behave in other experiments or tasks that they are doing outside of our own experiment.

*This could explain the surprisingly high rate of 23.5% of subjects that were inconsistent across just 2 scenarios.*

It seems that we explained ourselves poorly. The rate of 23.5% of inconsistencies referred to the number of people who filled in the price list in an inconsistent way by having more than a single switching point. For example, they might have been willing to to the work for all payments above (and including) \$2.00 and not willing to do it below, except that they also reported being willing to do it at \$1.50. We dropped people in our original study if they have more than one switch back and forth.

*Also, the randomization into treatments was not done within the session, leading to some systematic differences across when the treatments took place (e.g. some were on Friday and some on Saturday) potentially leading to different opportunity costs of effort.*

This is a justified criticism, and was an error on our behalf. We did however show that when we limited ourselves to those sessions where we did randomize at the session level, the results didn't change, see Appendix C.4 for details. Our new design, being within-subject, does not suffer from this issue.

*Ellis and Freeman (2021) as well as Rabin and Weiszacker (2009) show heterogeneity in bracketing across individuals. I do not think the authors do enough to adjust for this. While this is not inconsistent with the findings, taking heterogeneity seriously makes between subject analysis tricky to interpret. This could also explain why the authors come close to rejecting both. In general, this makes interpreting the results of the experiment problematic, especially since the results are on the border of significance.*

Since this is an important issue, we took this point to heart and decided to run a within-subjects design where we categorize individuals as consistent with narrow bracketing, broad bracketing, both, or neither depending on their behavior. As you correctly foresaw, we find substantial heterogeneity, and in the case where we generate variation in reservation wages by finding violations of non-linear utility — similar to our between-subjects design — we find a strong preponderance of narrow bracketing: 3 times as many subjects are categorized as narrow bracketers than as are categorized as broad bracketers, although this does not hold for the other two domains where we find more broad than narrow bracketers. See Result 3 on page 23.

*The secondary contribution is a theoretical analysis of when narrow bracketing and broad bracketing give identical predictions. This analysis is problematic and should be greatly reduced in size. It attempts to be overly general and as a consequence loses its way. The results require the existence of a dimension*

*that leads to a metric  $M$  that acts essentially as a utility function. In general, the theoretical analysis is orthogonal to the experiment conducted, which seems to require an additively separable utility function to properly interpret. However, that model is never explicitly provided or discussed, despite the fact that objects like a disutility  $d(\cdot)$  are used in the figures.*

You are correct that we, confusingly, referred to disutility  $d(\cdot)$  in our plots, despite not requiring a disutility.

More importantly, your comments and those of the editor and other referees made us realize that our theory needed to be substantially revamped. Further work allowed us to simplify and improve our theory by moving away from observable choice correspondences towards observable prices at which a person is willing to buy or sell a given good. Specifically, we say that a person is willing to buy a good  $X$  for price  $m$  if they choose  $Y + X - \$m$  when  $Y$  was available; and similarly that they sell the good for price  $m$  if in the same situation they choose  $Y$  over  $Y + X - m$ . Since we allow for other options to be available, such prices can be defined even when a person does not have a utility function representing some transitive preferences.

In this new framework, we show that we are not identified when the highest buying price equals the lowest selling price, in which case this price provides indeed a utility representation that is additive. In our context, this can happen if a person has linear utility over work and money, for instance. Conversely, if the highest buying price exceeds the lowest selling price, then a narrow bracketer can be arbitrated: thus if bracketing is identified, it can be identified by using arbitrage tests for which narrow bracketers throw away money. Note that in this case, people need not have a utility function that represents transitive preferences. In fact, if they do not satisfy the weak axiom of revealed preference, then the highest buying price must exceed the lowest selling price. Our new experiments now directly rely on this type of arbitrage test, and highlight it for three domains, including violations of WARP.

We hope that you will agree with us that the new theory is clearer, better connected to the experiment, and of a more reasonable length.

*The paper has a lot of sloppy writing and the exposition could be much better. Some examples: the papers under “monetary choices” on page 1 are about bracketing risks. The authors refer to a “pizza example” on page 3 but there is no other mention of “pizza” in the paper.*

Thank you for pointing out these examples, we now state these papers under “monetary and lottery choices” and keep them together as they are choices between monetary lotteries. We have also removed the reference to the pizza example.

## JEEA Manuscript # 13964: Response to Referee 2

(Report beginning with “This paper provides a theoretical framework for thinking about narrow bracketing. ...”)

Let us first thank you for your feedback and comments, both regarding what is good and what requires improvement. This has pointed us in the right directions for how to change the paper, in particular to improve and clarify the theory, and design a new experiment based on this theory to identify bracketing. The changes have been so substantial that we have renamed the paper to “Arbitraging Narrow Bracketers”.

*The experiment provides compelling support for the proposition that narrow bracketing is a descriptively accurate assumption about how people make choices in the domain of real effort tasks. The design of treatments is careful and allows for compelling population-level tests of both narrow and of broad bracketing. However, the between-subject design does not seem to allow any individual-level inferences about bracketing.*

The point regarding individual-level inferences about bracketing is a valid one, echoed by other referees. Based on these comments, we designed our new experiment as a within-subjects design allowing us to identify bracketing at the individual level, based on the result that if a person is a narrow bracketer whose willingness to work varies across different settings, then we can arbitrage this person, leading them to throw away money. We elaborate on this design further down in response to your specific comments.

*The idea behind narrow bracketing appears in Kahneman and Tversky’s (1979) original prospect theory paper, and this paper’s experimental design essentially ports that idea to the domain of real-effort tasks, while using the norms of economics experiments including appropriate instructions and incentives. Assumptions about bracketing are essential to just about every model in applied economics and how behavioral economists understand and model anomalies – yet is rarely discussed. I think this is largely because while this idea is old, the economics profession has been shockingly slow to get more convincing evidence beyond Kahneman and Tversky’s early work on the subject (as evidenced by the fact that Rabin and Weizsacker’s 2009 extension of their 1981 experiments is considered one of the principal experimental economics references on the subject). My assessment is that demonstrating an aggregate failure of broad bracketing and no aggregate statistical failure of narrow bracketing is important but does not sufficiently take us beyond past work (esp. Kahneman and Tversky 1979) for a top general interest journal like JEEA, I do think the principal contribution of the paper could be conveyed in a “shorter paper” format and would be appropriate for a very good and only slightly more specialized economics outlet.*

Thank you for the positive feedback, we agree that narrow bracketing deserves more attention. We also agree that our previous experiment did not go that far beyond previous work. We do however think that our new revision does

extend substantially further, both theoretically and empirically. We describe this below in response to your next major point.

*The theory and experimental sections felt essentially separate. Results on choice under risk (especially for independent lotteries) in Proposition 2 have little connection to the experiment. The idea behind Proposition 1 that is relevant to the experiment could be stated more simply in the context of a simple two-dimensional model in which people value \$m and work e according to  $v(m,e)$ , and state the conditions on v under which a narrow bracketer and broad bracketer would make the same choices (narrow and broad bracketing could be defined in terms of the model rather than choice). Such a setup isn't really that different from what's in the paper – what I'm really asking is to reduce the amount abstraction and notation so that an experimental reader can get to the insight more easily (in my experience, good theorists can see how to generalize a simple and intuitive result, more casual readers and experimentally/empirically oriented economists can struggle with high levels of abstraction and generality). As the paper points out, there are precedents for the theoretical results – for example parts of Proposition 2 have precedents in prior work by Rabin and Weizsacker (2009) and Mu et al. (2021). The ideas in Section 2.4 that show that bracketing can be identified even for a person who violates WARP do not seem particularly deep as far as I can tell – and in any case, are not formally developed. This is all to say that I think the theory section should be shortened considerably, though the authors could perhaps further develop a theoretical contribution and spin it off into a separate theory paper.*

This sentiment was shared by other referees as well as the editor, and we have to agree that our previous version did not link theory and experiment as tightly enough, and we did not stress this relation sufficiently. After receiving these comments, we worked hard to clarify what the relation was — which we did perceive to be there — and this led us to our new, simplified theoretical framework.

The basic idea is the following. Suppose that a person is willing to do extra work for different amounts of money in different situations — because they have to do the extra work on top of some higher baseline (effectively our experiment in the previous submission), or because their willingness changes as we add a decoy option to the choice set, which is a treatment in our new experiment. Then if a person is a narrow bracketer, we can arbitrage them with respect to these different willingnesses to work. We can offer both of these choices as separate yet simultaneous choices, changing the payments just so that a narrow bracketer is not willing to work for the higher payment, yet willing to work for the lower payment, and thus chooses an option that loses them money overall. In our theoretical framework, we show that if narrow bracketing can be identified at all, then it can also be identified via such arbitrage, and we implement such a design to identify bracketing at the individual level.

Returning to your comments, we have changed our framework in a way where we define bracketing with respect

to a value function  $v$ , although we still allow  $v$  to depend on the choice set and thus we allow for violations of WARP. You were correct to point out that we had not developed formally how violations of WARP can be used to identify bracketing, which we have now done, and we now also show how to practically implement these to test bracketing. While we cannot claim that this insight is necessarily deep — we believe that the deeper insight is the fact that we need price variation (variation in the willingness to work in our context) to identify bracketing and that narrow bracketers can be arbitrated once we identify such variation — the more important practical point is that our approach allows us to measure and identify bracketing in situations where we observe violations of WARP. Thus our framework allows us to separately identify violations of WARP from violations of broad or narrow bracketing, something which is not possible for example in the revealed preference approach of Ellis and Freeman (2024).

We now list your minor comments, using the same numbering as you did:

1. *Why doesn't Figure 2 include the Money treatment? Money vs Both provides an additional test of broad bracketing. I didn't see the Money treatment mentioned until 4.3. Is this an oversight?*

This was not an oversight, it is rather that we had added the "Money" treatment in order to have a robustness test in our original design to test whether endowments of money vs endowments of work are bracketed differently. The reason we did not include it was that this would have led to two tests of broad bracketing, while we had not included an additional test of narrow bracketing. Thus we would have subjected broad bracketing to more strict tests than narrow bracketing. Given that in the revised paper we even more strongly emphasize testing both narrow and broad by identifying a variation in the willingness to work that identifies both, we have not added this treatment in the main text, but include it in Appendix C.10. Figure 6 shows the differences in means across the treatments NONE, BOTH, and MONEY and highlights that broad bracketing is rejected.

2. *Study 2 is interesting even though the intervention had a small effect relative to statistical variation in the data. It suggests that framing different tasks as temporally sequenced vs. as part of a concurrent task-set does not affect behavior (much). This angle could also have been much more developed and I find it interesting.*

We agree and would have liked to expand further on this. However, given that this angle does not link directly with our theoretical framework, we have moved these treatments and results to Appendix E.

3. *I personally prefer tight writing and found Section 5 a bit long.*

We have removed this Section in the new draft given the new focus on arbitrage and since our design does not allow us to cleanly distinguish mechanisms.

## JEEA Manuscript 13964: Responses to Referee 3

(Report beginning with “This study sets out to identify (theoretically and empirically) “bracketing,” which describes how much information people pay attention to when they make decisions.”)

*I enjoyed reading this paper. The experiments were well-done and the analyses were carefully executed. It addresses an important yet relatively understudied topic. Nearly every single behavioral model requires some form of narrow bracketing assumption, yet systematic studies on narrow bracketing are still fairly rare. This paper presents both theory that outlines conditions when narrow bracketing can be identified and experimental tests for narrow bracketing that attempt to illustrate these techniques. The combination of theory and empirics is certainly a strength of the paper, but sometimes the link between the two was a bit unclear. I also thought the authors could do a better job highlighting the specific contribution of each component.*

Thank you for the careful reading and thorough feedback. While it was a lot of work to address the main concerns, it has enormously improved our paper in our opinion. The changes have been so substantial that we renamed the paper to “Arbitraging Narrow Bracketers”.

*Proposition 2 is the main theoretical result and the primary contribution of the paper in light of prior work. Namely, part 4 of the statement, where the author show that the additivity of  $M(\cdot)$  is the unifying feature behind the other three parts, each of which has been discussed and formally presented in the literature. I think this is conceptually interesting, but I don't feel like the paper does a sufficient job discussing its importance for the literature.*

Following your feedback and that of the other referees and the editor, we have worked hard to improve this result. We now highlight how the unifying feature is really the fact that in order to identify bracketing, we need variation in people's buying and selling prices for items, which in our context means variation in people's willingness to work. Thus if the willingness is constant, then for every bundle  $X$  there is a constant price  $P(X)$  which also turns out to be an additive utility representation for the choices. Then as our previous result shows, such additivity implies, among other things, that the utility is linear and satisfies expected utility.

The main insight here is that there are several situations where we might fundamentally be unidentified if people's choices can be rationalized via such a price representation, and that any identification of bracketing thus requires that we observe some variation in buying and selling prices. Different papers in the literature have relied on different sources to generate such variation, Kahneman and Tversky (1979) on the change from risk aversion in gains to risk seeking in losses, Rabin and Weizsäcker (2009) on changes in the degree of risk aversion, our own experiment on

increasing marginal disutility from effort, and so on. See page 9 in Section 2 for a discussion on this. This also allows us to derive a universal arbitrage test for narrow bracketing when we are identified, similar to the one in Rabin and Weizsäcker.

*Section 2 ends with “This identification is how our experimental design works. It relies on the fact that the willingness to pay for the same \*additional\* bundle X varies as we change some fixed endowment bundle Y.” I emphasize \*additional\* because this is key to both the theoretical result and the contribution of the experimental investigation. There have been many papers that have changed the endowment such that choice outcomes are the same but observed that behavior is different (e.g. K&T, 1979). And this indeed the main manipulation in the current study: the NONE and BOTH conditions change the endowment and observe a difference in behavior. Given the extent to which endowment bracketing has been explored in the past in conceptually similar environments, it is not clear to me what the experiment contributes beyond offering another demonstration of endowment bracketing, albeit now in the domain of effort; this translates to difficulty in interpreting the importance of the theoretical result. It seems to me that the paper needs to illustrate the utility of their theoretical approach to identify bracketing in a manner that is conceptually novel and important from an economic perspective.*

You are correct that our primary contribution was to show that people display endowment bracketing in the domain of effort, while showing that people truly bracket narrowly, rather than just showing that they do not bracket broadly (as some papers have done). We consider that both our new theoretical result that bracketing is identified if and only if people display variation in buying and selling prices (here, willingness to work) across choices, which implies that it is identified via arbitrage, is conceptually novel as it can guide researchers to search for such price variation. We discuss the design in our reply to your next comment, and why this framework is helpful in our later reply to your point that ‘smaller treatment differences imply less power’.

*As it currently stands, there is a potentially interesting theoretical result with an experiment that follows closely to other work and does not highlight the current paper’s contribution (as described further in point 1 below). I would urge the authors to design an experiment where, for example, they show a domain of choices that does not satisfy the additivity property and another domain of choices that does. After this, the experiment can illustrate how the theoretical tools can be used to either identify bracketing or not. As noted below, the gender result somewhat gets at approach, but is a bit ad hoc to be satisfactory.*

We thank you a lot for this suggestion, which was a key starting point for our new experimental design. Since whether or not bracketing is identified depends on a person’s choice behavior, our new design works roughly as follows. First, at the individual level, we search for choice pairs such that the subject is more willing to work in one

choice than in the other. We try to generate such variation — which may not be possible! — by either identifying non-linear utility (expecting convexity), violations of expected utility, and violations of WARP. In each case, if we identify such pairs of choices, we know that we are identified, in which case we offer participants these two choices as separate but simultaneous choices in a way that we could arbitrage narrow bracketers, thereby identifying bracketing.

We now list and respond to your main comments:

1. *Experiment: The paper argues that Scenario 2 is a valuable demonstration of the framework because lower convexity implies lower power to detect bracketing. However, it is not clear how the theory informs this test. In the experiment, less convexity coincides with smaller treatment differences. But smaller treatment differences imply less power to detect \*anything\*, whether it is related to bracketing or not. So this result can be also be interpreted as “smaller treatment differences implying less power to detect any effect, whether it is specific to bracketing and the current model or not.” This leaves me wondering how much the current experiment adds to highlighting the utility of the theory.*

You mention an interesting point: that ‘smaller treatment differences imply less power’, which is indeed a general point. In our previous framework and submission we both did not fully grasp ourselves how our theory provided additional insight, and did an even worse job at explaining it. In our new framework this is much clearer.

Concretely, we show that when there is no change in buying and selling prices across choices — in our context, when there is no change in willingness to pay — it is impossible to identify bracketing. Notice that this implies that for some people, we might be *fundamentally* incapable of identifying bracketing, for example if they satisfy WARP, expected utility, and have linear utility. In this case, it is not merely that we are saying that ‘smaller treatment differences imply less power’: we are saying that every researcher would have zero power to identify bracketing, no matter how much data they collect. Moreover, this identification assumption is observable in principle (with infinite data).

Of course, past researchers have intuitively grasped this, Kahneman and Tversky for example by realizing that they needed the switch from risk aversion to risk seeking. What our theory contributes is, first, that such price variation is *necessary* to identify narrow bracketing, which is something that we can test in any given dataset; second, that in order to be identified, we need to either have a violation of WARP or we need to have a non-additive utility function, such as violations of expected utility or of linearity of utility, which can guide our search for price differences; and third, if we are identified, that we can then identify bracketing via an arbitrage test whereby narrow bracketers throw away money. Each of these points provides guidance for researchers interested in identifying bracketing by suggesting ways to *increase* the treatment differences, rather than base their search on intuition, which might have worked for Kahneman and Tversky, but is less promising for Fallucchi and



Kaufmann.

We hope that we make the contributions of our upgraded theory clearer, so you can evaluate better how valuable it is.

2. *Writing: The paper was not written very clearly. I don't mean that with respect to the writing style—the exposition is clear and concise. But I often found myself confused about a concept or example that is being referred to, only to find it being introduced/defined later in the paper. There were also several seemingly contradictory statements for fairly important concepts. Here are some examples that I think should be addressed:*

Thank you for providing so many specific examples!

*a. I was generally fairly lost in the introduction. The theory section is briefly sketched out, but the specific concepts weren't defined and the main identification tests were barely outlined. Results were referred to without the setup, so it was unclear what they followed from (e.g. why one needed changes across choices for identification). The beginning of the theory section was much clearer in outlining the setup; some of that writing should be imported to the introduction, I think.*

We believe that our theory and its exposition is substantially clearer, not least because it is clearer in our own heads. We have spent a lot of time on formalizing our vocabulary, so that "buying a good  $X$  at price  $m$ " has a formal meaning that corresponds well to the usual meaning, and thereby describe our results in a way that is more approachable than our previous submission. In particular, we aimed to provide the gist of the intuition via simple examples already in the introduction, while keeping the examples grounded in the formal framework of Section 2.

*b. On p. 3: What is  $M(\text{cot})$  in footnote 3? What pizza example? What does "our definition does not assume people have preferences" mean?*

$M(\text{cot})$  was a typo, it should have been rendered as  $M(\cdot)$  in Latex.

We had forgotten to update the reference to our example from pizza to work.

"our definition does not assume people have preferences" was meant to state that we allow for violations of WARP.

*c. The gender differences paragraph seemed a bit ad hoc in the introduction. It seems like the paper uses this result to demonstrate when narrow bracketing can be identified or not (i.e. the test has more power for men than for women), but this was not made very clear in the text.*

We have removed most mentions to gender differences, other than in the results section where we simply report some gender differences, see Table 4. Since we now have better examples from our new experiment on situations where we are not identified, we no longer discuss this issue via gender differences.

*d. I don't think Example 2 is correct as currently written. A person is willing to work 1 hour for 2 or 2 hours for 3. Thus the marginal WTW is 1, not 3 as it is currently written.*

You are correct - although we deleted this example when we rewrote the entire introduction and theoretical section.

*e. Section 2 begins with "The examples show how either a violation of WARP or a violation of the additivity of the money metric can identify whether a person is bracketing broadly or narrowly." On p. 6 it is written "... we should not interpret such violations of WARP as evidence for or against either type of bracketing." On p. 9 it is written "The general proof shows that any violates of WARP can be leveraged..." to facilitate identification. These statements seem contradictory. Having read the paper several times I understand what these claims mean, but it should be clarified better earlier on.*

We apologize for the confusion. We believe part of it stemmed from the fact that we did not yet clearly understand ourselves the crucial role of variation in prices, whereas violations of WARP are merely one potential source for such variation. Our current version should be clearer in these aspects.

3. *3. Complexity: The paper claims that one of the contributions of the experiment is that the setting makes it very easy to broadly bracket. Therefore, the experiment is a conservative test of this condition. For example, in the introduction: "Broad bracketing in these situations is either complicated (when combining lotteries), or not necessarily the relevant target (when making social choices). The ease of broad bracketing in the experimental setting is also used to rule out alternative explanations found in other work. I did not agree with this claim. Looking at the experimental instructions, it is actually quite complicated to figure out that the choice outcomes for the NONE and BOTH treatments are the same: the treatments feature different wording, numbers are in different places, and it requires sitting down and combining all those numbers to realize that the outcomes are the same. Now compare this to the narrow bracketing test in the original demonstration of Kahneman and Tversky (1979)—which should be cited. In Problem 11 participants are told they have a 1000 endowment and choose between a lottery (1000, .5; 0, .5) or (500) with certainty. In Problem 12 participants are given 2000 and choose between (-1000, .5; 0, .5) or (-500). These conditions are conceptually isomorphic to the NONE and BOTH treatments in the current paper but, to me, seem much simpler. Participants in K&T also ignore the endowment and make different choices between the Problems, which is a clear violation of broad bracketing. Given the simplicity in this test, it seems like the authors should at least collect explicit data to argue that their setting is simpler than prior work. Otherwise it is not clear to me why complexity is not the explanation for bracketing, since there is no direct evidence for social preferences either (which is the authors' preferred explanation).*

This is a fair criticism, and we pushed the point on complexity too far. We have removed section 5 discussing

potential causes of bracketing, focusing instead on our theoretical contribution and on the new design using arbitrating to identify individual-level bracketing. Our theoretical framework provides a description of bracketing as well as results how to measure bracketing, but is silent on potential mechanisms, so based on your feedback and that of other referees, we decided to stay away from these discussions.

You further had one minor comment: “P. 4: “tests” should be “tasks”.” This part was rewritten completely.

## JEEA Manuscript 13964: Responses to Referee 4

(Report beginning with “This paper first rigorously defines narrow bracketing and provides some theoretical conditions about whether narrow bracketing can be identified.”)

*I like the experiment in the paper. It is beautifully designed and provides evidence for narrow bracketing in a simple but profound matter. So I think that the paper deserves to be published in a good journal, like this one. However, I feel that the current draft of the paper can be improved in a few doable dimensions.*

Thank you for your kind words and for providing actionable feedback. We have made substantial changes to the paper, which also led to changing its title due to the new focus to “Arbitraging Narrow Bracketers”.

*Writing. The paper’s quality of writing is not up to the standard. Overall, I feel that sentence structures and wording in the paper are not standard English. This makes the paper hard to read. This problem is particularly pronounced in the introduction. I rarely bring this up as a main point in referee report, but unfortunately this is a nontrivial problem for this paper (I almost rejected the paper for this reason). I would encourage the authors to spend serious efforts and use professional help to improve the writing.*

We have done our best to improve our writing and while we did not get professional help to improve the writing due to time constraints (we only ran the new experiment in early 2025), we are happy to do so if this is still perceived to be a problem.

*Theory. I feel that the main contribution of the paper is the experiment. The current theory part is somewhat disconnected with the experiment. It introduces many unnecessary notations/complications that are irrelevant to the experiment. I think that the authors can significantly simplify the theory part to make the main point relevant for the experiment: if disutility of work is convex, narrow bracketing is identified; if disutility of work is linear, narrow bracketing is unidentified. – Specifically, some propositions (propositions 1 and 2) seem irrelevant for the experiment. They focus on the “unidentified” case, while the experiment focuses on the identified case.*

You highlight a point raised by other referees and the editor, and we agree that our theoretical contribution wasn’t clear and impactful enough. Prompted by these criticisms, we have worked hard to improve the theory, which led to new insights. This in turn led to a clearer theory showing that if bracketing is identified, then it can be identified by arbitrage, with narrow bracketers falling prey to such arbitrage. We further designed a new experiment based on arbitrage.

We have also clarified our results on when we can identify bracketing, showing that this requires that we identify changes in buying and selling prices, that is, changes in willingness to work in our setting. In our new experiment, we

show how such changes can be driven by convexity (or non-linear disutility), but also by violations of expected utility or of the weak axiom of revealed preference (WARP). In this experiment, we focus on those subjects for whom we can identify such variation in price changes, however without our theoretical framework and the results on identification, it wouldn't be clear who these individuals are. We hope that you find the new theory more interesting, and that we relate it better to our experiments, including our previous one.

*One example of the authors is quite unnatural.*

*"We can identify bracketing by offering the following two (joint) choices: work for \$3.5 or do nothing (X); and work for \$2.5, study for \$2, or do nothing (Y). If the person brackets narrowly, they choose to work for \$2.5 in the second choice set, but turn down work for \$3.5 in the first choice set – they work for little when they could work for more, which a broad bracketer would never do."*

*I doubt that people will choose to work for \$2.5 in the second choice set, but turn down work for \$3.5 in the first choice set.*

We removed this example. In our new draft, we have tried to use examples similar in structure to those we have in our experiment, albeit with amounts that are larger to illustrate the points more simply.

*Discussion of Causes of Bracketing. I agree with authors' conclusion that the experimental findings come from some cognitive "mistakes," i.e., people fail to realize that they should combine outcomes in this situation. But I feel that Section 5 has room for improvement. Writing there needs to be streamlined. The current Section 5 is not focused and is hard to read. Moreover, I do not agree with the authors' interpretation of the literature regarding cognitive costs of broad bracketing. First, noise in information is often viewed as a proxy for cognitive noise/uncertainty, instead of narrowly defined information friction. Second, some of those papers do not impose that cognitive noise is rationally chosen given the cognitive costs (as in rational inattention models).*

Given how much our new paper has expanded with the addition of our new experiment, we decided to remove Section 5 discussing the causes of bracketing. Our framework is silent on causes of bracketing, defining bracketing as a behavior independently of its cause, and none of our treatments can speak clearly and convincingly to mechanisms. In order to distinguish between the models that you mention — cognitive costs that can stem from rational inattention as well as broader forms such as cognitive uncertainty — we would have to model mechanisms more carefully, which is beyond the scope of the current paper. We hope that this decision meets with your approval.