Stochastic Dominance and Demand for Surprise*

Séverine Toussaert†

March 26, 2024

Abstract

Decision theorists usually take a normative view on stochastic dominance: a decision maker who chooses a dominated lottery must be making a mistake. This paper provides evidence that stochastic dominance violations may naturally occur in situations where anticipatory utility is high, such as going on a holiday trip. In such a situation, the decision maker may trade the certainty of going to their favorite destination for the excitement of not knowing where they will go. I document this phenomenon in an experiment in which participants make choices between a sure destination and a "surprise lottery" over holiday trips, with the lottery outcome revealed close to the date of travel. I vary lottery characteristics to understand when violations are most likely to occur and analyze their properties. I discuss the implications for the design of goods with a surprise element and for the modelling of anticipatory utility.

JEL classification: C93, D01, D81, D91

Keywords: surprise; dominance; violation; anticipation; experiment

*This project greatly benefited from conversations with Johannes Abeler, Sandro Ambuehl, Aurélien Baillon, Miguel Ballester, Lukas Bolte, Emir Kamenica, George Loewenstein, Pietro Ortoleva, Collin Raymond, and many others. I would like to extend my thanks to participants at many seminars and workshops. I am very grateful to Gessienne Grey, Matthew Henderson, Stefania Merone, Jessica Milligan and Le Wu for their outstanding research assistance, and to St John's College (Oxford) and the Economic and Social Research Council (ES/V003461/1) for their financial support. Last but not least, I would like to thank Felix Tan and Zelia Leong from the (former) travel company Anywhr and Charlotte Cole from `travelcounsellors.co.uk` for their logistical support.

†Department of Economics, University of Oxford: severine.toussaert@economics.ox.ac.uk.

1 Introduction

> *[...] dominance rules command virtually unanimous assent [...] even from those who sometimes violate them in practice [...]. If a theory of decision under uncertainty is to be consistent with any of the large body of economic theory which has already been developed [...] it must satisfy these rules.*
>
> John Quiggin (1982)

When it comes to decision-making under uncertainty, stochastic dominance is considered to have real normative bite: a decision maker (DM) who violates such rule must be making a mistake. This stance has a strong theoretical basis, via money-pump arguments: a DM who violates stochastic dominance in the space of monetary gambles may end up with no money through a sequence of trades. Due to its normative appeal, extensive efforts have been made to preserve this property when relaxing expected utility theory. For instance, models of rank-dependent probability weighting such as rank-dependent expected utility or cumulative prospect theory (Quiggin, 1982; Tversky and Kahneman, 1992) allow for non-linear probability weighting while ensuring that first-order stochastic dominance is preserved, a major reason for their success relative to the original prospect theory model of Kahneman and Tversky (1979).

This normative property also tends to perform well descriptively, for stochastic dominance violations are typically rare. When violations do occur, they usually involve complex gambles with more than two outcomes, suggesting that they might be the result of a cognitive error (Birnbaum, 2005; Nielsen and Rehbeck, 2022; Puri, 2023). On the other hand, straightforward stochastic dominance violations are rarely revealed directly and often require transitivity arguments or between-subject designs. For instance, Gneezy, List, and Wu (2006) document violations of stochastic dominance in the direction of a preference for certainty using a between-subject design, reporting that violations disappear in a within-subject comparison.

While the normative and descriptive appeal of stochastic dominance is clear in the contexts typically studied, it is by no means universal. If a DM exhibits other concerns than maximizing material gains, violations of stochastic dominance could emerge as the natural expression of those concerns. One such context is when the DM derives utility not only from material consumption but also from anticipation.[1]

[1] Other contexts in which violations might result from a trade-off between maximizing material gains and addressing non-material concerns are when the DM faces the risk of disappointment or regret, or when they are affected by social or self-image concerns.

By anticipatory utility, I will broadly refer to the utility derived ahead of the realization of an event (e.g., prize draw, consumption of a good, enjoyment of an activity) from simply *thinking* about the outcome. In such contexts, the DM might trade the guarantee of their favorite outcome for the excitement of not knowing what it is and the pleasure of wondering what it might be. In other words, *surprise enjoyment* might lead to violations of stochastic dominance in the direction of a preference for randomization. The objective of this paper is to document this phenomenon in an experiment and to derive its theoretical and practical implications.

The market for goods and services with a surprise element has expanded over the years, so much so that it is now possible to buy a surprise box with pretty much anything in it.[2] While this market expansion suggests a positive demand for surprises, such products often come with a bundle of characteristics that could make them attractive to customers for other reasons than the surprise, such as personalization (products tailored to individual tastes), delegation of search and decision-making costs, price discounts, and novelty elements (ability to discover new products or events). The presence of these confounding factors makes it difficult to estimate from observational data how much customers value the surprise component of such goods. I thus employ experimental variation to address this identification problem.

To this end, I consider a positive consumption event that typically generates a lot of anticipatory utility: going on a holiday trip. In the experiment, participants are presented with a list of 10 European destinations, which they can possibly travel to for 4 days and 3 nights. I measure their preferences over the 10 destinations via a ranking combined with their valuation of each destination. Based on the elicited preference ordering, I then present participants with decision problems of the type:

OPTION A: Go to *Prague* for sure.

OPTION B: Take a 50/50 chance of either *Prague* or *Maastricht*.
Destination revealed the week of departure.

For a DM with the strict preference $Prague \succ Maastricht$, choosing Option B reveals a violation of stochastic dominance. The experiment systematically varies three

[2] The unboxing of surprise items has itself become a YouTube phenomenon, especially among young children, with one video generating over 94 million views: https://www.npr.org/2014/09/13/348241139/surprise-kids-love-unboxing-videos-too. Beyond, many goods and services have a surprise component already built in, such as books, movies, live sports events, dinner experiences, or games. Surprise rewards and loot boxes have also become key features of video game design.

features of the lotteries: (i) the number of destinations in the support of the lottery; (ii) the rank of each destination in the DM's preference ordering; (iii) the probability distribution over destinations. I exploit this variation to study the structure of the observed violations and categorize them using simpler monotonicity properties. Violations are linked to a class of utility representations in which the DM trades off their expected value from the trip with the uncertainty of not knowing where they will go.

To better understand randomization behavior, participants are also offered to design their favorite lottery and choose their preferred date for resolving the uncertainty. If violations are only due to factors such as noise, indecision, or regret, participants should not choose to delay the resolution of uncertainty. Finally, I ask whether violations subsist in the world of money by considering decision problems in which the trips are replaced with their valuations. My main findings are as follows:

1. On average, respondents violate stochastic dominance in favor of randomization in 1 out of 5 decision problems. However, individual heterogeneity is large.

2. Violations in favor of certainty occur as well, but less frequently; the two tendencies are negatively correlated, suggesting they reflect different phenomena.

3. Violations are economically significant: participants sacrifice an average of £46 to preserve the surprise, about 11% of the market value of a trip.

4. Decision noise or measurement error cannot account for the observed violations as they have a specific structure: violations occur more often for lotteries with a higher entropy i.e., when the outcome is more uncertain.

5. Randomization behavior comes with a preference for delay: 74% of those who prefer a lottery choose to postpone the realization of uncertainty by at least two days, with about half giving up £5 or more to preserve the surprise.

6. Violations drop to nearly zero for lotteries over monetary prizes, suggesting that the experiential value and/or multidimensionality of the good play a key role.

Taken together, these findings pose a challenge for existing theories of choice under risk and uncertainty. The vast majority of deterministic models satisfy first-order stochastic dominance, including expected utility and non-expected utility models

(Quiggin, 1982; Gul, 1991; Tversky and Kahneman, 1992; Cerreia-Vioglio, Dillenberger, and Ortoleva, 2015). At the exception of models assuming utility from gambling (Diecidue, Schmidt, and Wakker, 2004), deterministic models that allow for violations generate a preference for certainty under standard parametrizations (Kahneman and Tversky, 1979; Bell, 1985; Loomes and Sugden, 1986; Kőszegi and Rabin, 2007). While standard stochastic choice models allow for violations, they cannot explain the specific violations observed (i.e., for higher-entropy lotteries) or the preference for delayed resolution of uncertainty. Overall, existing models cannot jointly account for all facts presented here without introducing new degrees of freedom. The findings of this paper thus call for models that explicitly take into account the domain of randomization, including the multi-attribute nature of the choice objects.

Besides the theoretical interest, this research has practical relevance. First, if people value goods with a surprise element, then rewarding good behaviors with surprises could have powerful motivational effects. Because surprises create an information gap that triggers curiosity (Golman and Loewenstein, 2018), rewarding a DM with access to non-instrumental information may foster motivation and goal pursuit. Speaking to this point, Shen, Fishbach, and Hsee (2015) find that people expand more resources to obtain an uncertain reward than a certain reward, even if the former is dominated.[3] Despite this, information preferences have received much less attention in the literature on behavioral incentive design than other aspects of non-standard preferences pertaining to time, risk, or social considerations (Levitt, List, Neckermann, and Sadoff, 2016; Carrera, Royer, Stehr, and Sydnor, 2020). This research suggests that surprises, if optimally structured, could increase motivation at a lower cost.

Second, the present findings have potentially important implications for welfare. One question debated every year concerns the welfare effects of Christmas gift giving.[4] A provocative article published in 1993 proposed that Christmas gifts are a source of deadweight loss because recipients often derive less value from gifts than the amount of money spent to purchase them (Waldfogel, 1993). While this interpretation has been contested subsequently (Solnick and Hemenway, 1996; List and Shogren, 1998; Ruffle and Tykocinski, 2000), no work so far has pushed forward, let alone attempted

[3]Relatedly, Corgnet, Gächter, and Hernán González (2020) find that persistence in an effort task for a negligible reward is higher if the reward is stochastic, and the effect is stronger for high-entropy (more unpredictable) rewards. This effect is attributed to an increase in stress and attentional costs.

[4]See for instance this New York Times article: https://www.nytimes.com/2014/12/21/upshot/an-economist-goes-christmas-shopping.html.

to price, the utility benefits of anticipating gifts and waiting to unwrap them.[5] Anticipatory utility concerns might also partly explain why people enter sweepstakes or purchase lottery tickets even when their chances of winning are virtually zero. For instance, lottery players might derive utility benefits from dreaming about the possibility of a better life or the excitement of watching a televised draw. If utility from anticipation is larger for lower-income households e.g., reflecting a desire to escape from everyday-life anxieties, then conclusions about lotteries imposing a tax on the poor might not be warranted (Lockwood, Allcott, Taubinsky, and Sial, 2021).

My work lies at the intersection of several empirical and theoretical literatures. There is (well, surprisingly) little work in economics on preferences for surprise. At a theoretical level, Ely, Frankel, and Kamenica (2015) formalize the notions of "surprise" and "suspense" in a model where the DM derives utility from their beliefs changing over time, as the uncertainty gets resolved. Unlike Ely et al., I only consider one-shot resolution of uncertainty, thus ignoring the dynamics of belief updating; on the other hand, I allow preferences to depend not only on belief uncertainty but also on material outcomes, and study how the two are traded off depending on the measure of uncertainty used. By considering the Shannon entropy as one such measure, this paper also connects to the notion of "surprisal" in information theory (Shannon, 1948), which quantifies the uncertainty of an event based on its probability.[6]

At an empirical level, multiple attempts at quantifying surprise and measuring its hedonic value have been made in psychology, neuroscience, and computational biology (Modirshanechi, Brea, and Gerstner, 2022). Within behavioral science, and perhaps most connected to this work, is a recent marketing paper on the demand for "mystery consumption" goods (Buechel and Li, 2022). In a series of studies with different products, this paper shows that people tend to prefer goods with a mystery or surprise element to goods of known content that have equal expected value. I study the revealed preference implications of such a preference for surprise, by showing that it can translate into violations of stochastic dominance and characterize their properties.

More generally, this work speaks to a growing literature on preferences for randomization (Agranov and Ortoleva, 2022). At a theoretical level, most models were

[5] In addition, gift givers may themselves derive utility from the anticipation of surprising others, especially if they project their own preferences for surprise onto the recipients of their gifts.

[6] In a different vein, the desire to surprise others and its strategic implications have also been examined within the framework of psychological games (Geanakoplos, Pearce, and Stacchetti, 1989).

written with monetary outcomes in mind; as such, they usually satisfy stochastic dominance. This study moves outside of the traditional domain of money to consider how studying experiential goods with multiple attributes might expand the scope for randomization behavior. At an empirical level, many papers have documented a preference for coin flipping that cannot be rationalized by simple indifferences or mistakes (Agranov and Ortoleva, 2017; Dwenger, Kübler, and Weizsäcker, 2018; Levitt, 2020; Zhang and Zhong, 2020). Besides motives such as indecision, regret, or disappointment, I propose anticipatory utility from surprises as another driver to consider.

Finally, this paper speaks to a large literature incorporating the role of thoughts and feelings as a key driver of decisions (Schelling, 1987). Within this literature, many papers have sought to understand how anticipatory feelings might affect the dynamics of consumption decisions, the demand for or against receiving information, or people's beliefs about future outcomes (Loewenstein, 1987; Caplin and Leahy, 2001; Thakral and Tô, 2022). Most empirical studies have been conducted in the lab, often with monetary amounts or negative consumption events such as medical tests or electric shocks (Ganguly and Tasoff, 2017; Engelmann, Lebreton, Salem-Garcia, Schwardmann, and Weele, 2022; Falk and Zimmermann, 2023; Masatlioglu, Orhun, and Raymond, 2023). Instead, this paper studies a positive consumption event in a field setting, suggesting dreaming as a source of utility.

The rest of this paper is organized as follows. Section 2 introduces the theoretical framework used to structure the experiment and analyses. Section 3 describes the experimental design. Section 4 presents findings on the measurement of preferences over destinations, while Section 5 discusses the prevalence, size, and shape of stochastic dominance violations in the experiment. Section 6 investigates mechanisms. Section 7 concludes with a discussion of the results and open issues left for future work. Additional results and metadata can be found at https://osf.io/ya7x6/.

2 Theoretical framework

Setup I consider a DM who trades off the expected outcome of a lottery (i.e., what destination they will go to) with the excitement or anxiety produced by the uncertainty of not knowing the outcome. Let $X := \{x_1, x_2, ..., x_n\}$ be a set of n possible outcomes and let $\mathbf{p} := (p_1, p_2, ..., p_n) \in \Delta^n(X)$ be a probability measure on X, with p_k the probability of x_k. The interior of the simplex is $\text{int}\Delta^n(X) :=$

$\{\mathbf{p} := (p_1, p_2, ..., p_n) | \sum_{k=1}^{n} p_k = 1,\ p_k > 0\ \forall k = 1, ..., n\}$ and the support of $\mathbf{p}$ is $\text{supp}(\mathbf{p}) := \{x_k \in X \mid p_k > 0\}$. Below I will abuse notation and identify with x_k the dirac measure $\delta_{x_k} = 1$ if x_k is realized. When useful, I will enumerate a lottery $\mathbf{p}$ with its prizes as $(..., x_j, p_j; ...; x_k, p_k; ...)$. The primitive is a weak order $\succeq$ on $\Delta^n(X)$.

Definitions For any $\mathbf{p}, \mathbf{q} \in \Delta^n(X)$, say that $\mathbf{p}$ *stochastically dominates* $\mathbf{q}$, denoted $\mathbf{p} \rhd_{SD} \mathbf{q}$, if $\mathbf{p}\{x \mid x \preceq x_k\} \leq \mathbf{q}\{x \mid x \preceq x_k\}$ for all $x_k \in X$ (with at least one strict inequality). The DM *satisfies stochastic dominance* if $\mathbf{p} \rhd_{SD} \mathbf{q} \ \Rightarrow \ \mathbf{p} \succ \mathbf{q}$. In the experiment, I will focus on decision problems in which either $\mathbf{p}$ or $\mathbf{q}$ is a degenerate outcome, x_i. Let $\mathcal{D}^+$ denote the set of dominance problems in which the lottery is dominated by the sure outcome (i.e., $x_i \succeq x$ for all $x \in \text{supp}(\mathbf{p})$ and $x_i \succ x$ for some $x \in \text{supp}(\mathbf{p})$). Similarly, denote by $\mathcal{D}^-$ the set of problems in which the lottery is dominant. I will use d, d', d'' to refer to generic members of $\mathcal{D}^+ \cup \mathcal{D}^-$. The DM reveals a *preference for randomization* (*certainty*) at some $d \in \mathcal{D}^+$ $(\mathcal{D}^-)$ if $\mathbf{p} \succeq x_i$ $(x_i \succeq \mathbf{p})$.

Utility representation The DM has a valuation for each outcome $x_k \in X$ denoted by $v_k := v(x_k) \in [0, \bar{v}]$. Letting $\mathbf{v} := (v_1, v_2, ..., v_n) \in [0, \bar{v}]^n$ be the corresponding vector of valuations, I assume that the DM's preferences $\succeq$ on $\Delta^n(X)$ can be represented as the outcome of the following maximization problem:

$$\max_{\mathbf{p} \in \Delta^n(X)} U_{\alpha,\Psi}(\mathbf{p}) := \mathbf{p} \cdot \mathbf{v} + \alpha \Psi(\mathbf{p}, \mathbf{v})$$

where $\Psi(.) \geq 0$ summarizes the uncertainty contained in $\mathbf{p}$ and α is a taste parameter. The parameter α could a priori take any value to allow for both attraction $(\alpha > 0)$ or aversion $(\alpha < 0)$ towards the unknown.[7] Depending on the shape of Ψ, the DM may violate stochastic dominance for $|\alpha|$ large enough i.e., favor a lottery over its best outcome if $\alpha > 0$, or reject a lottery in favor of its worst outcome if $\alpha < 0$.

I allow the value of uncertainty Ψ to depend on $\mathbf{p}$ (e.g., a uniform distribution brings more uncertainty) and/or $\mathbf{v}$ (e.g., more spread in the possible outcome values

[7] Ψ could refer to the discounted sum of per period excitement (anxiety) generated by the surprise until the date t at which the uncertainty is resolved, $\Psi(\mathbf{p}, \mathbf{v}; t) = \sum_{\tau=0}^{t} \delta^\tau \psi_\tau(\mathbf{p}, \mathbf{v})$. In principle, the excitement or anxiety produced each period, $\psi_\tau(.)$, might differ over time e.g., (i) if the DM acquires new information and updates their beliefs, or (ii) if proximity to the date at which the uncertainty is resolved raises attention to the unknown so that $\psi_t \geq \psi_\tau,\ \forall\ \tau < t$. For our purposes, it will be enough to consider the aggregate measure Ψ and ignore the specific resolution date t. Implicit in the representation is the assumption that $\Psi = 0$ if $t = 0$ (uncertainty resolved immediately).

makes a lottery more uncertain). I will primarily, but not exclusively, focus on functions of the form $\Psi(\mathbf{p}, \mathbf{v}) = \psi(H(\mathbf{p}, \mathbf{v}))$ where $\psi(0) = 0$, $\psi'(.) > 0$, and $H(\mathbf{p}, \mathbf{v})$ is a "valid" measure of uncertainty in the sense of Frankel and Kamenica (2019) i.e., a measure such that (i) $H(x_k, \mathbf{v}) = 0$ for all $x_k \in X$ (uncertainty is null for degenerate outcomes); (ii) H is globally concave in $\mathbf{p}$ (uncertainty increases when mixing two distributions), and H is smooth. To give examples, a valid measure of uncertainty that depends on both $\mathbf{p}$ and $\mathbf{v}$ is the variance in valuations $H(\mathbf{p}, \mathbf{v}) = \sum_{k=1}^{n} p_k (v_k - E_{\mathbf{p}}(v))^2$; one that only depends on $\mathbf{p}$ is the Shannon entropy $H(\mathbf{p}) = -\sum_{k=1}^{n} p_k \ln(p_k)$. For reasons that will become clear below, I allow H to enter non-linearly in the DM's utility through the monotone transformation ψ.

2.1 Monotonicity properties

To categorize violations due to preference for randomization ($\alpha > 0$) and relate them to features of the utility representation, I will test whether choices satisfy two simpler monotonicity properties. The first property requires that, for any two outcomes in a lottery $\mathbf{p}$, increasing the probability weight on the worse outcome cannot make the DM now prefer the lottery over the sure outcome it is compared to (all else unchanged). I call this property P-MON for "monotonicity in probabilities" since it focuses on the probability weights assigned to two outcomes kept fixed. Formally:

P-MON: For any $n \geq 2$, $\mathbf{p} \in \Delta^n(X)$, and x_i, x_j, $x_k \in X$ s.t. $x_j \succ x_k$,

$$x_i \succ (..., x_j, p_j; ...; x_k, p_k; ...) \Rightarrow x_i \succ (...x_j, p_j - \epsilon; ...; x_k, p_k + \epsilon; ...)$$

for all $\epsilon > 0$ s.t. $\mathbf{p}_\epsilon := (..., p_j - \epsilon, ..., p_k + \epsilon, ...) \in \Delta^n(X)$ and $\text{supp}(\mathbf{p}) = \text{supp}(\mathbf{p}_\epsilon)$.

The second monotonicity property, coined X-MON, is imposed on the outcomes while keeping the probabilities unchanged: for any lottery $\mathbf{p}$, this property requires that replacing a given outcome in $\mathbf{p}$ with a worse outcome cannot make the DM now prefer the lottery over the sure outcome it is compared to (all else unchanged):

X-MON: For any $n \geq 2$, $\mathbf{p} \in \Delta^n(X)$, and x_i, x_j, x_k, $\tilde{x}_k \in X$ s.t. $x_k \succ \tilde{x}_k$,

$$x_i \succ (..., x_j, p_j; ...; x_k, p_k; ...) \Rightarrow x_i \succ (..., x_j, p_j; ...; \tilde{x}_k, p_k; ...)$$

A counterpart to these properties can be easily formulated for violations in the direction of a preference for certainty ($\alpha < 0$), by requiring the DM to still prefer a lottery over a sure outcome if the lottery assigns a higher weight to better ranked outcomes.[8] I will use these monotonicity properties to provide non-parametric tests of several classes of preferences compatible with a $U_{\alpha,\Psi}$-representation. Clearly, a standard DM ($\alpha = 0$) will satisfy all properties, while a DM with $\alpha > 0$ (resp. $\alpha < 0$) may violate P-MON and/or X-MON (or their counterpart). Besides allowing to sign α, the tests provide valuable information on the shape of Ψ i.e., how it depends on $\mathbf{p}$ and/or $\mathbf{v}$.

To make this point, I introduce additional notations. Let $\mathcal{B}^+$ and $\mathcal{B}^-$ be the subset of dominance problems with binary-outcome lotteries i.e., problems of the type $\{x_i, (x_j, p; x_k, 1-p)\}$. Within this subset, it will be useful to distinguish *weak dominance problems*, in which $x_i \sim x$ for some $x \in \text{supp}(\mathbf{p})$, denoted $\mathcal{B}^+_\sim$ and $\mathcal{B}^-_\sim$, from *strict dominance problems*, in which $x_i \succ x$ or $x \succ x_i$ for all $x \in \text{supp}(\mathbf{p})$, denoted $\mathcal{B}^+_\succ$ and $\mathcal{B}^-_\succ$. For example, assuming $x_1 \succ x_2 \succ x_3$, the decision problem $\{x_1, (x_1, p; x_2, 1-p)\}$ is an element of $\mathcal{B}^+_\sim$, while $\{x_3, (x_1, p; x_2, 1-p)\}$ is an element of $\mathcal{B}^-_\succ$.

2.2 Preference classes

I now discuss how the above behavioral properties relate to the shape of Ψ. Throughout, I focus on the case $\alpha > 0$; obvious corollaries hold for $\alpha < 0$. Since these properties are trivially satisfied if the DM never violates stochastic dominance, I assume that α is large enough for $\succeq$ to violate stochastic dominance at some $d \in \mathcal{D}^+$.

Observation 1. Assume $\Psi(\mathbf{p}, \mathbf{v}) = \psi(H(\mathbf{p}, \mathbf{v}))$ where H is a valid measure of uncertainty, and ψ is such that $\psi(0) = 0$ and $\psi'(.) > 0$. Then $\succeq$ may violate P-MON at some $(d, d') \in \mathcal{D}^+ \times \mathcal{D}^+$.

In particular, a DM may switch from a sure outcome x_i to a dominated lottery $(x_j, p_j; x_k, p_k; ...; x_K, p_K)$ where $x_i \succ x_j \succ x_k \succ ... \succ x_K$ if ϵ-weight is reallocated

[8] Formally, the corresponding monotonicity properties are:

P-MON*: For any $n \geq 2$, $\mathbf{p} \in \Delta^n(X)$, and x_i, x_j, $x_k \in X$ s.t. $x_j \succ x_k$, $(..., x_j, p_j; ...; x_k, p_k; ...) \succ x_i \Rightarrow (...x_j, p_j + \epsilon; ...; x_k, p_k - \epsilon; ...) \succ x_i$, for all $\epsilon > 0$ s.t. $\mathbf{p}_\epsilon := (..., p_j + \epsilon, ..., p_k - \epsilon, ...) \in \Delta^n(X)$ and $\text{supp}(\mathbf{p}) = \text{supp}(\mathbf{p}_\epsilon)$.

X-MON*: For any $n \geq 2$, $\mathbf{p} \in \Delta^n(X)$, and x_i, x_j, x_k and $\tilde{x}_j \in X$ s.t. $\tilde{x}_j \succ x_j$, $(..., x_j, p_j; ...; x_k, p_k; ...) \succ x_i \Rightarrow (..., \tilde{x}_j, p_j; ...; x_k, p_k; ...) \succ x_i$

from x_j to x_k so that $\Psi(\mathbf{p}_\epsilon, \mathbf{v}) > \Psi(\mathbf{p}, \mathbf{v})$. A violation of P-MON then occurs if

$$\mathbf{p} \cdot \mathbf{v} - \epsilon(v_j - v_k) + \alpha\Psi(\mathbf{p}_\epsilon, \mathbf{v}) > v_i > \mathbf{p} \cdot \mathbf{v} + \alpha\Psi(\mathbf{p}, \mathbf{v})$$

i.e., the utility boost from uncertainty $\alpha\,(\Psi(\mathbf{p}_\epsilon, \mathbf{v}) - \Psi(\mathbf{p}))$ compensates the material loss $\epsilon(v_j - v_k)$. How pervasive those violations are however depends on the shape of ψ and the type of problem. In fact, if ψ is weakly concave, then $\succeq$ cannot violate P-MON on the set $\mathcal{B}^+_\sim$ of weak dominance problems with binary-outcome lotteries:

Observation 2. Assume $\Psi(\mathbf{p}, \mathbf{v}) = \psi(H(\mathbf{p}, \mathbf{v}))$ where H is a valid measure of uncertainty, and ψ is such that $\psi(0) = 0$, $\psi'(.) > 0$ and $\psi''(.) \leq 0$. Then $\succeq$ must satisfy P-MON on $\mathcal{B}^+_\sim$ for all $\alpha > 0$.

In other words, if ψ is globally (weakly) concave, then the DM will switch at most once from x_i to $(x_i, p; x_j, 1-p)$ as p increases for any $x_i \succ x_j$. For example, if $\Psi(\mathbf{p}) = -\sum_{k=1}^n p_k \ln(p_k)$ (i.e., $\psi = I$), then the DM cannot exhibit the preference $(x_1, 0.5; x_2, 0.5) \succ x_1 \succ (x_1, 0.9; x_2, 0.1)$. On the other hand, P-MON violations may occur on $\mathcal{B}^+_\sim$ by convexifying ψ (e.g., for $\psi(H) = H^\gamma$ with $\gamma > 1$). They can also occur when ψ is concave if $|\text{supp}(\mathbf{p})| \geq 3$ or if dominance is strict (i.e., the best lottery outcome is strictly worse than the sure option). Intuitively, a new degree of freedom must be introduced to allow for such monotonicity violations (see Appendix F).

Observation 3. If Ψ is independent of $\mathbf{v}$, then $\succeq$ must satisfy X-MON for all $\alpha > 0$.

This follows immediately from the representation: if some $x_k \in \text{supp}(\mathbf{p})$ is replaced by another $\tilde{x}_k$ such that $x_k \succ \tilde{x}_k$, $\Psi(\mathbf{p})$ is unchanged but the expected material gain from the lottery falls by $p_k(v_k - \tilde{v}_k)$. Thus, replacing a lottery outcome with a worse option cannot induce the DM to now prefer the lottery. Relatedly, if Ψ is independent of $\mathbf{v}$, the optimal weight p_k^* on outcome x_k must be increasing in the DM's valuation of x_k. In other words, the optimal $\mathbf{p}^*$ must be a negatively skewed distribution. As an example, for the entropy case $H(\mathbf{p}) = -\sum_{k=1}^n p_k \ln(p_k)$ (taking $\psi = I$), the optimal $\mathbf{p}^* \in \text{int}\Delta^n(X)$ assigns weight p_k^* to x_k according to the logit formula

$$p_k^* = \frac{\exp(v_k/\alpha)}{\sum_{j=1}^n \exp(v_j/\alpha)}$$

As an alternative measure, I also study in Appendix F what Ely, Frankel, and Kamenica (2015) call the "residual variance," $H(\mathbf{p}) = \sum_{k=1}^{n} p_k(1-p_k)$.[9] Unlike the entropy, the optimal $\mathbf{p}^*$ in this case need not belong to the interior of $\Delta^n(X)$. Both measures are part of the larger class of permutation-invariant measures, for which $\lim_{\alpha\to\infty} \mathbf{p}^* = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$ provided H is strictly concave.[10]

Observation 4. If Ψ depends on the spread in valuations, then $\succeq$ may violate X-MON at some $(d, d') \in \mathcal{D}^+ \times \mathcal{D}^+$.

For instance, if $\Psi(\mathbf{p}, \mathbf{v}) = \sum_{k=1}^{n} p_k(v_k - E_{\mathbf{p}}(v))^2$, the DM will exhibit the preference $(x_1, p; x_3, 1-p) \succ x_1 \succ (x_1, p; x_2, 1-p)$ provided that $\alpha \in \left(\frac{1}{p(v_1-v_3)}, \frac{1}{p(v_1-v_2)}\right)$. As $\alpha \to \infty$, such a DM will prefer to put equal weight on the two most extreme outcomes.

To complete the typology, consider measures outside of the class of valid measures of uncertainty. First, measures that capture the variability in valuations without accounting for the probability of each prize may violate X-MON but not P-MON. For instance, a DM who only cares about the average distance between valuations will prefer lotteries that put all the weight on the two most extreme outcomes, with $(1-\epsilon)$-weight on the top outcome. Second, a DM for whom the only relevant uncertainty is in the size of the support of the lottery i.e., $\Psi(\mathbf{p}, \mathbf{v}) = \psi(|\text{supp}(\mathbf{p})| - 1)$, will satisfy both P-MON and X-MON. Note that a measure of this type is only weakly concave as $\Psi(\lambda\mathbf{p} + (1-\lambda)\mathbf{q}) = \lambda\Psi(\mathbf{p}) + (1-\lambda)\Psi(\mathbf{q})$ for all $\mathbf{p}, \mathbf{q} \in \text{int}\Delta^n(X)$. Such a DM will prefer a lottery that assigns $(1-\epsilon)$-weight to their preferred outcome and distributes the remaining ϵ-weight on the other $n-1$ outcomes. Table 1 summarizes the relationship between the monotonicity properties and the shape of Ψ with examples in each cell.

[9] To interpret this measure, consider a DM who is offered lottery $\mathbf{p}$ and contemplates each possible outcome x_k separately, wondering whether it will materialize or not next period. Letting $X_k = 1$ if x_k is realized and $X_k = 0$ otherwise, the DM expects $p_k = \mathsf{P}\{X_k = 1\}$ under $\mathbf{p}$, with variance $p_k(1-p_k)$. Thus, the residual variance can be interpreted as the sum of the variances of Bernoulli random variables $X_k \sim \mathcal{B}(p_k)$, one for each outcome x_k that could be realized.

[10] H is permutation invariant if $H(\mathbf{p}) = H(\mathbf{p}_\sigma)$ for all $\mathbf{p}_\sigma = (p_{\sigma(1)}, ..., p_{\sigma(n)}) \in \Delta^n(X)$, where $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ is a permutation function. To see why $\lim_{\alpha\to\infty} \mathbf{p}^* = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$, note that if H is strictly concave, it has a unique maximizer, $\mathbf{p}^*$; if H is also permutation invariant, $H(\mathbf{p}^*) = H(\mathbf{p}^*_\sigma)$ implies $\mathbf{p}^* = \mathbf{p}^*_\sigma$ for all permutations σ. The only distribution that satisfies this condition is $\mathbf{p}^* = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.

Table 1: Typology of $\Psi(\mathbf{p}, \mathbf{v})$ measures

	✓ X-MON	✗ X-MON
✓ P-MON	Measures $\Psi(\mathbf{p},\mathbf{v}) = \psi(h(\mathbf{p}))$ s.t. - $\psi(h(\mathbf{p})) = \psi(h(\mathbf{q}))\ \forall\, \mathbf{p}, \mathbf{q} \in \text{int}\Delta^n(X)$ - h weakly concave - $h(\delta_x) = 0\ \ \forall x \in X$ <u>Example</u>: Cardinality of support $\psi(\lvert\text{supp}(\mathbf{p})\rvert - 1)$	Measures of distance between valuations $\Psi(\mathbf{p},\mathbf{v}) = \psi\Big(\frac{1}{\varphi(\lvert\mathcal{J}\rvert)} \sum\limits_{j\in\mathcal{J}} \sum\limits_{k>j} d(v_j, v_k)\Big)$ where $\mathcal{J} := \{j \in \mathbb{N} \mid x_j \in \text{supp}(\mathbf{p})\}$ and $\varphi(\lvert\mathcal{J}\rvert) \geq 1$ is a weighting function <u>Example</u>: Average squared distance $\psi\Big(\frac{1}{\lvert\mathcal{J}\rvert(\lvert\mathcal{J}\rvert-1)/2} \sum\limits_{j\in\mathcal{J}} \sum\limits_{k>j} (v_j - v_k)^2\Big)$
✗ P-MON†	Measures $\Psi(\mathbf{p},\mathbf{v}) = \psi(H(\mathbf{p}))$ s.t. - H globally concave - $H(\delta_x) = 0\ \ \forall x \in X$ <u>Example 1</u>: Shannon entropy $\psi(-\sum_{k=1}^{n} p_k \ln(p_k))$ <u>Example 2</u>: Residual variance $\psi(\sum_{k=1}^{n} p_k(1-p_k))$	Measures $\Psi(\mathbf{p},\mathbf{v}) = \psi(H(\mathbf{p},\mathbf{v}))$ s.t. - H globally concave in $\mathbf{p}$ - $H(\delta_x, \mathbf{v}) = 0\ \ \forall x \in X$ - H increasing in distance btw. valuations <u>Example</u>: Variance in valuations $\psi(\sum_{k=1}^{n} p_k(v_k - E_{\mathbf{p}}(v))^2)$

Notes: †Satisfied on $\mathcal{B}^{+}_{\sim}$ if $\psi' > 0, \psi'' \leq 0$

3 Experimental Design

3.1 Recruitment and incentives

Recruitment period An initial pilot took place in December 2019 with 9 participants recruited via Facebook ads. The main data was collected using a more focused version of the survey administered to 83 participants in March 2020. All responses were completed over March 5$^{\text{th}}$ - 19$^{\text{th}}$, at a time of growing concerns over the COVID-19 pandemic, with travel restrictions and lockdowns spreading across the world.[11] As

[11] Italy was the first European (and worldwide) country to impose a national lockdown on 9 March 2020. The pilot survey included an Italian destination, which was replaced by a destination in Portugal for the main study. However, by March 19$^{\text{th}}$, several destinations advertised in the survey were in countries that had imposed travel restrictions (https://ourworldindata.org/coronavirus). By contrast, the UK had not imposed any restrictions yet, which might have made concerns less salient; as a matter of fact, only one respondent raised concerns about the pandemic in the survey.

the crisis was still in its infancy, it is however plausible that most people expected to be able to travel during the year; in fact, about 75% of respondents anticipated a travel date before end of 2020.[12] In an exploratory analysis, I assess how the survey completion date interacts with decisions, finding a limited impact (Appendix A.2).

Recruitment procedure About 87% of respondents were recruited via the Behavioural Research Lab of the London School of Economics, with the rest recruited on social media or some other channel. The survey was advertised as an online study on preferences for travel. The eligibility requirements were: being at least 18 years old, living in London, and holding valid documents to travel to the European Union. Once started, the survey had to be completed within 2 hours in order to limit the scope for surprise enjoyment within the survey. Respondents had to correctly answer comprehension questions throughout the survey. The average completion time was 44 minutes. Summary statistics about the sample are presented in Table A1.

Incentives All participants who completed the study received a £25 voucher valid for a trip with the travel partner, who was kept anonymous until the end of the survey. In addition, participants were told that they could win a free (non-transferable) trip to a European destination for 4 days and 3 nights worth £420. They knew that the survey was capped at 100 responses and that we had 5 such trips to give away, implying a 1 in 20 chance of winning. The other participants received either a discount voucher for a specific trip or a monetary payment. The prize received was based on one randomly selected decision in one of the sections of the study.[13] Ignoring the trip vouchers, the above incentives translate into expected earnings of around £34.80.

[12]In line with this, a survey on public expectations about the lifting of COVID-19 restrictions in Ireland ($N = 800$) found that over 60% (80%) of respondents surveyed in April 2020 expected restrictions on non-essential international travel to be lifted by the end of 2020 (by June 2021) (https://www.esri.ie/system/files/publications/SUSTAT88.pdf).

[13]The use of the Random Incentive System (RIS) is standard in experimental economics. This procedure for eliciting preferences is incentive compatible under expected utility (Azrieli, Chambers, and Healy, 2018). However, respondents who seek to maximize uncertainty could in principle try to randomize across decision problems, thus treating the experiment as one big lottery. To examine whether this is a real concern, I asked respondents how they made their choices after a sequence of decisions problems. With no exception, respondents reported treating each problem in isolation.

3.2 Structure of the study

Respondents first indicated a date at which they considered traveling. Afterwards, they were taken through 5 parts detailed below and summarized in Table 2.

Table 2: Structure of the survey

PART 1	**Elicitation of preferences $\succeq$ over holiday trips** ▷ Ordinal ranking of 10 destinations, $x_1, x_2, ..., x_{10}$ ▷ Valuation $v_k \in [0, 500]$ for each destination x_k
PART 2	**Choices in 45 binary decision problems (DPs)** A: $(x_i, 1)$ vs. B: $(x_j, p_j; x_k, p_k; ...; x_l, p_l)$
PART 3	**Design of favorite lottery** ▷ Selection of support and probability distribution ▷ Selection of date at which to resolve the uncertainty ▷ [If chose a positive delay] Valuation of delay
PART 4	**Preferences for a "wildcard trip"** ▷ Valuation of a trip to an unknown destination ▷ Questions to understand motives
PART 5	**Preferences over monetary gambles** 3 DPs with each destination replaced by its valuation Valuation of one risky and one ambiguous monetary bet
	Questions about travel history and preferences
END	Selection of choice problem

3.2.1 Elicitation of a preference ordering over 10 holiday trips

In Part 1, I elicited respondents' preferences over a set of trips to one of 10 European destinations. Respondents saw a short description of each trip with a picture and some information about the location, activities available on site, and accommodation offered (see example in Figure 1). The 10 destinations were: Düsseldorf, Gothenburg, Maastricht, Midi-Pyrénées, Porto, Prague, Santiago de Compostela, Sofia, Turku, Zakopane. The trip packages were designed by the travel partner with 4 concerns in mind: (i) no two destinations should be in the same country; (ii) they should offer a diversity of experiences; (iii) they should be easy to travel to and from; (iv)

Figure 1: Sample trip description

Destination: **Turku** (Finland)

Description: Step back through the ages in this medieval town, and explore the maze-like chambers of the impressive Turku Castle. Enjoy nature in national parks, sunbathe on beaches, and explore museums, art galleries, historic sites, restaurants aplenty, and shopping hotspots, all around town.

Accommodation: Arrive at Helsinki airport and take a train to Turku. For the best experience of living local, stay in a cosy Airbnb that you'll have entirely to yourself, located conveniently in the central of town.

packages should all have the same value for money (£420).[14] After reading each trip description, respondents rated their degree of familiarity with the destination. Their preferences $\succeq$ over the 10 trips were then elicited in two steps:

Step 1: Participants ranked the 10 destinations (listed in a random order), from top to bottom (options ranked x_1 to x_{10}).

Step 2: Right next to each destination x_k in the list, participants entered their minimum price v_k between 0 and 500 GBP for giving up the trip if they won it.

To incentivize the ranking in Step 1, respondents were informed that their chances of being offered a given trip (or equivalent amount of money) were higher if the trip was listed higher in their ranking. Under fairly mild assumptions, this procedure gives an incentive to truthfully report a strict preference.[15] However, the data from Step 1 is not rich enough to know whether preferences are strict or weak. To identify potential

[14] Participants were informed that "*the travel partner operates within a fixed budget such that if a destination is cheap to travel to, more money will be spent on the accommodation and vice versa.*"

[15] The exact probability distribution was $\mathbf{p} = (0.25, 0.20, 0.15, 0.12, 0.10, 0.08, 0.05, 0.03, 0.02, 0)$, where $p_1 = 0.25$ is the probability of drawing the destination ranked #1 and so on; this information was available by clicking on a link. Incentive compatibility requires the DM to satisfy FOSD over the subset of lotteries $\{\mathbf{p}_\sigma \in \Delta^n(X) \mid \sigma : \mathbb{N} \to \mathbb{N} \text{ is a permutation function}\}$. In principle, this is not guaranteed e.g., a DM who prefers lotteries with a larger variance in valuations might put more weight on more extreme (less valued) options. However, this property is satisfied for most measures considered (entropy, residual variance, support). In addition, since the realized draw was revealed right at the end of the survey, the anticipatory utility benefits of randomization were limited.

indifferences, I use the cardinal information collected in Step 2. Valuations were elicited using the standard Becker-DeGroot-Marschak (1964) procedure (henceforth, BDM): if a respondent won a trip to the destination ranked #k, x_k, their valuation v_k was compared to a random number $V \sim \mathcal{U}_{[0,500]}$; if $V \geq v_k$, the respondent received £V instead of the trip (and otherwise kept the trip).[16] Monotonicity of valuations with respect to the ranking was not strictly enforced, but respondents were told that "*logically they should have entered a higher price for options they ranked higher*" and asked to check their answers. I combine the data from Steps 1 & 2 to construct a respondent's preference ordering $\succeq$ over destinations (see Section 4.2).

3.2.2 Binary choice data

In Part 2, respondents faced a series of 45 decision problems (henceforth, DP), which were tailored to each respondent based on their ranking of destinations x_1, x_2, ... x_{10} in Part 1 (see Section 3.2.1, Step 1). Most problems involved a choice between the following two options:

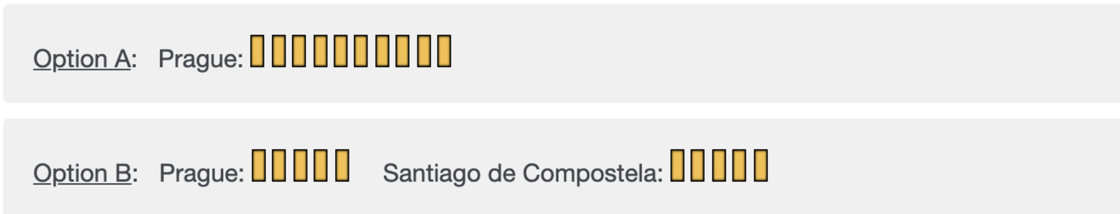
Option A: a trip to a given destination for sure (either x_1, x_2, x_3, x_5 or x_{10})

Option B: a "surprise lottery" between two or more destinations.

With a surprise lottery, participants could only discover their travel destination the week of departure: all the information would be contained in a sealed envelope received about 7 days prior to departure, with a suggestion to only open it at the airport. No information would be sent in between, except for a confirmation email after booking (with travel dates/times and airport), to focus on one-shot resolution of uncertainty. To keep the two options as comparable as possible, the email and envelope would be sent following the same schedule. In addition, choices were framed in exactly the same way, with 10 lottery tickets either entirely allocated to the sure destination (A) or split between two or more destinations (B). An illustration of one decision problem is provided in Figure 2. The surprise lotteries varied in 3 ways:

[16]While the BDM mechanism is incentive compatible in a standard EU framework, this is no longer guaranteed if the DM values uncertainty. With $V \sim \mathcal{U}_{[0,500]}$, reporting $\tilde{v} = 250$ maximizes the uncertainty over the realized prize (trip or money); depending on the respondent's true valuation v, the bias $\tilde{v} - v$ may be either positive or negative. Given the early resolution of uncertainty, the bias should be at most small and lead (if anything) to underestimating the size of α, due to the compression in valuations induced. See Appendix B.1 for a full analysis.

Figure 2: Sample decision problem



Notes: The destinations presented in a given DP depended on the respondent's ranking. Choosing Option B entailed learning the outcome of the lottery no earlier than one week before travel.

1. *Number of destinations*: The lottery contained 2, 3, 5 or all 10 destinations.

2. *Ranking of the destinations*: Fixing the support size, I varied the rank of the destinations in a respondent's preference ordering (e.g., $\{x_1, x_2\}$ vs. $\{x_1, x_{10}\}$).

3. *Chances of each destination*: The probability distribution $\mathbf{p}$ over destinations varied across DPs. For two-outcome lotteries, $p \in \{0.1, 0.5, 0.9\}$. For lotteries with 3, 5 or 10 outcomes, the weights were (close to) uniform.[17]

In 6 of the 45 DPs, Option B was a degenerate lottery i.e., respondents simply chose between two destinations (with Option A always ranked higher); I use this information to assess measurement error in the original ranking. In 24 DPs, Option A was ranked (weakly) higher than all destinations in the Option B lottery; choosing B thus implied a violation of stochastic dominance (up to indifferences and mistakes) in the form of a preference for randomization, $\alpha > 0$ (set $\mathcal{D}^+$). In 11 DPs, the opposite was true, with Option A ranked (weakly) lower than all destinations in Option B, thus allowing the respondent to express a preference for certainty, $\alpha < 0$ (set $\mathcal{D}^-$). In addition, there were 4 DPs where Option B contained both higher- and lower-ranked destinations relative to Option A. Table 3 presents the full list of DPs.

The 45 DPs appeared in a fixed order with simple lotteries (two destinations, equal probability) presented first and progressively moving to more complex choices. All problems appeared on one page. Before making their choices, respondents were reminded of their ranking from Part 1; in addition, they could consult again the trip descriptions. Respondents were asked to review their 45 selections before moving on.

[17]For lotteries with 3 options, x_i, x_j and x_k such that $i \leq j < k$, the weights were $(p_i, p_j, p_k) = (0.4, 0.3, 0.3)$.

Table 3: List of 45 decision problems

DP #	Degenerate lotteries (6 choices)
1, 4, 10, 13	A: $(x_1, 1)$ vs. B: $(x_k, 1)$ for $k \in \{2, 3, 5, 10\}$
7	A: $(x_2, 1)$ vs. B: $(x_3, 1)$
16	A: $(x_5, 1)$ vs. B: $(x_{10}, 1)$
	Dominated lotteries (24 choices)
2, 18, 27	A: $(x_1, 1)$ vs. B: $(x_1, p; x_2, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
5, 20, 29	A: $(x_1, 1)$ vs. B: $(x_1, p; x_3, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
8, 22, 31	A: $(x_1, 1)$ vs. B: $(x_2, p; x_3, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
11, 24, 33	A: $(x_1, 1)$ vs. B: $(x_1, p; x_5, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
14, 25, 34	A: $(x_1, 1)$ vs. B: $(x_1, p; x_{10}, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
36	A: $(x_1, 1)$ vs. B: $(x_1, 0.4; x_2, 0.3; x_3, 0.3)$
37	A: $(x_1, 1)$ vs. B: $(x_1, 0.4; x_3, 0.3; x_5, 0.3)$
38	A: $(x_1, 1)$ vs. B: $(x_2, 0.4; x_3, 0.3; x_4, 0.3)$
39	A: $(x_5, 1)$ vs. B: $(x_6, 0.4; x_7, 0.3; x_8, 0.3)$
40	A: $(x_1, 1)$ vs. B: $(x_1, 0.2; x_2, 0.2; x_3, 0.2; x_4, 0.2; x_5, 0.2)$
41	A: $(x_1, 1)$ vs. B: $(x_1, 0.2; x_3, 0.2; x_5, 0.2; x_7, 0.2; x_{10}, 0.2)$
42	A: $(x_1, 1)$ vs. B: $(x_6, 0.2; x_7, 0.2; x_8, 0.2; x_9, 0.2; x_{10}, 0.2)$
43	A: $(x_5, 1)$ vs. B: $(x_6, 0.2; x_7, 0.2; x_8, 0.2; x_9, 0.2; x_{10}, 0.2)$
44	A: $(x_1, 1)$ vs. B: $(x_1, 0.1; x_2, 0.1; ...; x_9, 0.1; x_{10}, 0.1)$
	Dominant lotteries (11 choices)
3, 19, 28	A: $(x_2, 1)$ vs. B: $(x_1, p; x_2, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
6, 21, 30	A: $(x_3, 1)$ vs. B: $(x_1, p; x_3, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
9, 23, 32	A: $(x_3, 1)$ vs. B: $(x_1, p; x_2, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
12	A: $(x_5, 1)$ vs. B: $(x_1, p; x_5, 1-p)$ for $p = 0.5$
15	A: $(x_{10}, 1)$ vs. B: $(x_1, p; x_{10}, 1-p)$ for $p = 0.5$
	Other lotteries with no dominance (4 choices)
17, 26, 35	A: $(x_5, 1)$ vs. B: $(x_1, p; x_{10}, 1-p)$ for $p \in \{0.5, 0.9, 0.1\}$
45	A: $(x_5, 1)$ vs. B: $(x_1, 0.1; x_2, 0.1; ...; x_9, 0.1; x_{10}, 0.1)$

Notes: DP # refers to the decision problem number as shown on the respondent's screen; x_k is the destination ranked # k by the respondent (see Section 3.2.1, Step 1).

3.2.3 Design task and WTP for delay

While binary choices allow to directly reveal stochastic dominance violations, they do not allow to pin down respondents' optimal lottery, nor identify whether they value preserving the surprise. In Part 3, respondents were offered to design their favorite lottery and select their preferred date at which to learn the destination. As a first step, respondents decided whether they preferred a sure destination (no surprise) or a surprise lottery of their choice. Those who chose to build a lottery (i) picked the support from the set of 10 destinations; (ii) allocated 100 lottery tickets to the selected options. As a second step, respondents selected the date at which to discover their destination from a list of options (e.g., "*Today, after I completed the survey*," "*About 2 weeks from now*," "*About 4 weeks before going to the airport*," etc.), or by indicating a specific day of their choice.[18] Those who initially preferred a sure destination were asked whether they would now prefer a lottery if the destination discovery date could be chosen. Finally, those who chose to delay the resolution of uncertainty (i.e., did not pick "*Today, after I completed the survey*") indicated how much they valued this option by making choices in a Multiple Price List of the form:

$$\underset{\text{[Chosen date]}}{\textsf{Revealing Later}} \textsf{ vs. } \underset{\text{[End of survey]}}{\textsf{Revealing Now}} + \pounds\text{X}$$

where $X \in \{5, 10, 15, 20, 30, 40, 50\}$. If implemented, one of the rows was randomly selected for payment, thus ensuring incentive compatibility.

3.2.4 Additional data

In the last two parts, I collect data to better understand randomization decisions. In Part 4, I elicit respondents' valuation $v^* \in [0, 500]$ for a "wildcard trip" to a new and unknown destination, which was closer in spirit to what the company sold.[19] I collect information on the determinants of these valuations by measuring respondents' motives and perceived chances of certain outcomes. In Part 5, I study whether stochastic dominance violations remain when prizes are monetary and measure respondents' risk and ambiguity attitudes in the domain of money. Finally, I collect information about respondents' travel history and travel preferences as well as their overall assessment of the concept of "surprise trip," including their reasons for (dis)liking the concept. The decision selected to count was revealed right at the end.

[18] Respondents were told that they would be informed of their destination by email at the specified date, but still receive the sealed envelope and travel guide one week before going to the airport.

[19] Participants were told that the wildcard trip was to another European destination (possibly in one of the 10 countries already presented or a different country), and came with the same features as the other 10 trips (same duration, market value, date the surprise is revealed, etc.). Valuations were again elicited via a BDM mechanism. See Appendix D.1 for more information.

4 Preliminaries: preferences over destinations

I start this results section by examining aggregate and individual preferences over the set X of 10 destinations, and explain how I construct the preference relation on X.

4.1 Aggregate preferences

Using the average rank or average valuation across respondents as a measure of aggregate preferences, no trip clearly dominates or is clearly dominated (Figure B1). Gothenburg and Düsseldorf show up as the most and least popular destinations for both the ranking and valuations, but the null hypothesis of randomly assigned ranks can only be rejected for these two destinations.[20] The average valuation is similar across destinations (between £226.8 and £273.6), and standard deviations are large (between £112.4 and £125.4). These findings reflect the fact that the 10 trips offered a diversity of experiences and had the same market value (£420). Interestingly, the average valuation of the wildcard trip at £275.4 exceeds the other 10 trips, a notable result given that no information on the wildcard trip was provided except for its market value. This finding suggests that a travel company unable to cater to travelers' heterogeneous preferences (e.g., due to a lack of data or capacity constraints) could maximize total revenue by selling trips to an unknown destination.

While no destination is a clear winner or loser, there is a significant difference in the average valuation of higher- vs. lower-ranked options (Table B1). On average, the difference is £38 when comparing the destinations ranked #1 and #2, and goes up to £218 when comparing the #1 and #10. Importantly, the vast majority of respondents greatly valued the possibility to go on at least one of these holiday trips: the trip of highest value was worth at least £200 for all respondents and at least £320 for 75% of them. These amounts are non trivial given that many respondents were students. On the other hand, the lowest-rank destination seemed quite unattractive for many respondents, with 25% valuing it at £80 or less and only 25% valuing it at £220 or more; this suggests that not all destinations might be a good surprise.

[20] If respondents assigned a rank at random to each destination, the average rank for each destination would be 5.5. A t-test rejects the null of equality of means for Gothenburg and Düsseldorf, with average ranks of 4.8 and 6.4 (p = 0.017 and p = 0.003) and fails to reject the null for all others.

4.2 Construction of individual preference orderings

Combining the ranking and valuations Having presented aggregate data on the ranking of trips and valuations, I now combine this ordinal and cardinal information to construct a first preference relation $\succeq$ on X for each respondent (further refined in the next paragraph). Since monotonicity of valuations with respect to the ranking was not strictly enforced, a respondent could have in principle entered a higher valuation for a destination ranked lower. Indifferences could also be expressed by entering the same valuation for two destinations. For any two trips x_j and x_k with x_j ranked above x_k ($j < k$), I thus construct $\succeq$ as: (i) $x_j \succ x_k$ if $v_j > v_k$; (ii) $x_j \sim x_k$ if $v_j = v_k$; (iii) $x_j \bowtie x_k$ if $v_j < v_k$.

Adding the binary choice data In addition to the ranking and valuations, respondents made direct choices between two destinations in 6 DPs: $\{x_1, x_2\}$, $\{x_1, x_3\}$, $\{x_2, x_3\}$, $\{x_1, x_5\}$, $\{x_1, x_{10}\}$, $\{x_5, x_{10}\}$. Although not all binary comparisons could be examined, data on these 6 comparisons allows to account for possible classification errors on $\mathcal{B}^+ \cup \mathcal{B}^-$ i.e., the vast majority of DPs. I combine this data with information on $\succeq$ to construct a second relation $\succeq^*$ that accounts for possible inconsistencies between the ranking and direct choices and is thus more incomplete. More precisely: (i) $x_j \succ^* x_k$ if $[x_j \succ x_k] \cap [x_j$ chosen from $\{x_j, x_k\}]$; (ii) $x_j \sim^* x_k$ if $x_j \sim x_k$; (iii) $x_j \bowtie^* x_k$ if $x_j \bowtie x_k$ or $[x_j \succ x_k] \cap [x_k$ chosen from $\{x_j, x_k\}]$.

Result 1. *The number of inconsistencies between the various preference measurements is low, as is the number of indifferences expressed.*

Table 4 presents the breakdown of preferences $(\succeq, \succeq^*)$ on X. Over 80% of comparisons between two destinations (x_j, x_{j+1}) adjacent in the respondent's ranking involve a strict preference ($v_j > v_{j+1}$). Indifferences represent 12% of comparisons, while valuations conflict with the ranking in only 4% of comparisons. Overall, 52% of respondents have a strict ordering $\succ$ on X. The ranking and valuations never contradict each other for 88% of respondents. Binary choices are also highly consistent with $\succeq$, as inconsistencies occur only about 8% of the time. In total, nearly 80% of respondents provided preference information that is fully consistent across all types of measurements ($\bowtie^* = \varnothing$); see Appendix B.3 for more information.

Table 4: Breakdown of preferences ($\succeq, \succeq^*$) on X

Preference	$x_j \succ x_k$	$x_j \sim x_k$	$x_j \bowtie x_k$	$x_j \succeq^* x_k$
% comparisons with $k = j + 1$	84.1	12.1	3.9	92.2
(frequency)	(628/747)	(90/747)	(29/747)	(153/166)
% comparisons overall	91.4	4.3	4.3	91.8
(frequency)	(3415/3735)	(161/3735)	(159/3735)	(457/498)

Notes: $x_j \succ x_k$ (resp. $x_j \sim x_k$) means that x_j ranked above x_k and $v_j > v_k$ (resp. $v_j = v_k$); $x_j \bowtie x_k$ means that x_j ranked above x_k and $v_j < v_k$; finally, $x_j \succeq^* x_k$ means either (i) $x_j \succ x_k$ and x_j chosen from $\{x_j, x_k\}$ or (ii) $x_j \sim x_k$. Total number of binary comparisons is $|X|(|X|-1)/2 \times N = 45 \times 83 = 3735$; total number of comparisons between two adjacent options ($k = j + 1$) is $9 \times 83 = 747$. For the last column, there are 6 DPs $\{x_j, x_k\}$ (including 2 DPs $\{x_j, x_{j+1}\}$); therefore, the denominator is $2 \times 83 = 166$ at the top and $6 \times 83 = 498$ at the bottom.

5 Violations of stochastic dominance

Having constructed preferences over destinations, I now examine stochastic dominance (SD) violations in the binary decision problems. First, I study the overall *prevalence* of these violations at the individual level. Second, I assess their *size* by computing the implied monetary cost. Third, I examine their *shape* by assessing their compatibility with the monotonicity properties presented in Section 2.1.

Computation of SD violations Before proceeding, I explain how I calculate for each respondent the total number and fraction of SD violations on $\mathcal{D}^+$ (all dominated lotteries), $\mathcal{B}^+$ (dominated lotteries with two outcomes) and $\mathcal{B}^-$ (corresponding dominant lotteries).[21] To avoid overestimating the prevalence of SD violations (i.e., making Type I errors), I apply the strictest possible classification given the available data. First, I identify a subset of clear dominance problems for each respondent by relying on the most coarse relation $\succeq^*$. More precisely, for any DP d between a sure option x_i and a lottery $\mathbf{p}$: (i) $d \in \mathcal{D}^+$ (or $\mathcal{B}^+$) if $x_i \succeq^* x$ for all $x \in \text{supp}(\mathbf{p})$ and $x_i \succ^* x$ for some $x \in \text{supp}(\mathbf{p})$; (ii) $d \in \mathcal{B}^-$ if $x \succeq^* x_i$ for all $x \in \text{supp}(\mathbf{p})$ and $x \succ^* x_i$ for some $x \in \text{supp}(\mathbf{p})$. I then calculate the number and fraction of violations only on this subset of clear dominance problems.[22] As this procedure generates more Type II

[21] When presenting aggregate statistics that contrast behavior on $\mathcal{B}^+$ vs. $\mathcal{B}^-$, I restrict attention to the subset of 11 problems in $\mathcal{B}^+$ with a symmetric counterpart in $\mathcal{B}^-$ so as to maximize comparability (thus removing problems $\{x_1, (x_1, p; x_5, 1-p)\}$ and $\{x_1, (x_1, p; x_{10}, 1-p)\}$ for $p \in \{0.1, 0.9\}$).

[22] As inconsistencies across preference measurements are quite rare, the number N^* of clear dominance problems is simply equal to the total number of DPs for 80% of respondents. However, $\text{N}^* = 0$

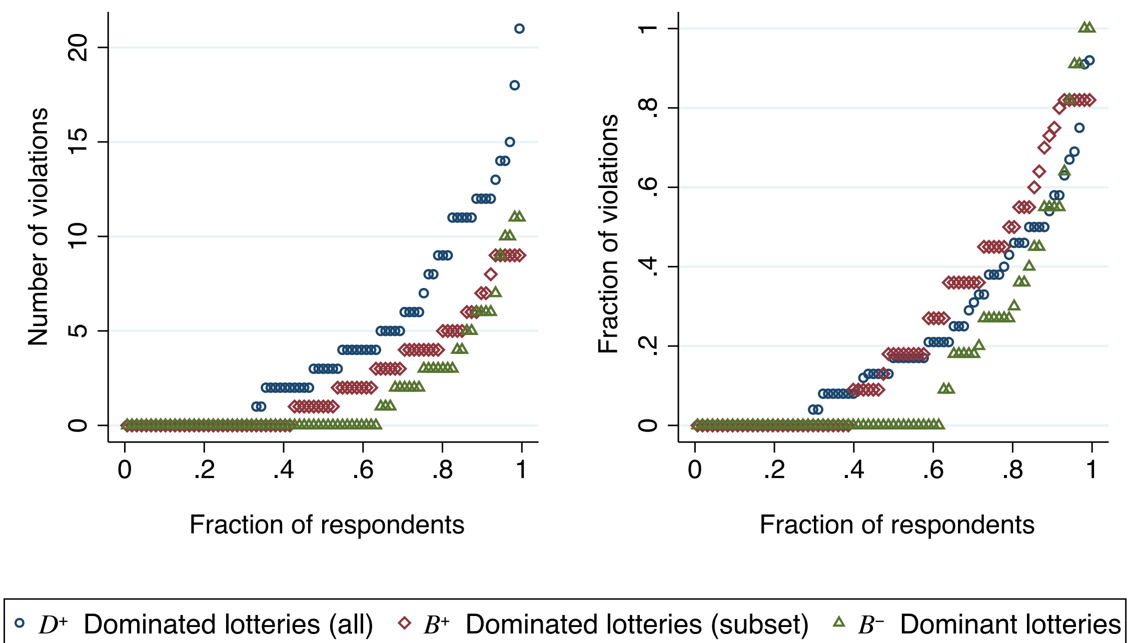
errors, Appendix B.3 presents statistics for a classification based only on the ranking, thus providing upper and lower bounds.

5.1 Prevalence of SD violations

Result 2. *On average, respondents violate stochastic dominance in favor of randomization in about 25% of decision problems on $\mathcal{B}^+$ (22% on $\mathcal{D}^+$) vs. 16% of problems on $\mathcal{B}^-$ i.e., in favor of certainty. However, individual heterogeneity is large.*

Figure 3 presents quantile plots of respondents' number (left panel) and fraction (right panel) of SD violations for decision problems in $\mathcal{D}^+$, $\mathcal{B}^+$ and $\mathcal{B}^-$. Each dot represents a respondent. On $\mathcal{D}^+$ ($\mathcal{B}^+$), 29% (39%) of respondents never violate stochastic dominance, while the top 20% do so at least 46% (50%) of the time (≥ 9 (5) violations), thus revealing a strong preference for randomization. On $\mathcal{B}^-$, respondents are

Figure 3: Prevalence of SD violations at the individual level



Notes: Quantile plots of the number (left panel) and fraction (right panel) of SD violations for each respondent on $\mathcal{D}^+$ (all 24 dominated lotteries), $\mathcal{B}^+$ (subset of 11 dominated lotteries), and $\mathcal{B}^-$ (11 dominant lotteries), defined according to $\succeq^*$. N = 83 for left panel and N = 79 for right panel.

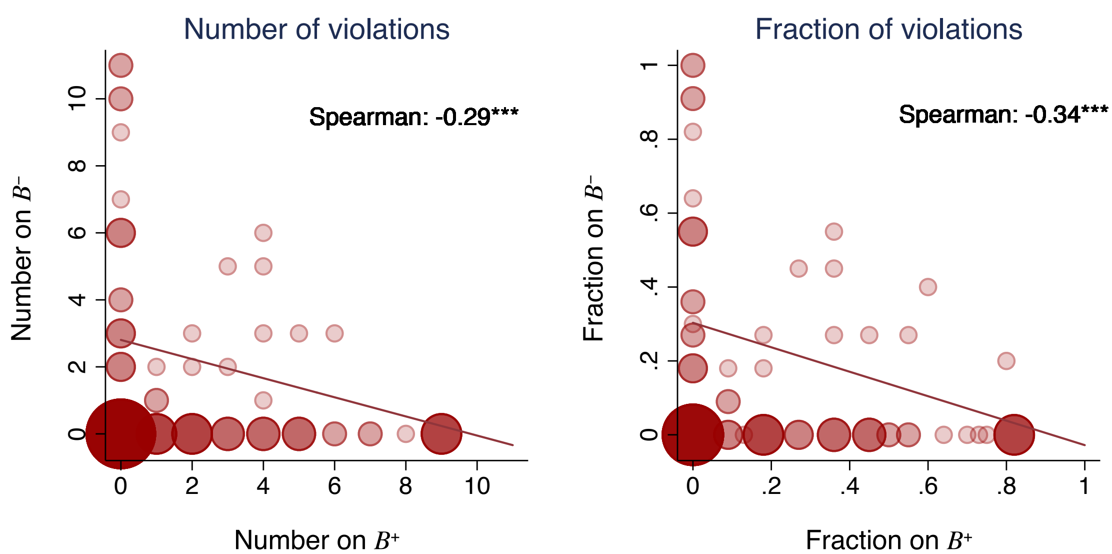
for 4 respondents with fully contradictory measurements; thus N= 79 for the fraction of violations.

generally less inclined to violate stochastic dominance, with 62% always selecting the dominant lottery; however, the top 20% still exhibit violations in favor of certainty at least 30% of the time ($\geq$ 3 violations). Overall, violations occur more frequently in favor of randomization than certainty, with a near first-order stochastic dominance relationship between the distributions on $\mathcal{B}^+$ vs. $\mathcal{B}^-$.[23]

Result 3. *Respondents who violate stochastic dominance in favor of randomization are less likely to exhibit violations in favor of certainty and vice versa, suggesting the two types of violations reflect largely distinct phenomena.*

Figure 4 contrasts respondents' propensity to violate stochastic dominance in each direction. Over 80% of respondents are on one of the axes, meaning that they exhibit at most one type of SD violations. The rank-order correlation between violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ is negative at about -0.3 ($p < 0.01$).

Figure 4: Relationship between SD violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$



Notes: Scatter plot of the number (left panel) and fraction (right panel) of violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$, with the bubble size proportional to the number of respondents. The red line is from a linear regression between the two variables (left panel: $\hat{\beta} = -0.29$, s.e.$= 0.11$; right panel: $\hat{\beta} = -0.30$, s.e.$= 0.10$). N$= 83$ for left panel and N$= 79$ for right panel. *** $p < 0.01$.

[23] A Kolmogorov-Smirnov test rejects that the two distributions are the same both for the number ($p = 0.04$) and fraction ($p = 0.03$) of violations. A t-test of equality of means yields $p = 0.11$ for the number of violations (2.3 vs. 1.6) and $p = 0.04$ for the fraction of violations (0.25 vs. 0.16).

Table 3: Typology of respondents

Number of SD violations		% (N) of respondents
$=0$	for all dominance problems	15.7 (13)
>0	for dominated lotteries only ($\mathcal{D}^+$)	48.2 (40)
>0	for dominant lotteries only ($\mathcal{B}^-$)	16.9 (14)
>0	for both dominated and dominant lotteries	19.3 (16)

Notes: On the subset of two-outcome lotteries $\mathcal{B}^+ \cup \mathcal{B}^-$, 42% (20%) of respondents only violate stochastic dominance for dominated (dominant) lotteries and 22% fully respect stochastic dominance.

As shown in Table 3, 48% (17%) of respondents only violate stochastic dominance when lotteries are dominated (dominant) and 16% fully respect stochastic dominance. Importantly, if noise was a predominant factor, the correlation between the two types of violations should be either positive or zero.[24]

Link to randomization decisions in the design task Since choosing a lottery over a sure destination entailed no extra monetary or effort cost in the binary choice exercise, respondents who were indifferent could have chosen either option.[25] To assess the robustness of randomization decisions, respondents were also directly offered to design their own lottery or select a sure destination instead. Designing a lottery took several steps and was therefore more time-consuming than selecting a sure destination. In total, 55% chose to design a lottery as their favorite option.[26] This number is in the same ballpark as the proportion of respondents who revealed a preference for randomization in the binary choice exercise (see Table 3). Furthermore, the two are related: those who chose to design a lottery were significantly more likely to exhibit SD violations on $\mathcal{B}^+$ and significantly less likely to do so on $\mathcal{B}^-$ (see Section C.3).

[24]To investigate this, I simulated 4 stochastic choice benchmarks described in Appendix E. The first two assume a fixed probability of mistake at each DP, which is either homogeneous (Benchmark 1) or heterogeneous (Benchmark 2) across respondents (*fixed error models*). Benchmark 3 assumes that respondents have a noisy perception of the valuation of each trip, with the noise varying at each DP (*random valuations model*). Benchmark 4 assumes expected utility with additive shocks at each DP (*random utility model*). In all cases, I pick the noise to roughly match the proportion of preference reversals observed in DPs involving two sure destinations ($\approx 10\%$), and assess the role of sampling error by simulating 1,000 datasets. The median correlation between the two types of violations is positive and large for Benchmarks 2, 3 and 4 and close to zero for Benchmark 1.

[25]For instance, choosing $(x_1, 0.5; x_2, 0.5)$ over x_1 only implies $(x_1, 0.5; x_2, 0.5) \succeq x_1$, strictly-speaking; nevertheless, such a choice implies a violation of stochastic dominance if $x_1 \succ x_2$.

[26]15% chose to build a lottery only after being offered to customize the date at which to learn the destination, while 40% chose to build a lottery even with a fixed delay of one week before travel.

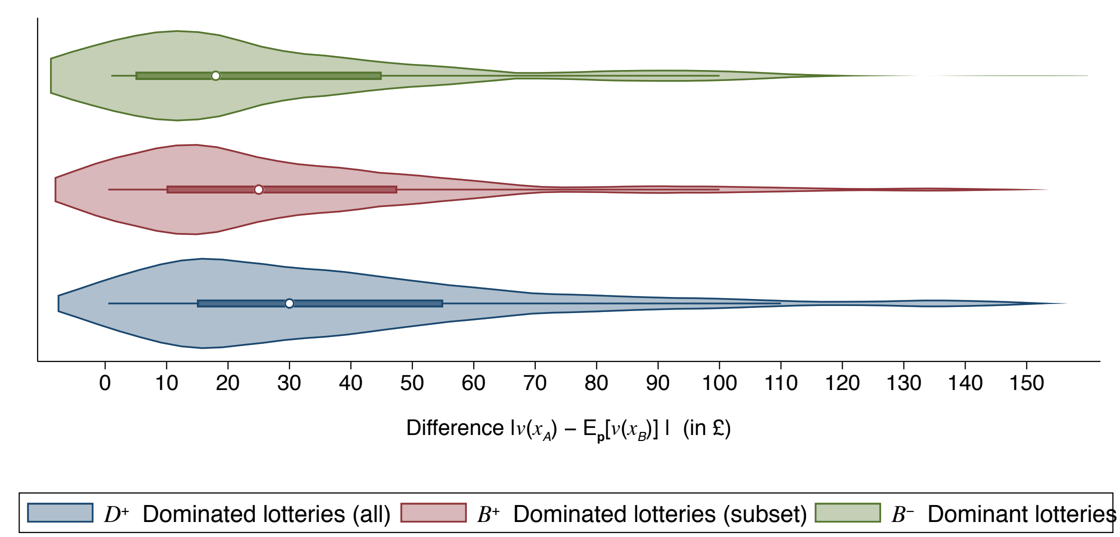
5.2 Size of SD violations

Even if they are quite prevalent, SD violations would be of low importance if they happened to be negligible in size. For each DP at which a violation occurred, one can assess this by comparing a respondent's value for the sure destination (Option A) to the expected value of the lottery (Option B).

Result 4. *The observed SD violations are economically significant. On average, respondents were willing to sacrifice £41 across all problems at which a violation occurred, or about 10% of the market value of a trip.*

Figure 5 presents violin plots of the monetary loss $|v(x_A) - E_{\mathbf{p}}[v(x_B)]|$ for each type of DP. The mean loss was £46 for violations on $\mathcal{D}^+$, £33 on $\mathcal{B}^+$, and £29 on $\mathcal{B}^-$. The loss distributions exhibit a large spread, reflecting heterogeneity across respondents and DPs. Given the large positive skew, median losses are smaller, but remain sizable (£32 on $\mathcal{D}^+$, £25 on $\mathcal{B}^+$, and £18 on $\mathcal{B}^-$). Although violations in favor of randomization appear slightly larger in size, the differences in monetary loss for SD violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ are not statistically significant.

Figure 5: Expected value difference with the sure option (in £)



Notes: Violin plots of the difference in monetary value between Options A and B, $|v(x_A) - E_{\mathbf{p}}[v(x_B)]|$, for all DPs in which a respondent violated stochastic dominance on (i) $\mathcal{D}^+$ (all dominated lotteries; N= 369); (ii) $\mathcal{B}^+$ (subset of dominated lotteries; N= 194); (iii) $\mathcal{B}^-$ (dominant lotteries; N= 136). In each plot, the white dot corresponds to the median, the box to the interquartile range, and the spikes extend to the upper- and lower-adjacent values. Differences above £150 were removed for visibility (11 observations for "Dominated lotteries (all)" and 1 observation for "Dominant lotteries").

Among respondents who designed a lottery as their favorite option, the average monetary loss relative to the highest-valued destination is £49 ($\approx$ 12% of value), very close to the average loss on $\mathcal{D}^+$. In sum, randomization decisions have non-trivial monetary implications.

5.3 Shape of SD violations

I now examine the shape of violations by assessing their compatibility with the monotonicity properties defined in Section 2.1.

Result 5. *The observed SD violations have a specific structure: dominated lotteries are more likely to be chosen when they have a higher entropy, in violation of P-MON; on the other hand, they are less likely to be chosen when they contain lower-ranked options i.e., X-MON is globally respected.*

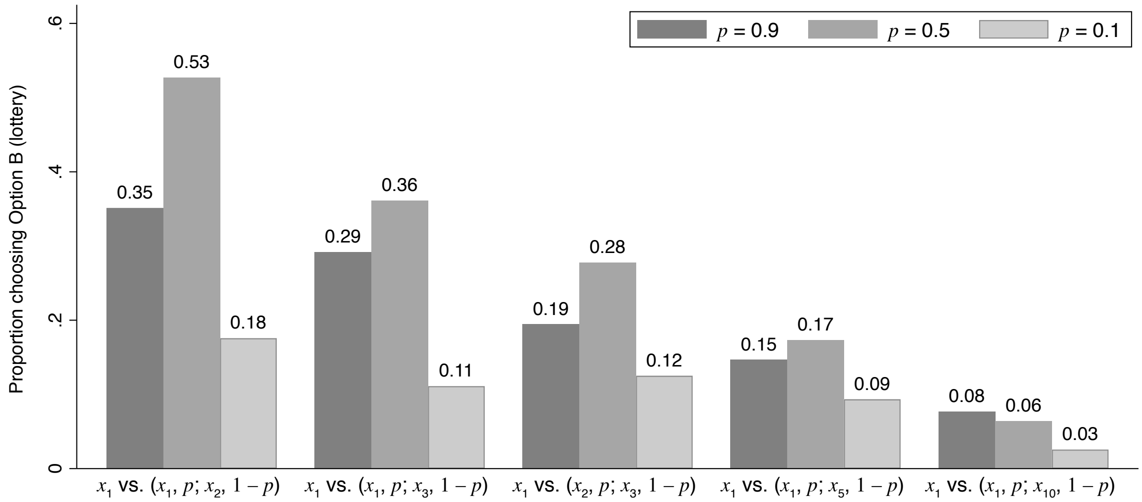
Fixing the support, Figure 6 shows how the propensity to choose a dominated lottery depends on the probability p of its higher-ranked outcome. While P-MON would dictate a (weakly) lower proportion of SD violations as p goes down, respondents were generally more likely to choose the dominated lottery for $p = 0.5$ than $p = 0.9$. Most strikingly, while 53% of respondents preferred the lottery $(x_1, 0.5; x_2, 0.5)$ to x_1, only 35% preferred $(x_1, 0.9; x_2, 0.1)$ to x_1.[27] On the other hand, respondents were significantly less likely to choose the lottery for $p = 0.1$ compared to $p \in \{0.5, 0.9\}$, when lowering p does not increase outcome uncertainty.[28] An individual-level analysis confirms the aggregate patterns: 37% of respondents violated P-MON at least once, and 81% of P-MON violations occurred when moving from $p = 0.9$ to $p = 0.5$.

Fixing the probability p, Figure 6 also allows to examine violations of X-MON. Respondents' propensity to choose the dominated lottery decreases monotonically as a given destination is replaced with a lower-ranked outcome in the lottery. In other words, SD violations occur in a way that satisfies X-MON. This property continues to be globally satisfied for lotteries with larger supports (Figure B3). In total, 72% of respondents never violate X-MON (85% do so at most once).

[27] A two-sample test of proportions yields p = 0.03. While differences for the other DPs are not individually significant, the pooled difference of 6.7 percentage points is significant at p = 0.02 in a regression of lottery choice on probability indicators (with $p = 0.9$ as the reference category), with standard errors clustered at the respondent level.

[28] The pooled differences of 17.3 and 10.5 percentage points with respect to $p = 0.5$ and $p = 0.9$ are statistically significant at p < 0.001 in a regression of lottery choice on probability indicators.

Figure 6: Proportion of SD violations on $\mathcal{B}^+$ as a function of p

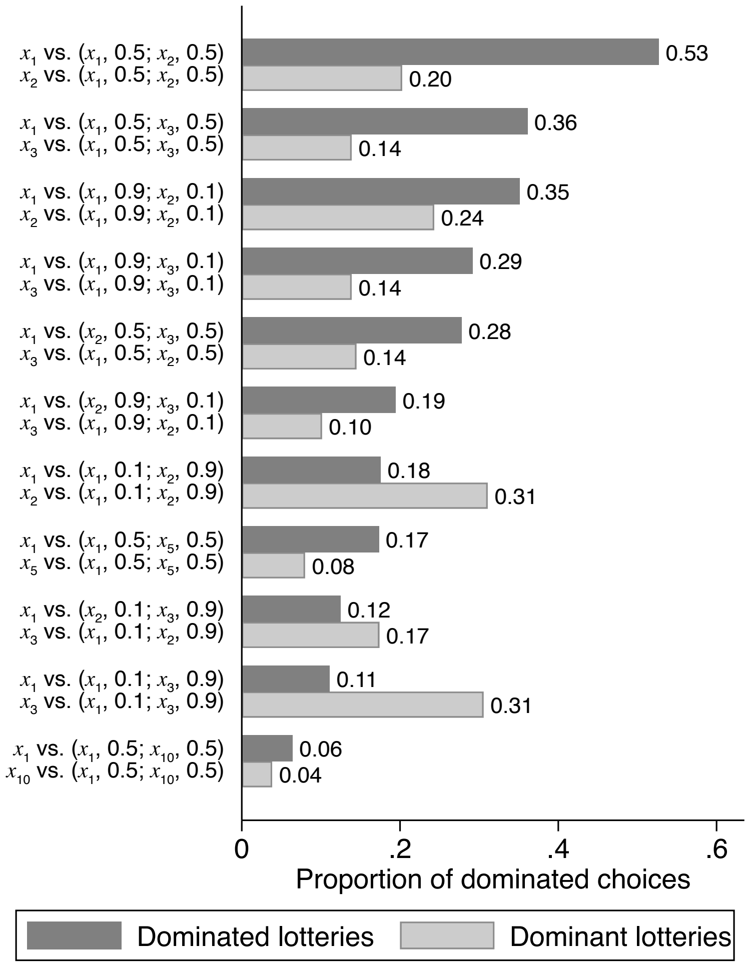


Notes: Each bar corresponds to the proportion of respondents who violated stochastic dominance in problem $\{x_i, (x_j, p; x_k, 1-p)\}$ for $p \in \{0.1, 0.5, 0.9\}$ (clear dominance problems only). See Table B3 for the number of respondents in each bar.

Taken together, these findings are inconsistent with $U_{\alpha,\Psi}$-representations such that $\Psi(\mathbf{p}, \mathbf{v}) = \psi(\sum_{k=1}^{n} p_k(v_k - E_{\mathbf{p}}(v))^2)$ (a mean-variance model would violate X-MON) or $\Psi(\mathbf{p}, \mathbf{v}) = \psi(|\text{supp}(\mathbf{p})| - 1)$ (a mean-support model would satisfy P-MON). On the other hand, they are consistent with models requiring that $\Psi(\mathbf{p}, \mathbf{v}) = \psi(H(\mathbf{p}))$ where H is a valid measure of uncertainty (e.g., Shannon entropy, residual variance) and ψ is not weakly concave (see Section 2.2).

To investigate this further, I now examine the design decisions of respondents who chose to build a lottery when offered the option. The median respondent included 4 destinations in the support, with only 9% of respondents choosing a lottery with full support (all 10 destinations). Respondents tended to design high-entropy lotteries: normalizing the entropy measure so it lies between 0 and 1 regardless of the size of the support (i.e., with $H(\frac{1}{n}, ..., \frac{1}{n}) = 1$ for all $n > 1$) gives a mean entropy value of 0.88. Importantly, consistent with a DM trading off a higher entropy with a lower expected realization, respondents generally built negatively-skewed lotteries: 59% assigned a higher probability weight to higher-ranked options, with a mean skew of -0.75 (significantly different from 0 at p = 0.006). Overall, the findings from the design task are very much in line with those from the binary choice exercise (see Appendix C.1 for more details).

Figure 7: Proportion of SD violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$



Notes: The dark vs. light gray bars correspond to the proportion of respondents who violated stochastic dominance on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ (clear dominance problems only). See Table B3 for the number of respondents in each bar.

Result 6. *Violations of stochastic dominance have a different shape for dominated vs. dominant lotteries. In particular, violations in favor of certainty are more likely to occur at near certainty i.e., for low-entropy lotteries.*

Figure 7 presents the proportion of SD violations for each problem in $\mathcal{B}^+$ (dominated lotteries) paired with its symmetric counterpart in $\mathcal{B}^-$ (dominant lotteries). While preference for randomization generally dominates, respondents were more inclined to express a preference for certainty for lotteries with a low probability of the higher-ranked outcome. For example, 31% of respondents preferred x_2 to $(x_1, 0.1; x_2, 0.9)$, compared to only 18% who preferred $(x_1, 0.1; x_2, 0.9)$ to x_1. Violations in favor of the sure outcome happen more often for $p = 0.1$ than for $p = 0.5$, that is, making the outcome more uncertain does not lead to a stronger preference for certainty. These differences in the shape of violations again suggest that preferences for randomization vs. certainty reflect different phenomena.

6 Interpretation of violations

Having documented the existence of SD violations, I now turn to the interpretation of these violations. The main explanation proposed is that violations in favor of randomization emerge because of the anticipatory utility benefits of surprises. In this section, I study support for this conjecture and alternative drivers of randomization behavior by combining a set of empirical facts on potential mechanisms (Section 6.1) with a discussion of theories that could rationalize the evidence (Section 6.2).

6.1 Empirical evidence on potential drivers

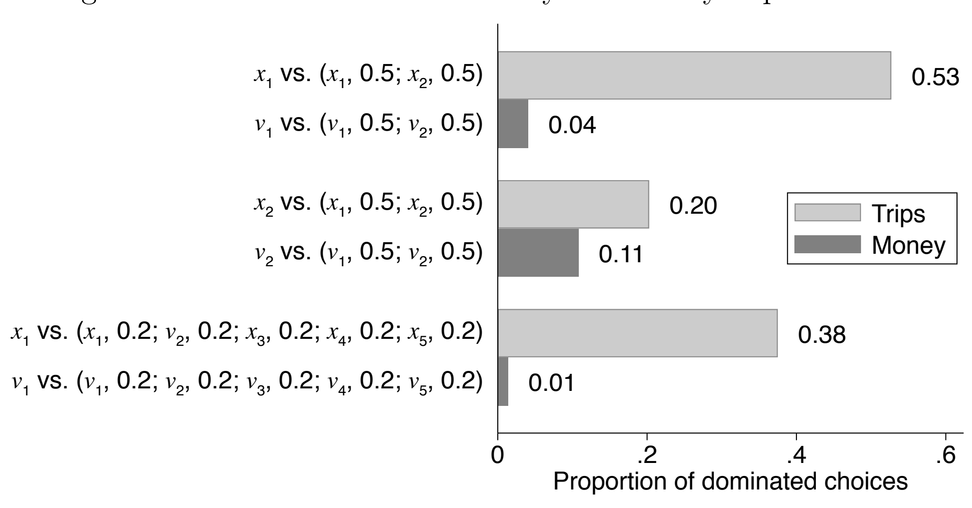
6.1.1 Preferences for delaying the resolution of uncertainty

Result 7. *Among respondents who designed a holiday trip lottery as their favorite option, about three quarters preferred to postpone learning about their destination and the majority were willing to pay for this option.*

Although the option of learning today was listed first, 74% of respondents preferred to maintain the surprise of the destination for 48 hours or more. Conditional on preferring a delay, the mean delay was 94 days (s.d. = 104 days), with chosen delays covering an average of 54% (s.d. = 44%) of the distance to the anticipated travel date. The chosen delays vary greatly, however, with respondents balancing the costs and benefits of maintaining uncertainty in different ways (Figure C2).

Many respondents sacrificed money for these non-trivial delays. In particular, 74% of the respondents who set a delay had a positive willingness to pay to learn on their chosen date instead of right away. Overall, this means that 54% of those who built a lottery (and 30% of all respondents) were willing to give up at least £5 for the delay option (Figure C3). Among them, the mean WTP for delay was £15.74,

Figure 8: SD violations for monetary vs. holiday trip lotteries



Notes: The light vs. dark gray bars correspond to the proportion of SD violations for the trip vs. monetary lotteries (clear dominance problems only). The p-value in a two-sample test of proportions for trips vs. money is p < 0.001 (top), p = 0.11 (middle) and p < 0.001 (bottom). See Table B3 for the number of respondents in each bar.

amounting to 5.8% of the expected value of the lottery they designed.[29]

6.1.2 Do violations survive in the world of money?

Result 8. *When the trip destinations are replaced in the surprise lotteries by their monetary valuations, SD violations almost entirely disappear.*

To test whether violations survive in the world of money, respondents were asked to consider 3 DPs that were identical to those presented earlier, except for one difference: each trip destination x_k was replaced by the respondent's valuation v_k for that destination. In other words, the prizes were monetary payments to be sent on the date at which the respondent had planned to travel and, for the lottery option, with the prize revealed one week before payment.[30] As shown in Figure 8, violations almost completely vanish for money. While over 50% of respondents chose $(x_1, 0.5; x_2, 0.5)$

[29]WTP for delay is also positively correlated with the fraction of SD violations on $\mathcal{D}^+$ (Spearman $\rho = 0.34$, p = 0.023, N = 46), but not with other measures of SD violations.

[30]The 3 DPs were: (i) A: $(x_1, 1)$ vs. B: $(x_1, 0.5; x_2, 0.5)$; (ii) A: $(x_2, 1)$ vs. B: $(x_1, 0.5; x_2, 0.5)$ and (iii) A: $(x_1, 1)$ vs. B: $(x_1, 0.2; x_2, 0.2; x_3, 0.2; x_4, 0.2; x_5, 0.2)$ (with x_k replaced by v_k). In case $v_1 \leq v_2$, I replaced (v_1, v_2) with $(400, 350)$ in the first two DPs and $(v_1, v_2, v_3, v_4, v_5)$ with $(400, 370, 350, 320, 300)$ in the last DP. To make one-to-one comparisons, I discard respondents for whom $v_1 \leq v_2$, and more generally, those for whom the DP in question was not a clear dominance problem. Numbers are very similar when pooling all respondents.

over x_1, only 4% chose $(v_1, 0.5; v_2, 0.5)$ over v_1. Delay of the resolution of uncertainty is thus not a sufficient condition for preference for randomization (or certainty); the domain of randomization appears to play a major role (see Section 6.2).

6.2 Competing theories

I now bring all the facts presented in this paper to discuss their compatibility with various classes of theories.

6.2.1 Surprise enjoyment and anticipatory utility

The leading hypothesis of this paper is that people may violate stochastic dominance to enjoy the surprise of some pleasurable outcome. In support of this interpretation, I find that (i) when considering trips, preference for randomization overall dominates preference for certainty, with the two being distinct (Results 2, 3, 6); (ii) randomization is more frequent for higher-entropy lotteries, increasing scope for surprise (Result 5); (iii) randomization decisions tend to come with a preference for delayed resolution of uncertainty (Result 7); (v) preference for randomization disappears when replacing trips with equivalent monetary amounts (Result 8).

Surprises may create enjoyment for at least two reasons. First, uncertainty about the destination creates thrill and excitement ahead of the resolution of uncertainty (e.g., the anticipated pleasure of opening the envelope). Second, uncertainty offers the possibility to daydream about different worlds. In line with this, when asked about their motives, a majority of respondents with a favorable view of surprise trips ranked either the excitement until the destination is revealed or the possibility of daydreaming as the #1 factor (out of a list of six - see Appendix D.2). Furthermore, these concerns are positive predictors of SD violations for dominated lotteries and valuations of the wildcard trip (Appendix D.1). Given the difference observed for trips vs. money, below I form conjectures on the necessary conditions for randomization behavior.

Outcome valence An important characteristic of holiday trips is that they are pleasurable events, which people typically enjoy thinking about. One conjecture is that surprise enjoyment ($\alpha > 0$) requires outcomes to have positive valence. In other words, preference for randomization should vanish for goods or experiences that trigger negative emotions (e.g., a medical appointment), possibly replaced by a preference for certainty ($\alpha < 0$). The effect of surprise should also disappear for ordinary goods that have neutral valence (e.g., laundry detergent), for they fail to trigger emotions ($\alpha = 0$). When presented abstractly, monetary payments likely belong to this category. An open question is whether violations would occur more often with cash if people first contemplated what they might do with various sums of money: while the psychological investment might trigger an emotional response, its effect might be limited in the absence of a commitment to purchasing the contemplated goods.

Outcome valence has been found to play a role in many areas of decision-making. For instance, people may prefer for negative (vs. positive) consumption events to occur earlier (vs. later) due to anticipatory utility concerns (Loewenstein, 1987). Similarly, they usually prefer to savor good news and ignore bad news (Golman, Hagmann, and Loewenstein, 2017; Golman, Loewenstein, Molnar, and Saccardo, 2022). To account for the importance of valence, a range of theories have put forward the role of emotions triggered by the attention devoted to possible outcomes. In the context of risk, Rottenstreich and Hsee (2001) propose that affect-rich and affect-poor outcomes are compared differently depending on whether their occurrence is certain or uncertain, because uncertainty triggers stronger emotional responses (e.g., hope of gain and fear of loss) for hedonic goods than utilitarian goods.[31] More recently, Bolte and Raymond (2022) consider a model in which attention has both an instrumental and an emotional value, such that the DM chooses to deploy more attention towards higher-payoff outcomes in order to raise their "attention utility".

Outcome multidimensionality The positive valence of outcomes might be a necessary but not sufficient condition for preference for randomization. After all, if the DM enjoys dreaming about a given outcome, offering certainty will allow them to fully specialize in the contemplation of this outcome. Thus, it appears important that (i) each outcome presents something pleasant to think about (*no dominance across attributes*); (ii) the DM derives utility from entertaining diverse possibilities (*variance in attributes*). Supporting this conjecture, Buechel and Li (2022) find that preference for "mysterious consumption" disappears when there is no variance in the type of good (e.g., gift certificates to two bookstores) or there is a clear ordering on quality. In line with this, the proposed trips offered a range of experiences at the

[31] For instance, the authors find that even if two coupons to cover university tuition vs. go on a European trip are equally valued, a 1% chance of the European trip coupon is more highly valued than a 1% chance of the tuition coupon.

same market value, creating both variance and no obvious dominance (Section 4.1).

Surprises open information gaps (Golman and Loewenstein, 2018; Golman, Gurney, and Loewenstein, 2021), which creates wonder by opening new questions on a pleasurable topic (e.g., will I go to the beach or the mountains?). By contrast, committing to a sure outcome greatly constrains the variety of dreams that can be entertained. If there are diminishing returns to constantly dreaming about the same place, the DM might spend more time dreaming about trips with more upside, but do some balancing so as to avoid boredom. Such a model would generate the preference for negatively-skewed lotteries observed in the experiment. As another implication of this theory, randomization behavior should go down if people are presented abstractly with the choice between (4* hotel 1, beach 1) and a lottery over (4* hotel 1, beach 1) and (4* hotel 2, beach 2), which limits the mind's exploration.

6.2.2 Decision noise and inability to choose

Evaluating (lotteries over) holiday trips is an unusual task, which begs the question: do violations simply reflect mistakes and/or an inability to decide?

Decision noise While decision noise is unavoidable in such an experiment, multiple reasons suggest that it cannot be the primary factor behind the observed violations. First, respondents reported preferences over destinations in a consistent manner across elicitation methods, suggesting they could evaluate the trips coherently. Second, SD violations were computed only on the subset of clear dominance problems, thus minimizing the chances of Type I error. Third, violations have a specific structure e.g., those in favor of randomization occur more often for high-entropy lotteries. Forth, the correlation between SD violations on $\mathcal{B}^+$ and $\mathcal{B}^-$ is negative, with each type reflecting a different psychology. Importantly, none of the standard stochastic choice models (fixed error model, random valuations, random utility) would predict P-MON violations, and they would produce either a positive or zero correlation between violations of each type (Appendix E). Finally, while revealed perturbed utility models (Fudenberg, Iijima, and Strzalecki, 2015) could produce the observed violations for a carefully chosen cost function, they make no predictions about preference for delay.

Inability to decide and choice delegation One possibility is that respondents chose the trip lotteries to resolve their indecisiveness (Agranov and Ortoleva, 2017).

However, if a randomization device was needed to help them decide, they could have simply selected the sure option in each DP and let the random incentive system make the final selection. Indeed, the problems were presented in alternating order, with the sure option being dominant in one case (e.g., x_1) and dominated in the next case (e.g., x_2). Achieving randomization in this way was easy because all problems were evaluated on the same page. Importantly, randomizing across (instead of within) DPs offered the benefits of randomization without the cost of learning the destination with a long delay. Despite this, no respondent reported adopting this strategy. In addition, those who viewed surprise trips favorably ranked their ability to delegate the decision as one of the least important reasons for liking the lotteries (Figure D4). Finally, models of incompleteness would need to explain (i) why randomization behavior is associated with a preference for delay and (ii) why respondents generally preferred negatively-skewed lotteries (suggesting a preference for higher-ranked outcomes).

6.2.3 Non-standard preferences for risk and uncertainty

Having discussed stochastic choice (whether deliberate or due to mistakes), I now examine deterministic theories of choice under risk, which may generate SD violations.

Utility from gambling The most straightforward approach is to assume that people derive utility from the act of gambling per se, and use a different utility function to evaluate certain and uncertain outcomes. Assuming a discontinuity at certainty as in Diecidue, Schmidt, and Wakker (2004) can generate SD violations in favor randomization. However, without making further assumptions, these models cannot explain (i) why violations occur for trips but not money; (ii) why they sometimes happen for $(x_1, 0.5; x_2, 0.5)$ but not for $(x_1, 0.9; x_2, 0.1)$. In fact, because Diecidue, Schmidt, and Wakker (2004) maintain the independence axiom over the set of risky gambles, $(x_1, 0.9; x_2, 0.1) \succ (x_1, 0.1; x_2, 0.9)$ would imply $(x_1, 0.9; x_2, 0.1) \succ (x_1, 0.5; x_2, 0.5)$ i.e., P-MON must be satisfied.

Non-linear probability weighting The class of representations in this paper posit that the DM correctly perceives probabilities. While this is a restriction, probability weighting appears insufficient to explain the evidence. First, for SD violations to be generated, the weights $\tilde{\pi}_k$ on each x_k must be such that $\sum_{\{k \,:\, x_k \,\in\, \text{supp}(\mathbf{p})\}} \bar{\pi}_k \neq 1$, which immediately rules out models of rank-dependent probability weighting (Quiggin, 1982; Tversky and Kahneman, 1992). Second, although simple non-linear probability weighting as in Kahneman and Tversky (1979) (without an editing stage) could rationalize $(x_1, 0.5; x_2, 0.5) \succ x_1 \succ x_2$ for $\pi(0.5) > 0.5$, this requirement would violate *subcertainty* ($\pi(p) + \pi(1-p) < 1$ for all p), a property necessary to generate the Allais Paradox. Third, even allowing for this, existing parametrizations (as listed in Stott (2006)) are not flexible enough to accommodate the observed P-MON violations. Fourth, a model with probability distortions would need to assume outcome dependence to rationalize the absence of violations when using money.[32]

Disappointment and regret Models of expectation-based reference dependence such as the Kőszegi and Rabin (2007) CPE model or the models of disappointment aversion of Bell (1985) and Loomes and Sugden (1986) can generate SD violations in favor of randomization (including P-MON violations) only if $\lambda < 0$; this assumption would contradict the empirical evidence on loss aversion and also generate violations of X-MON, which are not observed in the data.[33] Furthermore, data on self-reported motives suggests that disappointment and regret are at best secondary determinants of positive attitudes towards surprise trips; instead, disappointment emerges as the leading factor among those who dislike the trip lotteries (Appendix D.2).

Ambiguity aversion and hedging One could treat a trip x_j as a Savage act $\{f_j(\omega)\}_{\omega\in\Omega}$, yielding a different payoff in each state of the world $\omega \in \Omega$. The surprise trips in the experiment could thus be conceived as compound lotteries mixing layers of objective and subjective uncertainty. In such a framework, a DM might randomize due to hedging motives e.g., if they hold multiple priors $\pi \in \Delta(\Omega)$ and believe nature or the experimenter to be adversarial in the sense of Gilboa and Schmeidler (1989). Such a DM might say $pf_j + (1-p)f_k \succ f_j \succ f_k$ provided that $f_j(\omega) > f_k(\omega)$ and $f_k(\omega') > f_j(\omega')$ for some $\omega, \omega' \in \Omega$ (no statewise dominance of one act). However, such a model cannot generate the P-MON violations observed; furthermore, self-reported mistrust in the experiment is low overall and does not predict preferences for randomization or delay (Appendix F.4). In addition, if respondents simply wanted to

[32]Even models in which distortions come from motivated beliefs such as the wishful thinking model of Caplin and Leahy (2019) will typically satisfy stochastic dominance.

[33]In fact, Masatlioglu and Raymond (2016) show that the Kőszegi and Rabin (2007) CPE model with quadratic gain-loss utility admits an equivalent mean-variance representation; such a representation violates X-MON as discussed in Section 2.2 and Appendix F.

hedge against subjective uncertainty, choosing the risky lotteries should be associated with a lower valuation of the wildcard trip i.e., a very ambiguous lottery; instead, I find a (weakly) positive relationship between the two (Appendix D.1).

6.2.4 Other psychological motives

Planning aversion One reason why people might prefer a lottery with delayed resolution of uncertainty to the guarantee of their favorite destination is that giving up control lowers the anxiety and pressure of planning the perfect holiday. Instead of rehearsing the script of their upcoming holiday, planning-averse people might want to pay *not* to have to think about it ahead of time. If planning aversion was the primary driver of randomization decisions, then most respondents should have preferred to postpone learning about their destination until the last moment. Against this, about two thirds of the respondents who designed a lottery chose to learn the location at least 3 weeks prior to departure, thus leaving plenty of time to plan (Figure C2).

Unawareness and forced experimentation Another benefit of letting fate decide on the destination is the chance to encounter situations that one would have never chosen to enter spontaneously. In other words, randomization offers a commitment to novelty and experimentation, thus creating unexpected opportunities to learn and challenge one's prior beliefs. Wildcard trips are extreme versions of that: they are not only ambiguous lotteries, but they entail an element of unawareness, maximizing the potential to be pleasantly surprised by an unexpected experience. Data on self-reported motives provides suggestive evidence in this direction (Appendix D).

7 Discussion

This paper studies whether people may violate stochastic dominance in favor of randomization in order to enjoy the anticipatory utility benefits of surprises. I generate the evidence using experiential goods that typically trigger a lot of anticipation i.e., holiday trips. The opportunity for surprise is operationalized by creating risky lotteries over destinations with a delayed resolution of uncertainty. I study how people compare various trip lotteries to the guarantee of a better (or worse) destination.

I find that violations occur on average about 20% of the time, but with significant heterogeneity across respondents. Preference for randomization and preference for certainty appear to be largely distinct phenomena, with the former somewhat dominating the latter in this context. Violations have a specific structure e.g., the choice of dominated lotteries is more frequent if they have a high entropy, a finding which cannot be simply rationalized by noise. Results from a lottery-building exercise show that those who randomize generally prefer to postpone the resolution of uncertainty. Finally, violations almost entirely disappear in a control task in which each trip is replaced by its corresponding valuation, suggesting that the experiential and/or multidimensional nature of the good might be an important precondition for randomization behavior. By stepping outside of traditional domains of inquiry, this paper raises many new questions. Below I come back to some of the limitations of this study and articulate intriguing open questions left for future research:

Going from objective to subjective uncertainty The surprise trips in this experiment were restricted to the space of risky lotteries (objective probabilities and known support). While randomization behavior in this risky setting is predictive of attitudes towards more radical uncertainty as measured through valuations of a "wild-card trip," the correlation is modest in size. An open question is how introducing ambiguity and/or unawareness of the possible outcomes might affect people's preferences for surprise goods. Moving away from objective risk raises major challenges for the quantification of the uncertainty implied by a choice. From a product design perspective, one interesting question is how to optimally engineer uncertainty about the various outcomes. Some companies charge for the option to remove certain options from the set of possibilities, effectively allowing customers to insure themselves against certain undesirable realizations. One theoretical question is what the optimal insurance mechanism looks like in such a context.

Timing of the resolution of uncertainty This experiment only considers one-shot resolution of uncertainty with some fixed delay (i.e., no earlier than one week before travel). When offered the possibility to customize the delay, respondents exhibit highly heterogeneous preferences and many are willing to give up money for their chosen delay. Interestingly, existing companies rarely offer a delay customization option; the findings of this experiment suggest they could charge a premium for this opportunity. More research should investigate the determinants of chosen delays such as the distance to the consumption date, the amount of uncertainty to be resolved, or the associated costs and benefits of postponing its resolution. Relative to one-shot resolution, gradual resolution of uncertainty (e.g., by sending clues) might boost anticipatory utility by reminding the DM of the upcoming event and stimulating their curiosity, while minimizing the costs of a lack of preparation. Alternatively, surprise could be maximized by resolving the uncertainty at a random date, a type of time lottery (DeJarnette, Dillenberger, Gottlieb, and Ortoleva, 2020).

Preference for randomization vs. certainty While preference for randomization overall dominates in the experiment, a non-negligible fraction of respondents do violate stochastic dominance in favor of certainty. It is also worth noting that the design was not fully symmetric. Since the goal was to understand the link between positive surprises and randomization behavior, preference for certainty was examined only on a subset of problems (two-outcome lotteries). On this subset, preference for randomization generally dominates, but violations in favor of certainty are more frequent when the worse outcome has high probability. In addition, the experiment mostly examined violations for lotteries among higher-ranked destinations, again to focus on the positive value of surprises; preference for certainty might manifest itself more strongly for less desirable outcomes e.g., comparing $(x_{10}, 1)$ vs. $(x_9, p; x_{10}, 1-p)$. Beyond, a plausible conjecture is that preference for certainty would dominate for lotteries over negative consumption events (i.e., for which WTP < 0) or a mixture of positive and negative events (e.g., if the DM suffers from loss aversion).

Population prevalence and predictors of heterogeneity The vast majority of respondents were found to exhibit at most one of the two tendencies to either seek or avoid uncertainty. One question is whether the two kinds of violations capture fundamentally distinct personality types who adopt the same attitudes towards randomization across a range of domains (Agranov, Healy, and Nielsen, 2023). As an alternative, there might be substitution effects e.g., people seeking a surprise adventure to compensate for their lack of experimentation in everyday life. More research should be conducted to understand the individual characteristics correlated with each type of behavior and the level of consistency exhibited across decision domains.

References

Agranov, M., P. J. Healy, and K. Nielsen. (2023). "Stable Randomisation". *Economic Journal* 133.655, 2553–79. [40]

Agranov, M., and P. Ortoleva. (2017). "Stochastic Choice and Preferences for Randomization". *Journal of Political Economy* 125.1, 40–68. [7, 35]

Agranov, M., and P. Ortoleva. (2022). "Revealed Preferences for Randomization: An Overview". *AEA Papers and Proceedings* 112, 426–30. [6]

Azrieli, Y., C. P. Chambers, and P. J. Healy. (2018). "Incentives in Experiments: A Theoretical Analysis". *Journal of Political Economy* 126.4, 1472–503. [14]

Bell, D. E. (1985). "Disappointment in Decision Making Under Uncertainty". *Operations Research* 33.1, 1–27. [5, 37, 23]

Birnbaum, M. H. (2005). "A Comparison of Five Models that Predict Violations of First-Order Stochastic Dominance in Risky Decision Making". *Journal of Risk and Uncertainty* 31, 263–87. [2]

Bolte, L., and C. Raymond. (2022). "Emotional Inattention". Working paper. [34]

Buechel, E. C., and R. Li. (2022). "Mysterious Consumption: Preference for Horizontal (vs. Vertical) Uncertainty and the Role of Surprise". *Journal of Consumer Research* 49.6, 987–1013. [6, 34]

Caplin, A., and J. Leahy. (2001). "Psychological Expected Utility Theory and Anticipatory Feelings". *Quarterly Journal of Economics* 116.1, 55–79. [7]

Caplin, A., and J. Leahy. (2019). "Wishful Thinking". Working Paper 25707. National Bureau of Economic Research. [37]

Carrera, M., H. Royer, M. Stehr, and J. Sydnor. (2020). "The structure of health incentives: Evidence from a field experiment". *Management Science* 66.5, 1890–908. [5]

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva. (2015). "Cautious Expected Utility and the Certainty Effect". *Econometrica* 83.2, 693–728. [5]

Corgnet, B., S. Gächter, and R. Hernán González. (2020). "Working Too Much for Too Little: Stochastic Rewards Cause Work Addiction". Working Paper 12992. IZA Institute of Labor Economics. [5]

DeJarnette, P., D. Dillenberger, D. Gottlieb, and P. Ortoleva. (2020). "Time lotteries and stochastic impatience". *Econometrica* 88.2, 619–56. [40]

Diecidue, E., U. Schmidt, and P. P. Wakker. (2004). "The Utility of Gambling Reconsidered". *Journal of Risk and Uncertainty* 29, 241–59. [5, 36]

Dwenger, N., D. Kübler, and G. Weizsäcker. (2018). "Flipping a coin: Evidence from university applications". *Journal of Public Economics* 167, 240–50. [7]

Ely, J., A. Frankel, and E. Kamenica. (2015). "Suspense and Surprise". *Journal of Political Economy* 123.1, 215–60. [6, 11]

Engelmann, J., M. Lebreton, N. Salem-Garcia, P. Schwardmann, and J. J. van der Weele. (2022). "Anticipatory Anxiety and Wishful Thinking". *CEPR Press Discussion Paper No. 17665*, [7]

Falk, A., and F. Zimmermann. (2023). "Attention and dread: Experimental evidence on preferences for information". *Management Science, forthcoming*, [7]

Frankel, A., and E. Kamenica. (2019). "Quantifying Information and Uncertainty". *American Economic Review* 109.10, 3650–80. [9]

Fudenberg, D., R. Iijima, and T. Strzalecki. (2015). "Stochastic choice and revealed perturbed utility". *Econometrica* 83.6, 2371–409. [35]

Ganguly, A., and J. Tasoff. (2017). "Fantasy and dread: The demand for information and the consumption utility of the future". *Management Science* 63.12, 4037–60. [7]

Geanakoplos, J., D. Pearce, and E. Stacchetti. (1989). "Psychological games and sequential rationality". *Games and Economic Behavior* 1.1, 60–79. [6]

Gilboa, I., and D. Schmeidler. (1989). “Maxmin expected utility with non-unique prior”. *Journal of Mathematical Economics* 18.2, 141–53. [37, 24]

Gneezy, U., J. List, and G. Wu. (2006). “The Uncertainty Effect: When a Risky Prospect is Valued Less than its Worst Possible Outcome”. *Quarterly Journal of Economics* 121.4, 1283–309. [2]

Golman, R., N. Gurney, and G. Loewenstein. (2021). “Information gaps for risk and ambiguity.” *Psychological Review* 128.1, 86. [35]

Golman, R., D. Hagmann, and G. Loewenstein. (2017). “Information Avoidance”. *Journal of Economic Literature* 55.1, 96–135. [34]

Golman, R., and G. Loewenstein. (2018). “Information Gaps: A Theory of Preferences Regarding the Presence and Absence of Information”. *Decision* 5.3, 143–64. [5, 35]

Golman, R., G. Loewenstein, A. Molnar, and S. Saccardo. (2022). “The demand for, and avoidance of, information”. *Management Science* 68.9, 6454–76. [34]

Gul, F. (1991). “A Theory of Disappointment Aversion”. *Econometrica* 59.3, 667–86. [5]

Kahneman, D., and A. Tversky. (1979). “Prospect Theory: An Analysis of Decision under Risk”. *Econometrica* 47.2, 263–91. [2, 5, 37]

Kőszegi, B., and M. Rabin. (2007). “Reference-Dependent Risk Attitudes”. *American Economic Review* 97.4, 1047–73. [5, 37, 23]

Levitt, S. D. (2020). “Heads or Tails: The Impact of a Coin Toss on Major Life Decisions and Subsequent Happiness”. *Review of Economic Studies* 88.1, 378–405. [7]

Levitt, S. D., J. A. List, S. Neckermann, and S. Sadoff. (2016). “The behavioralist goes to school: Leveraging behavioral economics to improve educational performance”. *American Economic Journal: Economic Policy* 8.4, 183–219. [5]

List, J. A., and J. F. Shogren. (1998). "The deadweight loss of Christmas: comment". *American Economic Review* 88.5, 1350–55. [5]

Lockwood, B., H. Allcott, D. Taubinsky, and A. Y. Sial. (2021). "What Drives Demand for State-Run Lotteries? Evidence and Welfare Implications". 6061. NBER Working Paper Series. [6]

Loewenstein, G. (1987). "Anticipation and the valuation of delayed consumption". *Economic Journal* 97.387, 666–84. [7, 34]

Loomes, G., and R. Sugden. (1986). "Disappointment and Dynamic Consistency in Choice under Uncertainty". *Review of Economic Studies* 53.2, 271–82. [5, 37, 23]

Masatlioglu, Y., A. Y. Orhun, and C. Raymond. (2023). "Intrinsic information preferences and skewness". *American Economic Review* 113.10, 2615–44. [7]

Masatlioglu, Y., and C. Raymond. (2016). "A Behavioral Analysis of Stochastic Reference Dependence". *American Economic Review* 106.9, 2760–82. [37, 24]

Modirshanechi, A., J. Brea, and W. Gerstner. (2022). "A taxonomy of surprise definitions". *Journal of Mathematical Psychology* 110, 102712. [6]

Nielsen, K., and J. Rehbeck. (2022). "When Choices Are Mistakes". *American Economic Review* 112.7, 2237–68. [2]

Puri, I. (2023). "Simplicity and Risk". Working paper. [2]

Quiggin, J. (1982). "A theory of anticipated utility". *Journal of Economic Behavior & Organization* 3.4, 323–43. [2, 5, 36]

Rottenstreich, Y., and C. K. Hsee. (2001). "Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk". *Psychological Science* 12.3, 185–90. [34]

Ruffle, B. J., and O. Tykocinski. (2000). "The Deadweight Loss of Christmas: Comment". *American Economic Review* 90.1, 319–24. [5]

Schelling, T. (1987). "The Mind as a Consuming Organ". In *The Multiple Self*. Edited by J. Elster. Cambridge University Press, 177–96. [7]

Shannon, C. E. (1948). "A mathematical theory of communication". *Bell system technical journal* 27.3, 379–423. [6]

Shen, L., A. Fishbach, and C. K. Hsee. (2015). "The motivating-uncertainty effect: Uncertainty increases resource investment in the process of reward pursuit". *Journal of Consumer Research* 41.5, 1301–15. [5]

Solnick, S. J., and D. Hemenway. (1996). "The deadweight loss of Christmas: comment". *American Economic Review* 86.5, 1299–305. [5]

Stott, H. P. (2006). "Cumulative prospect theory's functional menagerie". *Journal of Risk and Uncertainty* 32.2, 101–30. [37]

Thakral, N., and L. T. Tô. (2022). "Anticipation and Consumption". Working paper. [7]

Tversky, A., and D. Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty". *Journal of Risk and Uncertainty* 5.4, 297–323. [2, 5, 37]

Waldfogel, J. (1993). "The Deadweight Loss of Christmas". *American Economic Review* 83.5, 1328–36. [5]

Zhang, X., and S. Zhong. (2020). "Putting preference for randomization to work". Working paper. [7]

Online Appendix

A Experimental procedures and sample

A.1 Instructions, recruitment, and sample characteristics

The experiment was conducted online using Qualtrics (preview link here); a PDF of the instructions is also available as a supplemental appendix and on the project OSF page: https://osf.io/ya7x6/.[34] Prospective participants first signed up for the study by completing a consent form (preview link here) and providing an email address. Enrollment was on a first-come, first-served basis and capped at 100 respondents. Those who signed up were emailed the survey within 24 hours, together with a unique ID. In total, 83 finished the survey (13 left the survey with ≤ 20 % completed and 4 were timed out). Table A1 presents descriptive statistics about the sample.

A.2 Changes in preferences with the spread of the pandemic

The data collection took place in March 2020, at a time of growing travel restrictions due to the spread of COVID-19. One question is whether preferences for travel and surprise holidays shifted as a result. While I do not have exogenous variation to give a causal answer, I exploit natural variation in the survey completion date to shed some light on this question. On 9 March 2020, Italy was the first country worldwide to impose a national lockdown, an event which made the consequences of the spread of the pandemic particularly salient. Coincidentally, it was on this same day that participants from the LSE lab pool were emailed about the study, generating a spike in survey completion (45% of responses on that day). Below I examine potential differences in respondents' decisions based on whether they completed the survey after the lockdown announcement of March 9$^{\text{th}}$ (i.e., from March 10$^{\text{th}}$ onwards, N = 42) or up to that date (N = 41).[35]

[34] Several features of the Qualtrics editor changed since the survey was programmed, implying minor differences in user experience relative to the survey taken by respondents (e.g., text width, character alignment, etc.); the text is identical.

[35] The official announcement of a nationwide lockdown was made by Giuseppe Conte, then Italian prime minister, at a press conference late on Monday 9 March 2020 (https://www.theguardian.com/world/2020/mar/09/coronavirus-italy-prime-minister-country-lockdown).

Table A1: Sample characteristics

	Mean	SD	Min	Max	N
Male respondent	0.59	0.49	0	1	83
Age	27.1	7.8	18	50	83
Number of trips abroad in last 12 months	4.0	3.0	0	20	83
Number of EU countries visited in list of 10	4.9	2.2	0	9	83
Time to chosen travel date (in days)	229.7	158.4	23	636	83
Anticipated number of travel partners					
None	0.14	0.35	0	1	83
1 partner	0.73	0.44	0	1	83
2 partners or more	0.12	0.33	0	1	83
Recruitment channel					
LSE Lab (email)	0.87	0.34	0	1	83
Social media or other ads	0.13	0.34	0	1	83
Survey completed after Italian lockdown	0.51	0.50	0	1	83
Total survey time (in minutes)	44.0	19.7	11	120	83

Notes: Total survey time censored at 120 minutes for one respondent.
Recruitment channel recoded for 9 respondents based on text responses.

Main takeaways A summary of mean differences for a large range of outcome variables is presented in Table A2. Overall, preferences appear to have remained stable over the survey time window (5-19 March 2020), with two exceptions: (i) respondents pushed back their anticipated travel date post announcement of the Italian lockdown; (ii) their preference for preserving the surprise of the destination appears to have been stronger post-announcement.[36] In addition, the trip valuations remained nearly identically distributed for most destinations (no decrease in valuations post announcement) and I find no difference in people's preferences on various holiday trip criteria (e.g., quietness vs. vibrant atmosphere).

[36] Preferences for keeping the surprise of the destination remain stronger post-announcement even after controlling for the anticipated travel date in a regression framework.

Table A2: Preferences pre- and post-lockdown announcement (March 9th)

	Pre-	Post-	Diff.	s.e.	N
SD violations in binary choices					
Number of violations on $\mathcal{D}^+$	4.0	4.8	-0.8	(1.1)	83
Fraction of violations on $\mathcal{D}^+$	0.18	0.27	-0.09*	(0.05)	79
Number of violations on $\mathcal{B}^+$	2.3	2.4	-0.1	(0.6)	83
Fraction of violations on $\mathcal{B}^+$	0.21	0.30	-0.09	(0.06)	79
Number of violations on $\mathcal{B}^-$	1.4	1.8	-0.4	(0.6)	83
Fraction of violations on $\mathcal{B}^-$	0.13	0.19	-0.06	(0.06)	79
Favorite option					
Sure destination	0.44	0.45	-0.01	(0.11)	83
Trip lottery with fixed delay	0.39	0.40	-0.01	(0.11)	83
Trip lottery with custom delay	0.17	0.14	0.03	(0.08)	83
Valuation of wildcard trip (in £)	266.7	283.9	-17.3	(29.6)	83
Chose to reveal the destination later	0.61	0.87	-0.26**	(0.13)	46
Length of delay (in days)	29.1	109.6	-80.5***	(26.7)	46
Time to chosen travel date (in days)	176.4	281.7	-105.3***	(33.0)	83
% delay relative to travel date	23.4	56.6	-33.2**	(12.4)	46
WTP for delay (in £)	5.2	18.0	-12.8***	(4.0)	46
Surprise trip like rating (0-100)	52.9	53.6	-0.7	(6.8)	83
Total survey time (in minutes)	43.1	45.0	-1.9	(4.4)	83

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. `Chose to reveal the destination later` is an indicator = 1 if a respondent chose not to reveal the destination of their favorite lottery right away; `Length of delay (in days)` is the number of days the respondent was willing to postpone and `% delay relative to travel date` is the amount of delay as a percentage of `Time to chosen travel date (in days)`. The variable `WTP for delay (in £)` is the amount of money a respondent was willing to forego to delay learning about the destination of their favorite lottery; it is set equal to 0 for respondents who preferred not to delay the resolution of uncertainty in the first place (diff. $= -12.2$, s.e. $= 5.0$, p$= 0.02$ among the 34 respondents who completed the MPL question to measure WTP). Results qualitatively unchanged when using a Wilcoxon rank-sum test.

B Preferences $\succeq$ on X and implied SD violations

B.1 Incentive compatibility of the BDM mechanism

The Becker-DeGroot-Marschak (1964) mechanism (henceforth, BDM) is not incentive compatible if the DM directly values uncertainty. Depending on the DM's true valuation v, the bias in the report $\tilde{v}$ will be either positive (for low valuations) or negative (for high valuations). As a result, the distribution of reports is more compressed than the true distribution. Below I show this point for $\Psi(\mathbf{p}, \mathbf{v}) = \sum_{\omega \in \Omega} p_\omega(1-p_\omega)$, but the intuition holds more generally. Let $V \sim \mathcal{U}_{[0,500]}$ be the random compensation. There are two states of the world: ω_1 = "V is less than $\tilde{v}$" (the DM gets a trip worth v to them) and ω_2 = "V is greater than or equal to $\tilde{v}$" (the DM gets £V). The DM solves

$$\max_{\tilde{v} \in [0,500]} \mathsf{P}\{V < \tilde{v}\}v + [1 - \mathsf{P}\{V < \tilde{v}\}]E[V|V \geq \tilde{v}] + 2\alpha \mathsf{P}\{V < \tilde{v}\}[1 - \mathsf{P}\{V < \tilde{v}\}]$$

$$\iff \max_{\tilde{v} \in [0,500]} \frac{\tilde{v}}{500}v + \left(1 - \frac{\tilde{v}}{500}\right) \frac{\int_{\tilde{v}}^{500} \frac{V}{500} dV}{\left(1 - \frac{\tilde{v}}{500}\right)} + 2\alpha \frac{\tilde{v}}{500}\left(1 - \frac{\tilde{v}}{500}\right)$$

The FOC is:

$$\frac{v}{500} - \frac{\tilde{v}}{500} + \frac{2\alpha}{500} - \frac{4\alpha\tilde{v}}{500^2} = 0$$

$$\boxed{\tilde{v} = p(\alpha)v + [1 - p(\alpha)]250 \ \text{ where } \ p(\alpha) := \frac{1}{1+\alpha/125}}$$

It follows immediately that $\tilde{v} > v \iff v < 250$. In other words, the bias in reporting is positive (negative) if the true valuation v is lower (higher) than the midpoint of the support of the distribution of V. Furthermore, $\lim_{\alpha \to 0} \tilde{v} = v$ and $\lim_{\alpha \to \infty} \tilde{v} = 250$, which is intuitive since reporting £250 gives a 50/50 chance of receiving either the compensation or the trip and thus maximizes the uncertainty about the prize.

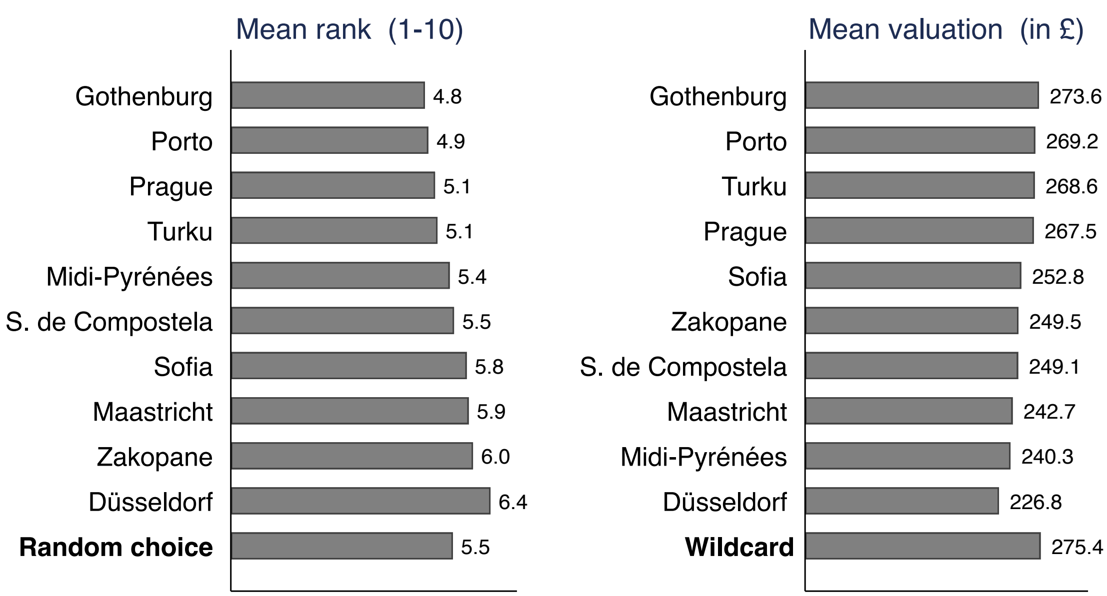
In addition, for most measures of uncertainty considered in this paper, $\tilde{v}$ is a strictly increasing function of v. In other words, the ranking implied by the distribution of reports coincides with the ranking that uses the true valuations, and should also coincide with the ordinal ranking procedure.[37]

[37] Assuming $\alpha > 0$, it can be shown that a sufficient condition for $\frac{\partial \tilde{v}}{\partial v} > 0$ is (i) $\frac{\partial^2 \Psi}{\partial \tilde{v} \partial v} \geq 0$ and (ii) $\frac{\partial^2 \Psi}{\partial \tilde{v}^2} \leq 0$. Part (ii) is satisfied for all valid measures of uncertainty (concavity in $\mathbf{p}$), while (i) is satisfied for all measures of uncertainty that do not depend on the distribution of valuations (e.g., entropy, residual variance, support); however, this condition is violated for measures such as the variance in valuations.

Practical implications and empirical tests If respondents indeed used the elicitation procedure as a randomization device, the findings of this paper will understate the importance of preferences for randomization. Given the compression in the reports induced by the BDM mechanism, the observed SD violations will appear less dramatic than they truly are (leading to underestimating $|\alpha|$). I also performed an empirical test: if respondents with a preference for randomization used the BDM mechanism as a randomization device, the distribution of reported valuations should be more compressed for respondents with a higher propensity to violate stochastic dominance in favor of randomization i.e., there should be a negative relationship between the prevalence of SD violations on $\mathcal{D}^+$ or $\mathcal{B}^+$ and the standard deviation of the reported valuations, $\text{std}(\{\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_{10}\})$. Instead, I find that the correlation is insignificant and, if anything, positive. There is also no relationship between preference for randomization in the design task and compression in the reported valuations.

B.2 Data on ranking and valuation of the destinations

Figure B1: Average rank and valuation across respondents



Notes: For the left panel, the random choice benchmark corresponds to the mean rank assuming the respondent randomly selects a rank from the set $\{1, 2, ..., 10\}$ (each with equal chance). N = 83.

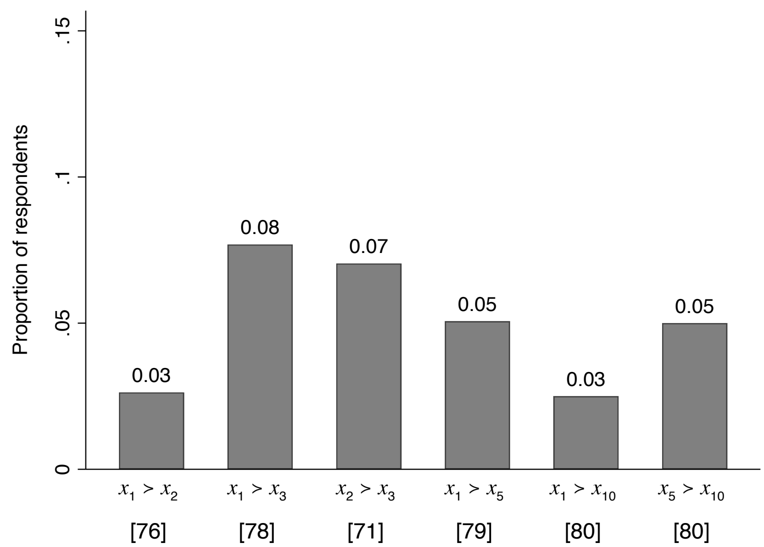
Table B1: Valuations by rank number

	Mean	SD	p25	p50	p75	Min	Max	N
v_1	373.2	88.6	320	400	450	80	500	83
v_2	335.5	93.9	300	350	400	10	480	83
v_3	310.6	90.3	260	300	380	50	490	83
v_4	278.1	100.2	200	290	350	8	454	83
v_5	260.8	96.4	190	270	340	7	480	83
v_6	242.1	98.9	170	240	320	6	498	83
v_7	216.8	100.7	150	200	300	4	485	83
v_8	193.6	104.7	120	180	250	3	500	83
v_9	174.1	104.5	100	150	240	0	499	83
v_{10}	155.2	105.0	80	135	220	0	475	83
v^*	275.4	134.4	180	295	400	10	500	83

Notes: Summary statistics about the distribution of valuations entered by respondents for the holiday trip they ranked in the k^{th} position in their ranking, where $k \in \{1, 2, ..., 10\}$; v^* is the valuation for the wildcard trip. Valuations expressed in £. Market value of each trip equal to £420.

B.3 Preference consistency and prevalence of SD violations

Figure B2: Proportion choosing x_k from $\{x_j, x_k\}$ when $x_j \succ x_k$



Notes: The number in square brackets under each bar is the number of respondents with a strict preference $x_j \succ x_k$ (i.e., x_j ranked above x_k and $v_j > v_k$).

Table B2: Prevalence of SD violations based on $\succeq^*$ vs. the ranking only

		Mean	SD	p25	p50	p75	Min	Max
Number of SD violations on								
$\mathcal{D}^+$:	using $\succeq^*$	4.4	4.9	0	3	7	0	21
	using rank	5.0	6.1	0	3	8	0	23
$\mathcal{B}^+$:	using $\succeq^*$	2.3	2.8	0	1	4	0	9
	using rank	3.3	3.5	0	2	6	0	11
$\mathcal{B}^-$:	using $\succeq^*$	1.6	2.8	0	0	3	0	11
	using rank	2.1	3.1	0	0	3	0	11
Fraction of SD violations on								
$\mathcal{D}^+$:	using $\succeq^*$	0.22	0.23	0	0.17	0.38	0	1
	using rank	0.27	0.28	0	0.17	0.46	0	1
$\mathcal{B}^+$:	using $\succeq^*$	0.25	0.28	0	0.18	0.45	0	1
	using rank	0.30	0.32	0	0.18	0.55	0	1
$\mathcal{B}^-$:	using $\succeq^*$	0.16	0.26	0	0	0.27	0	1
	using rank	0.19	0.28	0	0	0.27	0	1

Notes: Summary statistics of the distribution of SD violations at the individual level on $\mathcal{D}^+$ (all 24 dominated lotteries), $\mathcal{B}^+$ (subset of 11 dominated lotteries), $\mathcal{B}^-$ (11 dominant lotteries). N = 83.

Figure B3: Proportion of SD violations depending on size of the support

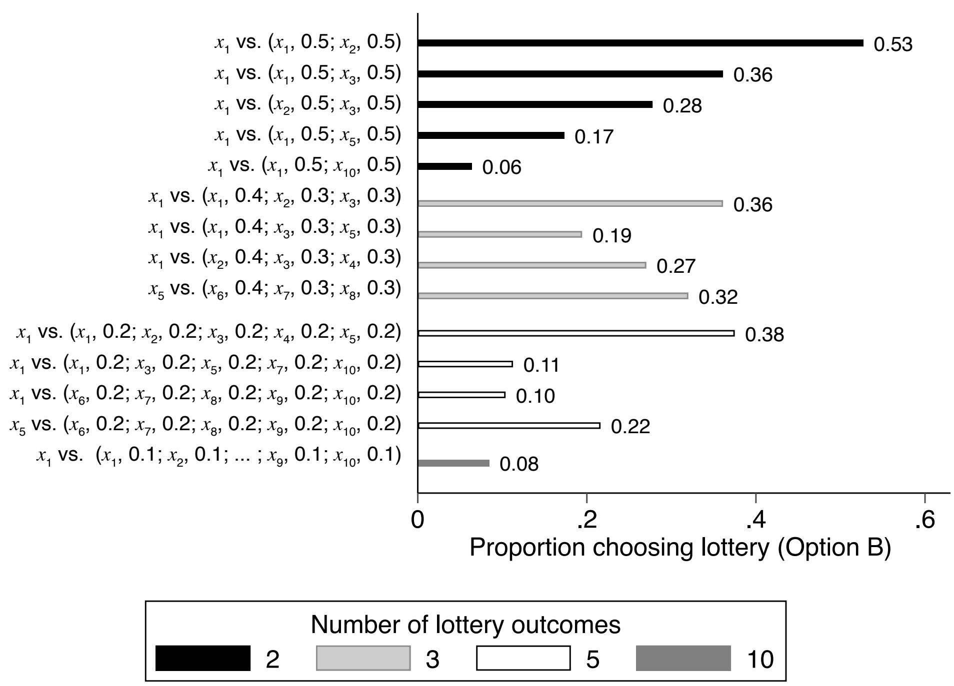


Table B3: Breakdown of SD violations

	n*	N*	n*/N*	n^R	N^R	n^R/N^R
SD violations on $\mathcal{D}^+$ (all dominated lotteries)						
x_1 vs. $(x_1, 0.5; x_2, 0.5)$	39	74	0.53	47	83	0.57
x_1 vs. $(x_1, 0.5; x_3, 0.5)$	26	72	0.36	36	83	0.43
x_1 vs. $(x_2, 0.5; x_3, 0.5)$	20	72	0.28	29	83	0.35
x_1 vs. $(x_1, 0.5; x_5, 0.5)$	13	75	0.17	20	83	0.24
x_1 vs. $(x_1, 0.5; x_{10}, 0.5)$	5	78	0.06	9	83	0.11
x_1 vs. $(x_1, 0.9; x_2, 0.1)$	26	74	0.35	30	83	0.36
x_1 vs. $(x_1, 0.9; x_3, 0.1)$	21	72	0.29	28	83	0.34
x_1 vs. $(x_2, 0.9; x_3, 0.1)$	14	72	0.19	22	83	0.27
x_1 vs. $(x_1, 0.9; x_5, 0.1)$	11	75	0.15	17	83	0.20
x_1 vs. $(x_1, 0.9; x_{10}, 0.1)$	6	78	0.08	8	83	0.10
x_1 vs. $(x_1, 0.1; x_2, 0.9)$	13	74	0.18	19	83	0.23
x_1 vs. $(x_1, 0.1; x_3, 0.9)$	8	72	0.11	16	83	0.19
x_1 vs. $(x_2, 0.1; x_3, 0.9)$	9	72	0.13	17	83	0.20
x_1 vs. $(x_1, 0.1; x_5, 0.9)$	7	75	0.09	14	83	0.17
x_1 vs. $(x_1, 0.1; x_{10}, 0.9)$	2	78	0.03	6	83	0.07
x_1 vs. $(x_1, 0.4; x_2, 0.4; x_3, 0.3)$	26	72	0.36	35	83	0.42
x_1 vs. $(x_1, 0.4; x_3, 0.4; x_5, 0.3)$	14	72	0.19	22	83	0.27
x_1 vs. $(x_2, 0.4; x_3, 0.4; x_4, 0.3)$	20	74	0.27	26	83	0.31
x_5 vs. $(x_6, 0.4; x_7, 0.4; x_8, 0.3)$	24	75	0.32	30	83	0.36
x_1 vs. $(x_1, 0.2; x_2, 0.2; x_3, 0.2; x_4, 0.2; x_5, 0.2)$	27	72	0.38	35	83	0.42
x_1 vs. $(x_1, 0.2; x_3, 0.2; x_5, 0.2; x_7, 0.2; x_{10}, 0.2)$	8	71	0.11	15	83	0.18
x_1 vs. $(x_6, 0.2; x_7, 0.2; x_8, 0.2; x_9, 0.2; x_{10}, 0.2)$	8	77	0.10	12	83	0.14
x_5 vs. $(x_6, 0.2; x_7, 0.2; x_8, 0.2; x_9, 0.2; x_{10}, 0.2)$	16	74	0.22	21	83	0.25
x_1 vs. $(x_1, 0.1; x_2, 0.1; ...; x_9, 0.1; x_{10}, 0.1)$	6	71	0.08	13	83	0.16
SD violations on $\mathcal{B}^-$ (dominant lotteries)						
x_2 vs. $(x_1, 0.5; x_2, 0.5)$	15	74	0.20	17	83	0.20
x_3 vs. $(x_1, 0.5; x_3, 0.5)$	10	72	0.14	14	83	0.17
x_3 vs. $(x_1, 0.5; x_2, 0.5)$	10	69	0.14	13	83	0.16
x_5 vs. $(x_1, 0.5; x_5, 0.5)$	6	75	0.08	10	83	0.12
x_{10} vs. $(x_1, 0.5; x_{10}, 0.5)$	3	78	0.04	6	83	0.07
x_2 vs. $(x_1, 0.9; x_2, 0.1)$	18	74	0.24	21	83	0.25
x_3 vs. $(x_1, 0.9; x_3, 0.1)$	10	72	0.14	15	83	0.18
x_3 vs. $(x_1, 0.9; x_2, 0.1)$	7	69	0.10	11	83	0.13
x_2 vs. $(x_1, 0.1; x_2, 0.9)$	23	74	0.31	25	83	0.30
x_3 vs. $(x_1, 0.1; x_3, 0.9)$	22	72	0.31	25	83	0.30
x_3 vs. $(x_1, 0.1; x_2, 0.9)$	12	69	0.17	17	83	0.20

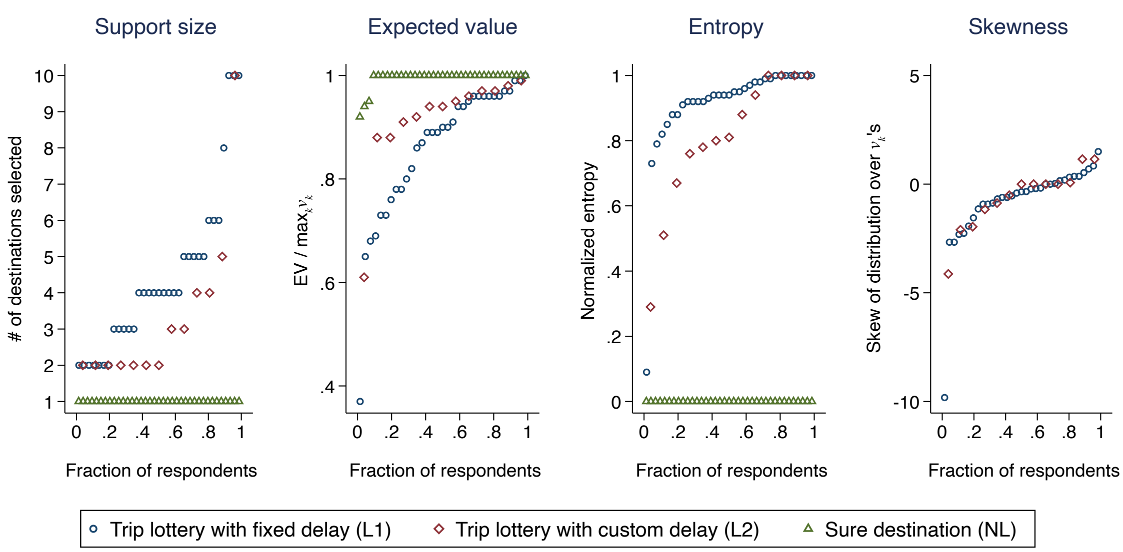
Notes: For each DP d, n*/N* denotes the fraction of SD violations according to the unambiguous relation $\succeq^*$ (see Sections 4.2 and 5 for definitions), where N* is the total number of respondents for whom d can be classified as a clear dominance problem and n* is the number of respondents who violated stochastic dominance among them; n^R/N^R is the fraction of SD violations according to the ranking i.e., n^R is the number of respondents who chose Option B (A) when $d \in \mathcal{D}^+$ ($d \in \mathcal{B}^-$).

C Decisions in the design task

C.1 Characteristics of the favorite option

Chosen destination(s) and probability weights In Part 3 of the survey, respondents were offered to design their own lottery or take a sure destination instead (see Subsection 3.2.3). Figure C1 presents the characteristics of the option chosen by each respondent. In total, 55% designed a lottery: 40% (33/83) chose a lottery even with a fixed delay (group L1) and 15% (13/83) only if the delay could be customized (group L2). Generally speaking, L1 respondents selected lotteries with more uncertainty than L2 respondents and were willing to sacrifice more in terms of expected value. Reassuringly, over 90% of those who preferred a sure destination selected the option they assigned the highest valuation to.[38] Among those who built a lottery, 93% (43/46) included their rank #1 destination in the support of their lottery. The last panel shows that the chosen lotteries tended to exhibit a negative skew.

Figure C1: Characteristics of respondents' favorite option



Notes: Quantile plots of (i) the size of the support of $\mathbf{p}$ chosen by respondents; (ii) the expected value of the chosen option relative to the highest valuation, $\mathbf{p} \cdot \mathbf{v} / \max_k v_k$; (iii) the entropy for the chosen probability $\mathbf{p}$ (normalized by $\ln(|\text{supp}(\mathbf{p})|)$ so the measure lies between 0 and 1); (iv) the skewness (3rd moment) of the distribution of valuations (lottery decisions only). N = 83.

[38]In total, 84% (31/37) of the respondents who preferred a sure destination selected the one they ranked #1. In 3 of the 6 cases where they did not, they in fact selected the option they valued the most (meaning that the ranking and valuations disagreed).

C.2 Preferences for delaying the resolution of uncertainty

Chosen delays and WTP for delay Among respondents who built a lottery (N = 46), Figure C2 shows the distribution of chosen dates for when to reveal the destination. Figure C3 shows the distribution of WTP for delay among these 46 respondents (left panel) and for the subset of 34 respondents who saw the MPL question (right panel). There is a positive relationship between amount of delay chosen and willingness to pay for delay (Spearman $\rho = 0.42$, p = 0.013, N = 34).

Figure C2: Distribution of chosen delays

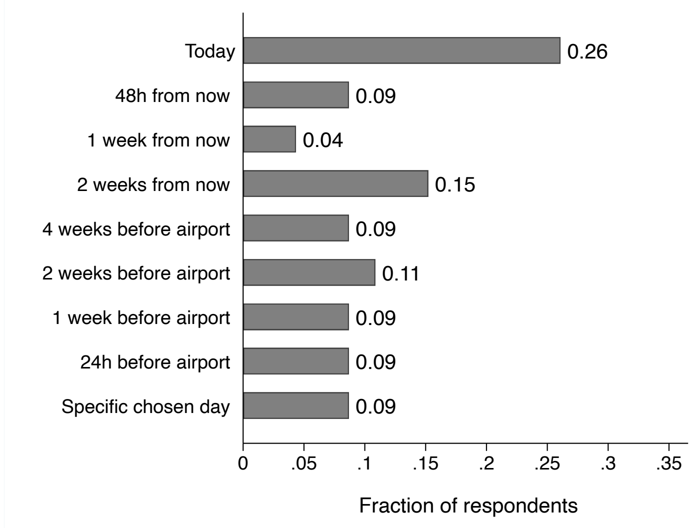
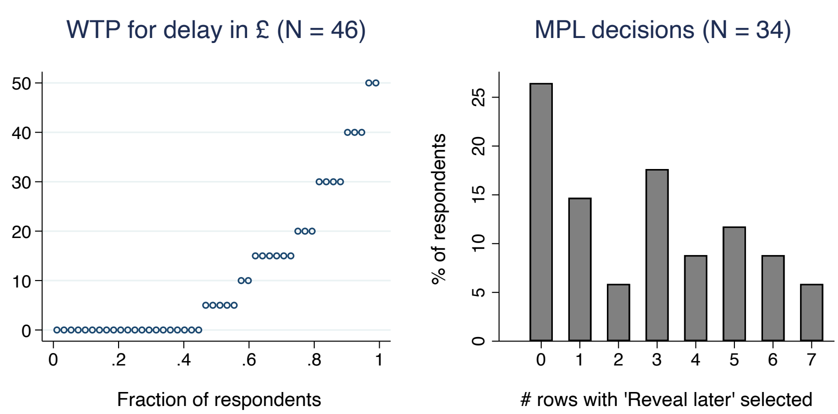


Figure C3: Distribution of WTP for delay



Notes: The left panel is a quantile plot (using uniform scaling) of `WTP for delay (in £)`; it is set to 0 for the 12 respondents who preferred to learn right away. The right panel is a histogram of the number of rows in the MPL at which the respondent preferred "*Revealing Later*" to "*Revealing Now*" + £X, where X $\in \{5, 10, 15, 20, 30, 40, 50\}$ (34 respondents who preferred a delay only).

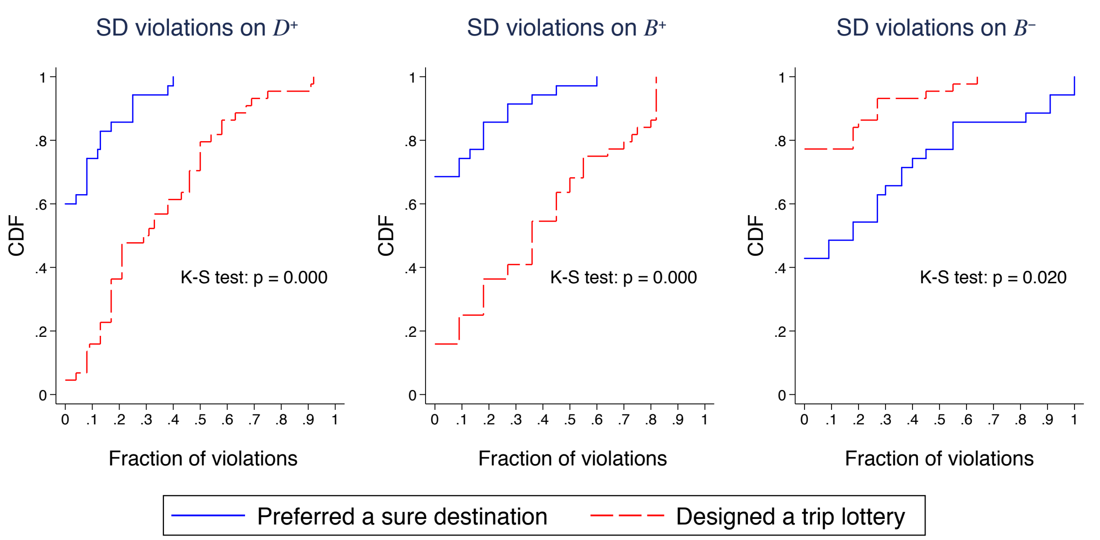
C.3 Preferences in the binary choice exercise vs. design task

Table C1: Prevalence of SD violations depending on favorite option

	Sure destination	Trip lottery	Diff.	s.e.	N
SD violations on $\mathcal{D}^+$					
Number of violations	1.4	6.9	-5.4***	(0.9)	83
Fraction of violations	0.07	0.34	-0.27***	(0.04)	79
SD violations on $\mathcal{B}^+$					
Number of violations	0.7	3.6	-2.9***	(0.5)	83
Fraction of violations	0.08	0.39	-0.31***	(0.05)	79
SD violations on $\mathcal{B}^-$					
Number of violations	2.8	0.7	2.0***	(0.6)	83
Fraction of violations	0.27	0.07	0.20***	(0.06)	79

Notes: *** $p < 0.01$. `Sure destination` refers to respondents who preferred not to build a trip lottery, while `Trip lottery` pools respondents who chose to build a lottery (before or after being offered to customize the delay). For the number (fraction) of violations, N = 37 (35) for the `Sure destination` group and N = 46 (44) for the `Trip lottery` group.

Figure C4: Relationship between binary choices and lottery-building exercise



Notes: p-values from Kolmogorov-Smirnov tests of equality of distributions. N = 37 for the respondent group “Preferred a sure destination” and N = 46 for the “Designed a trip lottery” group.

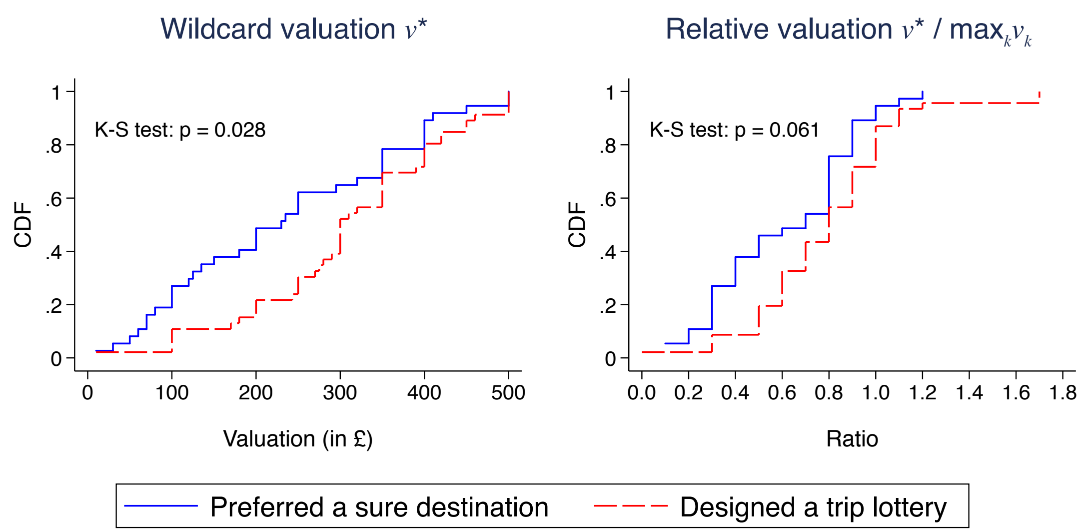
D Drivers of preferences for surprise trips

This section presents information on the psychological drivers of respondents' preferences for surprise trips (risky lotteries, wildcard trip, and link between the two).

D.1 Preferences for the wildcard trip

Link with risky trip lotteries Respondents who preferred to design a trip lottery typically placed a higher value on the wildcard trip (Figure D1). The average difference in wildcard valuations between the two groups is £74.1 (95% CI: [17.0, 131.2], p = 0.012).[39] The association between wildcard valuation and other measures of preference for randomization and delay is however generally weak and insignificant (highest correlation around 0.2, with the fraction of SD violations on $\mathcal{D}^+$).

Figure D1: Relationship between wildcard valuation and design choice

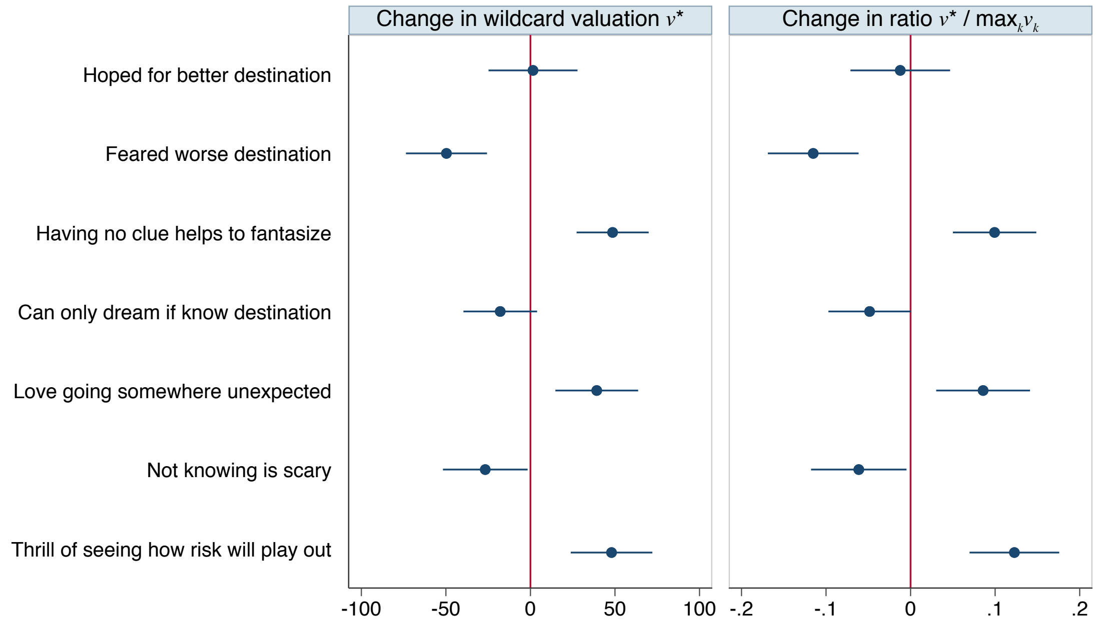


Notes: Distribution of wildcard valuations by whether the respondent designed a trip lottery as their favorite option (N = 46) or preferred a sure destination (N = 37); the left panel shows valuations in pounds and the right panel as a ratio relative to the highest valuation $\max_k v_k$ assigned to the 10 known destinations. The p-values are from Kolmogorov tests of equality of distributions. N = 83.

Motives After entering their valuations, respondents were asked to rate their agreement with various potential reasons for their preferences. Figure D2 shows that hoping

[39] The average difference is £64.6 ($p = 0.043$, 95% CI: [2.1, 127.2]) if one restricts attention to the set of respondents who had a clear preference for x_1 relative to all the other options (i.e., for whom $x_1 \succ^* x_j$ for all $j \in \{2, 3, ..., 10\}$).

Figure D2: Determinants of wildcard valuation



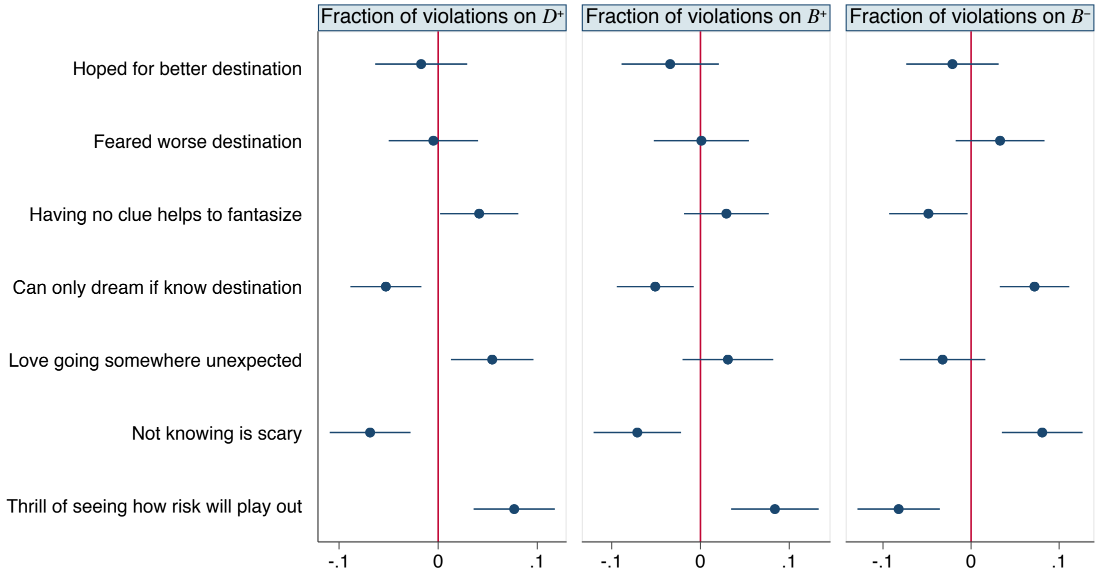
Notes: **Left panel**: Two-limit Tobit regressions of wildcard trip valuation v^* (in £) on each of 7 ratings (where 1 = "completely disagree" and 5 = "completely agree") entered separately; **right panel**: univariate linear regression of the ratio $v^* / \max_k v_k$ (where $\max_k v_k$ is the highest trip valuation that the respondent entered for the 10 destinations) on each rating separately. N = 83.

for a better destination has no predictive power, but fear of a worse destination does: a one-point increase in agreement with the statement "*I feared I would get a trip I like less than the other destinations*" is associated with a £49.6 decrease (95% CI: [-73.6, -25.7]) in valuations on average. In terms of positive factors, the ability to fantasize about different worlds and the thrill of seeing how risk will play out appear to be the strongest predictors, with a £48.6 increase (95% CI: [27.3, 69.9]) on average for the former, and a £48.0 increase (95% CI: [23.8, 72.1]) for the latter.[40]

Link with SD violations I also examined whether the motives underlying valuations of the wildcard trip underpin randomization behavior for risky lotteries. Predictors are very similar overall as shown in Figure D3.

[40]The various ratings tend to be highly correlated, with most correlations in the range [0.4, 0.6], at the exception of "Hoped for better destination," which is not correlated with any of the other factors. Putting all these factors in a multivariate regression produces some counterintuitive findings (positive and significant coefficient on "Can only dream if know destination") and, at the exception of "Feared worse destination," the estimated effects for the other reasons are non-robust and sometimes depend on the choice of outcome variable (v^* or $v^* / \max_k v_k$).

Figure D3: Link between motives for the wildcard and SD violations



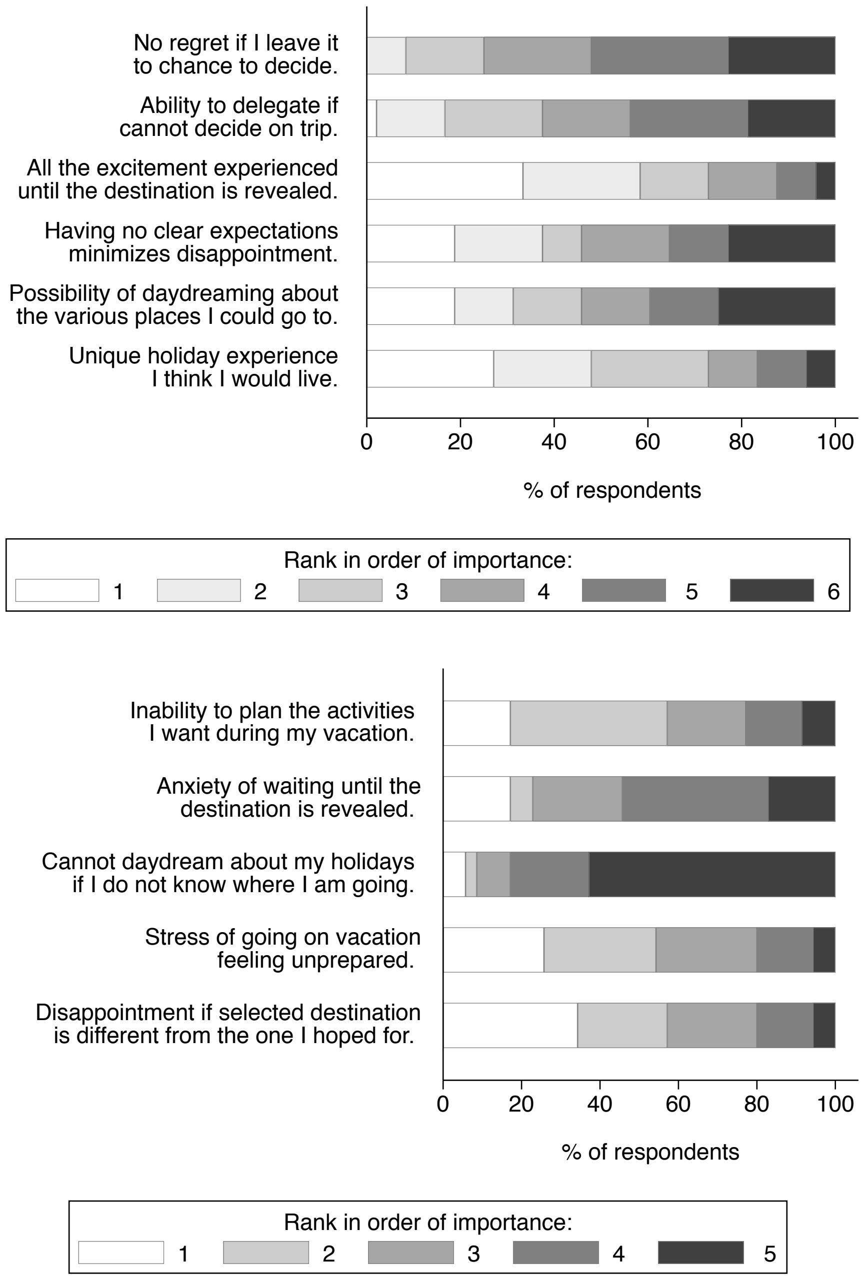
Notes: Coefficients from univariate linear regressions of the fraction of SD violations on $\mathcal{D}^+$, $\mathcal{B}^+$ and $\mathcal{B}^-$ on each of 7 ratings (where 1 = "completely disagree" and 5 = "completely agree"). N = 79.

D.2 Attitudes towards surprise trips

Rating of surprise trips At the end of the survey, respondents were asked to rate on a scale from 0 to 100 how much overall they liked the concept of "surprise trip" presented in the study. In line with the choice data, views tended to be positive but with large heterogeneity: the mean (median) rating was 53.3 (60), with a standard deviation of 30.6. As expected, this rating correlates very well with the various measures of preference for randomization and delay considered in this paper.

Motives for the rating Respondents who gave a rating of at least 50 (less than 50) to the concept of surprise trip were asked to rank a set of reasons for why they liked (disliked) the concept. Figure D4 shows the distribution of answers.

Figure D4: Motives for liking (top) vs. disliking (bottom) surprise trip lotteries



Notes: **Top panel**: Distribution of answers to the question: "*What do you like the most in the idea of a surprise lottery over holiday trips? Please rank the following aspects in order of importance [...].*"; only asked to the 48 respondents who gave a rating $\geq$ 50 (out of 100) to the concept of "surprise trip". **Bottom panel**: Distribution of answers to the question: "*What do you find unappealing in the idea of a surprise lottery over holiday trips? Please rank the following aspects in order of importance, starting with the aspect you find the most unappealing [...].*"; only asked to the 35 respondents who gave a rating $<$ 50 (out of 100) to the concept of "surprise trip".

E Stochastic choice benchmarks

Below I summarize simulation exercises conducted to understand the implications of 4 stochastic choice models for (i) the prevalence of SD violations; (ii) the correlation between violations in favor of randomization vs. certainty; (iii) the shape of violations (link to P-MON).

General approach To account for sampling error, I draw 1,000 simulations for each benchmark, which I compare to the actual distribution. For all benchmarks, the simulations take as given respondents' preferences over (sure) destinations, thus only perturbing their choices when comparing a trip lottery to a sure destination. Doing so allows me to compute the total number and fraction of SD violations at the individual level in exactly the same way for the actual and simulated choices i.e., using the unambiguous relation $\succeq^*$. For each respondent, I then generate counterfactual indices for the number and fraction of SD violations on $\mathcal{D}^+$ (all 24 dominated lotteries), $\mathcal{B}^+$ (symmetric subset of 11 dominated lotteries), and $\mathcal{B}^-$ (11 dominant lotteries).

E.1 Benchmark 1: Uniform probability of mistake

Assumptions and implementation This simple benchmark assumes that respondents have standard preferences (satisfying stochastic dominance), but make implementation mistakes with a 10% chance when choosing between Options A and B. The preference flips were simulated by taking iid draws (across respondents and DPs) from a Bernoulli distribution $X \sim \mathcal{B}(p)$ with $p = 0.1$, where $X = 1$ corresponds to a flip. The value of $p = 0.1$ was picked to be an upper bound on the observed proportion of inconsistencies in the DPs involving two sure options, which is 0.07 on average when the ranking and valuations agree (see Figure B2 for a breakdown).

Main findings The null of equality of distributions for the actual vs. each of the simulated indices is comfortably rejected for SD violations on $\mathcal{D}^+$ and $\mathcal{B}^+$, while the evidence is more mixed for violations on $\mathcal{B}^-$. Of note, the simulation fails to reproduce the long tail of high-frequency violations that appears in the actual data.[41] Looking

[41] For robustness, I also examine the distribution of individual-level indices assuming a probability $p = 0.2$ of preference flip, which approximately matches the average prevalence rate of SD violations observed for dominated lotteries. The distribution of simulated indices remains too compressed to match the empirical data, with clear statistical differences for both classes of dominance problems.

at the individual-level relationship between SD violations on $\mathcal{B}^+$ and $\mathcal{B}^-$ in the 1,000 simulated datasets, the median correlation is positive and equal to 0.13 (0.02) for the total number (fraction) of SD violations, with only 10% (39%) of correlation coefficients being negative. The null hypothesis of zero correlation on average ($\mu_\rho = 0$) is easily rejected in both cases in favor of $\mu_\rho > 0$, contradicting the actual data.

E.2 Benchmark 2: Heterogeneous probability of mistake

Assumptions and implementation Instead of imposing a uniform probability of mistake across respondents, this benchmark assumes that each respondent i makes a mistake with probability p_i drawn from some distribution F with mean $\mu_1 = 0.1$. The distribution F was chosen to approximately match the empirical distribution $\hat{F}$ of inconsistency rates on comparisons between two sure options. As this distribution is right-skewed ($\mu_3 = 2.9$), with most respondents exhibiting no inconsistency, I assume $p_i \sim \text{Beta}(\alpha, \beta)$ with $\alpha = 0.1$ and $\beta = 0.9$ (yielding $\mu_1 = 0.1$ and $\mu_3 = 2.5$).

Main findings This benchmark accommodates more heterogeneity in the distribution of indices and does a fairly good job at approximating the data on $\mathcal{B}^-$ (with an underestimation at the 75$^{\text{th}}$ percentile). However, major differences remain between the actual and simulated datasets for violations on $\mathcal{D}^+$ and $\mathcal{B}^+$. In addition, allowing for heterogeneity in the probability of mistake generates strong positive correlations between violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ i.e., above 0.5 in all 1,000 simulations, with a median of 0.82 (0.81) for the number (fraction) of SD violations.[42]

E.3 Benchmark 3: Random valuations model

A plausible conjecture is that valuations are measured with error because of the difficulty of the exercise. Benchmark 3 allows for stochastic valuation errors by assuming that respondent i decides in each DP l as if their valuation for destination k was given by $\tilde{v}_k^i(l) = v_k^i + \epsilon_k^i(l)$, where $\epsilon_k^i(l) \sim \mathcal{N}(0, \sigma^2)$ is a shock drawn iid across respondents and DPs (and v_k^i is respondent i's reported valuation for destination k). I conduct simulations with $\sigma = 20$ and truncate valuations at 0 and 500 so that:

[42]For completeness, I also re-ran the simulations for $p_i \sim \text{Beta}(0.2, 0.8)$, yielding $\mu_1 = 0.2$ and $\mu_3 = 1.4$. In this case, the exercise also does a good job at matching the data for problems in $\mathcal{B}^+$ (and $\mathcal{D}^+$ to a lower extent), while fit remains very good on $\mathcal{B}^-$. However, the positive correlation between SD violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ remains high (if not stronger) in all simulations.

$$\tilde{v}_k^i(l) = \begin{cases} 0 & \text{if} \quad v_k^i + \epsilon_k^i(l) \leq 0 \\ v_k^i + \epsilon_k^i(l) & \text{if} \quad v_k^i + \epsilon_k^i(l) \in (0, 500) \\ 500 & \text{if} \quad v_k^i + \epsilon_k^i(l) \geq 500 \end{cases}$$

Setting $\sigma = 20$ would generate a proportion of preference flips of $\approx 10\%$ in DPs involving two sure destinations (i.e., $\mathbb{P}\{\tilde{v}_j^i(l) < \tilde{v}_k^i(l)\} \approx 0.108$ across all 1,000 simulations), in line with the assumption made in Benchmarks 1 & 2.[43]

Main findings This benchmark greatly underestimates the prevalence rate of SD violations both for dominated and dominant lotteries, although differences between the actual and simulated distributions are only significant for $\mathcal{D}^+$ and $\mathcal{B}^+$. The correlation between violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ is positive in all samples, with a median of 0.46 (0.49) for the number (fraction) of SD violations. Importantly, according to this benchmark, the probability of choosing the dominated option should be (i) independent of p and (ii) lower for destinations that are further apart in the DM's ranking (i.e., for which the distance in valuations is larger). To see this, note that respondent i chooses $(x_j, p; x_k, 1-p)$ over x_j in DP l if $p[v_j^i + \epsilon_j^i(l)] + (1-p)[v_k^i + \epsilon_k^i(l)] > v_j^i + \epsilon_j^i(l)$ (breaking ties in favor of the sure option). Rearranging the expression yields $\epsilon_k^i(l) - \epsilon_j^i(l) > v_j^i - v_k^i$. While (ii) matches the data, (i) does not.[44]

E.4 Benchmark 4: Random utility model

Assumptions and implementation A related but different model assumes that additive shocks occur on utilities instead of valuations, as in standard random utility models. Since the model assumes risk neutrality, this assumption amounts to imposing an additive error on the expected value of a lottery. In this case, respondent i will choose Option B in DP l if $E_{\mathbf{p}}[v^i(x_B)] + \epsilon_B^i(l) > v^i(x_A) + \epsilon_A^i(l)$ where ϵ_A^i and ϵ_B^i

[43] Note that the model predicts more reversals (choosing x_k from $\{x_j, x_k\}$ despite $x_j \succ x_k$) for two trips that are closer in the DM's preferences. The data provides limited support for this prediction. First, while the difference $\Delta v := v_j - v_k$ is on average smaller for respondents who chose x_k from $\{x_j, x_k\}$ instead of x_j ($\Delta v =$ £37 vs. £98, $p = 0.001$), this difference is mostly driven by the small set of respondents for whom $v_j - v_k \leq 0$. Second, the probability of reversal between x_j and x_k does not increase as the distance in ranks $|j - k|$ goes down (see Figure B2).

[44] These observations do not rely on shocks being additive e.g., assuming multiplicative shocks to valuations, the DM chooses the lottery $(x_j, p; x_k, 1-p)$ over the sure option x_j in DP l provided that $p[\epsilon_j^i(l)v_j^i] + (1-p)[\epsilon_k^i(l)v_k^i] > \epsilon_j^i(l)v_j^i \;\Rightarrow\; \frac{\epsilon_k^i(l)}{\epsilon_j^i(l)} > \frac{v_j^i}{v_k^i}$ (independent of p).

are taste shocks drawn iid across respondents and DPs. To maximize comparability with Benchmark 3, I again draw iid shocks $\epsilon^i_A(l)$, $\epsilon^i_B(l) \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 20$, equivalent to $\approx 10\%$ of preference flips in DPs involving two sure destinations (i.e., $\mathbb{P}\{v^i(x_B) + \epsilon^i_B(l) > v^i(x_A) + \epsilon^i_A(l)\} \approx 0.111$ across all 1,000 simulations).

Main findings Benchmark 4 better matches certain moments of the data than Benchmark 3, especially for violations on $\mathcal{B}^-$ (mean and 75$^{\text{th}}$ percentile), and accommodates more heterogeneity. However, it generally underestimates the prevalence of SD violations, with statistically significant differences between the simulated and actual distributions detected in most cases (except on $\mathcal{B}^+$). In addition, the correlation between SD violations on $\mathcal{B}^+$ vs. $\mathcal{B}^-$ is again positive in (virtually) all samples, with a median correlation coefficient of 0.32 (0.44) for the number (fraction) of SD violations. Finally, this benchmark would generate a higher proportion of SD violations as the utility difference between the two options goes down (e.g., as p increases on $\mathcal{B}^+$). Indeed, respondent i chooses $(x_j, p; x_k, 1-p)$ over x_j provided that $pv^i_j + (1-p)v^i_k + \epsilon^i_B(l) > v^i_j + \epsilon^i_A(l)$, that is, $\epsilon^i_B(l) - \epsilon^i_A(l) > (1-p)(v^i_j - v^i_k)$. Thus, P-MON should not be violated in the aggregate i.e., the fraction of SD violations should not increase when p decreases from 0.9 to 0.5, which contradicts the data.[45]

F Theoretical appendix

F.1 Discussion of Observation 2

In Section 2.2 of the main text, I discuss the fact that if H is a valid measure of uncertainty and ψ is such that $\psi(0) = 0$, $\psi' > 0$, and $\psi'' \leq 0$, then $\succeq$ cannot violate P-MON on $\mathcal{B}^+_\sim$ (Observation 2). I present a proof of Observation 2 below and explain how P-MON may be violated (i) on $\mathcal{B}^+_\sim$ by convexifying ψ; (ii) on $\mathcal{B}^+_\succ$ for any monotone ψ; (iii) on $\mathcal{D}^+_\sim$ for any monotone ψ if $|\text{supp}(\mathbf{p})| \geq 3$.

Proof of Observation 2. If ψ is globally concave, then $\psi \circ H$ is globally concave, as ψ is strictly increasing and H is globally concave (by definition of a valid measure).

[45] At the individual level, violations of P-MON may still occasionally occur in such a model given that shocks are drawn iid across DPs. For instance, taking DP $l \in \{2, 18\}$ (see Table 3), one could simultaneously have $0.5v^i_1 + 0.5v^i_2 + \epsilon^i_B(2) > v^i_1 + \epsilon^i_A(2)$ and $v^i_1 + \epsilon^i_A(18) > 0.9v^i_1 + 0.1v^i_2 + \epsilon^i_B(18)$.

The DM strictly prefers x_j to the lottery $(x_j, p; x_k, 1-p)$ where $x_j \succ x_k$ if

$$v_j > pv_j + (1-p)v_k + \alpha\Psi(p) \iff \frac{\Delta v}{\alpha} > f(p) := \frac{\Psi(p)}{1-p}$$

where $\Delta v := v_j - v_k$. It can be easily seen that

$$f'(p) = \frac{\Psi'(p)(1-p) + \Psi(p)}{(1-p)^2} \geq 0 \quad \forall\, p \in (0,1)$$

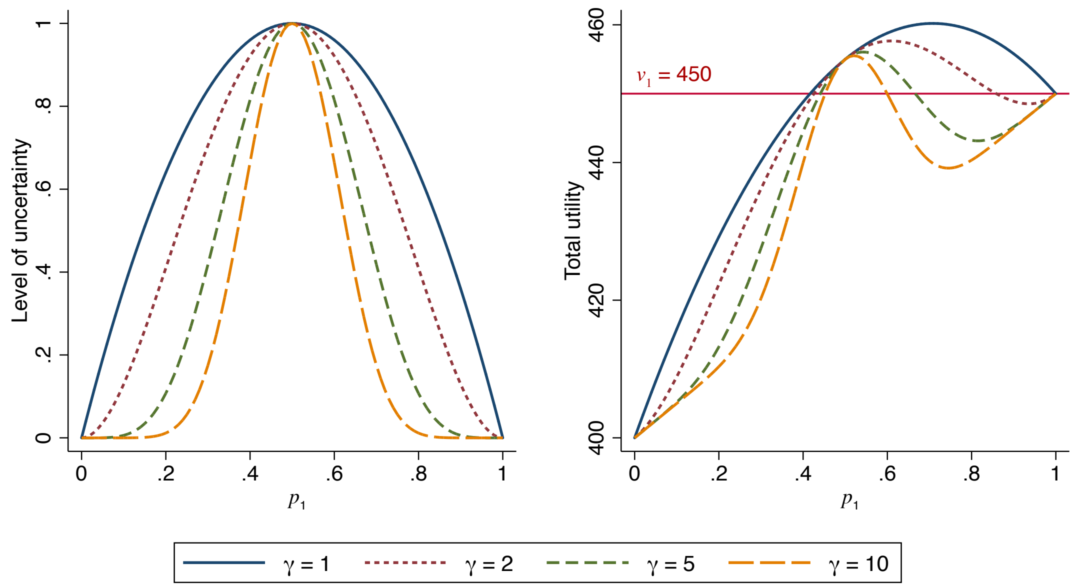
This follows from the concavity of Ψ, which implies $\Psi(1) \leq \Psi(p) + \Psi'(p)(1-p)$, together with the fact that $\Psi(1) = 0$. In other words, if ψ is globally (weakly) concave, the DM will switch at most once from x_j to $(x_j, p; x_k, 1-p)$ as p increases i.e., P-MON cannot be violated. Q.E.D.

P-MON violations on $\mathcal{B}^+_\sim$ by convexifying ψ Using the residual variance as an example, Figure F1 shows how the shape of uncertainty $\Psi = \psi \circ H$ and utility $U_{\alpha,\Psi}$ change for different levels of convexity of ψ (i.e., $\psi(H) = H^\gamma$ with $\gamma \in \{1, 2, 5, 10\}$). To facilitate comparisons, I normalize H to be between 0 and 1 by dividing it by its maximum value. The transformed measure is hump-shaped, allowing for violations of P-MON. The figure is nearly identical when using the Shannon entropy instead.

P-MON violations on $\mathcal{B}^+_\succ$ If the DM violates stochastic dominance at some $d \in \mathcal{B}^+_\succ$ (strict dominance problem with a binary-outcome lottery), then they must violate P-MON as well. To see this, suppose there exists $p^* \in (0,1)$ such that $(x_j, p^*; x_k, 1-p^*) \succ x_i$ for three prizes x_i, x_j, x_k with $x_i \succ x_j \succ x_k$. By the representation, $U_{\alpha,\Psi}(p^*) = p^*v_j + (1-p^*)v_k + \alpha\Psi(p^*) > v_i$. Note that $\lim_{p\to 1} U_{\alpha,\Psi}(p) = v_j < v_i$. Since $U_{\alpha,\Psi}(p)$ is continuous, there must be some $\bar{p} \in (p^*, 1)$ such that $U_{\alpha,\Psi}(\bar{p}) < v_i$, or equivalently, $x_i \succ (x_j, \bar{p}; x_k, 1-\bar{p})$, implying a violation of P-MON.

P-MON violations when $|\mathbf{supp}(\mathbf{p})| \geq 3$ By a similar logic, P-MON violations can occur on weak dominance problems if $|\text{supp}(\mathbf{p})| \geq 3$. For instance, for any $\mathbf{p}^*$ such that $U_{\alpha,\Psi}(\mathbf{p}^*) = p_i^*v_i + p_j^*v_j + p_k^*v_k + \alpha\Psi(\mathbf{p}^*) > v_i$ (with $v_i > v_j > v_k$), one can find $\epsilon > 0$ and $\mathbf{p}^*_\epsilon = (p_i^*, p_j^* + \epsilon, p_k^* - \epsilon)$ such that $v_i > p_i^*v_i + (p_j^* + \epsilon)v_j + (p_k^* - \epsilon)v_k + \alpha\Psi(\mathbf{p}^*_\epsilon)$,

Figure F1: Shape of uncertainty and SD violations for $\Psi(\mathbf{p}) = \left(\frac{n}{n-1} \sum_{k=1}^{n} p_k(1-p_k) \right)^{\gamma}$



Notes: Plots of the level of uncertainty $\Psi(p_1) = \left(\frac{2p_1(1-p_1)}{1/2}\right)^{\gamma}$ (left panel) and utility $U_{\alpha,\Psi}(p_1, v_1, v_2)$ $= p_1 v_1 + (1-p_1)v_2 + \alpha\Psi(p_1)$ (right panel) for $\alpha = 30$, $(v_1, v_2) = (450, 400)$ and $\gamma \in \{1, 2, 5, 10\}$.

which follows by continuity and the fact that $\lim_{\epsilon \to (1-p_j^*)} U_{\alpha,\Psi}(\mathbf{p}_\epsilon) = v_j$.[46]

F.2 Optimal solution for specific measures of uncertainty

Below I examine the optimal solution to the following maximization problem

$$\max_{\{p_k\}_{k=1}^n} \quad \sum_{k=1}^{n} p_k v_k + \alpha\Psi(\mathbf{p}, \mathbf{v}) \text{ s.t. } (1a): \sum_{k=1}^{n} p_k = 1 \text{ and } (1b): \; p_k \geq 0 \;\; \forall k$$

for 3 measures of uncertainty (assuming $\psi = I$): (i) residual variance $\Psi(\mathbf{p}, \mathbf{v}) = \sum_{k=1}^{n} p_k(1-p_k)$; (ii) Shannon entropy $\Psi(\mathbf{p}, \mathbf{v}) = \sum_{k=1}^{n} p_k \ln(p_k)$; and (iii) variance in valuations $\Psi(\mathbf{p}, \mathbf{v}) = \sum_{k=1}^{n} p_k(v_k - E_{\mathbf{p}}[v])^2$. Define the Lagrangian for this problem:

$$\mathcal{L}(\mathbf{p}) := \sum_{k=1}^{n} p_k v_k + \alpha\Psi(\mathbf{p}, \mathbf{v}) - \lambda(\sum_{k=1}^{n} p_k - 1) + \sum_{k=1}^{n} \mu_k p_k$$

[46]For example, if $\Psi(\mathbf{p}, \mathbf{v}) = \sum_{k=1}^{n} p_k(v_k - E_{\mathbf{p}}[v])^2$, $(v_1, v_2, v_3) = (450, 400, 350)$, and $\alpha = 0.08$, it is easy to verify that $(x_1, 0.2; x_2, 0.5; x_3, 0.3) \succ x_1 \sim (x_1, 0.2; x_2, 0.6; x_3, 0.2) \succ (x_1, 0.2; x_2, 0.7; x_3, 0.1)$.

where λ and $\{\mu_k\}_{k=1}^n$ are the Lagrange multipliers on (1a) and (1b).

Residual variance Differentiating the Lagrangian, the first-order condition for p_k is $v_k + \alpha(1-2p_k) - \lambda + \mu_k = 0$. A solution $\mathbf{p}^*$ with corners exists if $\mu_k > 0$ for some k (implying that $p_k = 0$ by the Kuhn-Tucker complementary slackness conditions). Let $\mathcal{J} := \text{supp}\{\mathbf{p}^*\}$ be the set of all destinations x_j such that $p_j^* > 0$. By Kuhn-Tucker, $\mu_j = 0$ for all $j \in \mathcal{J}$. Combining the FOCs for two $j, k \in \mathcal{J}$ yields:

$$v_j + \alpha(1-2p_j) = v_k + \alpha(1-2p_k) \iff p_j = p_k + \frac{v_j - v_k}{2\alpha}$$

Summing over all $j \in \mathcal{J}$ and using the fact that $\sum_{j\in\mathcal{J}} p_j = 1$:

$$\sum_{j\in\mathcal{J}}\left[p_k + \frac{v_j - v_k}{2\alpha}\right] = 1 \implies p_k = \frac{1}{|\mathcal{J}|} - \frac{1}{2\alpha|\mathcal{J}|}\sum_{j\in\mathcal{J}}(v_j - v_k)$$

If $\mathbf{p}^* \in \text{int}\Delta^n(X)$, the solution is thus $p_k^* = \frac{1}{n} - \frac{1}{2\alpha}[\bar{v} - v_k]$ where $\bar{v} = \frac{1}{n}\sum_{j=1}^n v_j$. Note that $p_k^* > \frac{1}{n} \iff v_k > \bar{v}$ and $\mathbf{p}^* \to \left(\frac{1}{n}, ..., \frac{1}{n}\right)$ as $\alpha \to \infty$. If $\mathbf{p}^* \notin \text{int}\Delta^n(X)$, because utility is strictly increasing in v_k for all k, the solution assigns positive probability only to the first $|\mathcal{J}|$ destinations in the DM's ordering. In the limit as $\alpha \to 0$, $\mathbf{p}^* = \delta_{x_1}$.

Shannon entropy I ignore constraint (1b), i.e., $p_k \geq 0$ for all k, and will show that it is never binding. Differentiating the Lagrangian, the first-order condition for p_k is $v_k - \alpha\ln(p_k) - \alpha - \lambda = 0$. Combining the FOCs for two $j, k \in \mathcal{J}$ yields:

$$\ln\left(\frac{p_j}{p_k}\right) = \frac{v_j - v_k}{\alpha} \iff p_j = \exp\left(\frac{v_j - v_k}{\alpha}\right)p_k$$

Summing over all $j \in \{1, ..., n\}$ and using the fact that $\sum_{j=1}^n p_j = 1$:

$$1 = \sum_{j=1}^n \exp\left(\frac{v_j - v_k}{\alpha}\right)p_k = \frac{\sum_{j=1}^n \exp(v_j/\alpha)}{\exp(v_k/\alpha)}p_k \implies p_k^* = \frac{\exp(v_k/\alpha)}{\sum_{j=1}^n \exp(v_j/\alpha)}$$

Note that $p_k^* \in (0,1)$ for any vector of valuations $\mathbf{v}$ and any α, so (1b) is indeed never binding. Furthermore, $\mathbf{p}^*$ tends to $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$ as $\alpha \to \infty$ or as the difference between any two valuations $(v_j - v_k)$ tends to 0. The entropy case has properties that are very similar to the residual variance case, except for the absence of corner solutions.

Variance in valuations The general problem does not admit a closed-form solution. For $n = 2$, the problem simplifies to $\max_{p\in[0,1]} \; pv_j+(1-p)v_k+\alpha p(1-p)(v_j-v_k)^2$. Taking the first-order condition yields $p^* = \frac{1}{2} + \frac{1}{2\alpha\Delta v}$ if $\alpha > \frac{1}{\Delta v}$ ($p^* = 1$ otherwise). Note that $p^* \to \frac{1}{2}$ as $\Delta v \to \infty$; in other words, the DM prefers more uncertainty as the two destinations over which they randomize are further apart in their preferences. It is easy to see that variance violates X-MON even for $n = 2$ by observing that $x_j \succ (x_j, p; x_k, 1-p) \iff v_j > pv_j + (1-p)v_k + \alpha p(1-p)(v_j - v_k)^2$, i.e., $\alpha p\Delta v < 1$. Thus, for Δv large enough, a decrease in v_k (holding v_j fixed) increases the chances that the DM will prefer the lottery over x_j.

F.3 Theories of expectation-based reference dependence

SD violations may occur if the DM judges each outcome of a lottery $\mathbf{p}$ relative to an expectation-based reference point and experiences gain-loss utility when the realized outcome deviates from it. I consider the simple case in which the reference point is unidimensional and depends only on the DM's valuation of each trip so that $U(\mathbf{p}|\mathbf{r}) = \sum_{k=1}^{n} p_k \left[v_k + \nu(v_k|\mathbf{r})\right]$, where $\mathbf{r}$ is a reference lottery and $\nu(v_k|\mathbf{r})$ is the gain-loss utility of receiving outcome x_k (valued at v_k), given $\mathbf{r}$. Below I study two classes of models, which assume that the reference lottery is the one that the DM chose ($\mathbf{r} = \mathbf{p}$).

Models of Disappointment Aversion (DA) The first class of models introduced by Bell (1985) and Loomes and Sugden (1986) takes the reference point to be the expected utility of the chosen lottery. The gain-loss utility of outcome x_k is thus $\nu(v_k|\mathbf{p}) := \mu(v_k - E_{\mathbf{p}}[v])$, where $\mu(.)$ is a gain-loss function such that $\mu(0) = 0$.

Kőszegi and Rabin (KR) The second class of preferences are those induced by the Choice-Acclimating Personal Equilibrium (CPE) concept of Kőszegi and Rabin (2007), in which the choice is the reference point. After every realization x_k, the DM compares v_k to what was expected given $\mathbf{p}$ i.e., $\nu(v_k|\mathbf{p}) := \sum_{j=1}^{n} p_j \mu(v_k - v_j)$.

To facilitate comparisons between models, I assume that the gain-loss function is of the form $\mu(\Delta v) = (\Delta v)^\gamma$ if $\Delta v \geq 0$ and $-\lambda(-\Delta v)^\gamma$ if $\Delta v < 0$. Below I discuss the two (most frequent) cases of linear and quadratic gain-loss i.e., $\gamma \in \{1, 2\}$.

CASE 1: Linear gain-loss utility ($\gamma = 1$) Consider any binary-outcome lottery $\mathbf{p}$ given by $(x_j, p; x_k, 1-p)$ where $x_j \succ x_k$ and let $E_{\mathbf{p}}[v] = pv_j + (1-p)v_k$. With two outcomes, DA and KR are equivalent. The DM's utility of lottery $\mathbf{p}$ is

$$U_{KR}(\mathbf{p}|\mathbf{p}) = E_{\mathbf{p}}[v] + (1-\lambda)p(1-p)(v_j - v_k) = U_{DA}(\mathbf{p}|\mathbf{p})$$

Note that $(x_j, p; x_k, 1-p) \succ x_j$ iff $E_{\mathbf{p}}[v] + (1-\lambda)p(1-p)(v_j - v_k) > v_j$, i.e., $(1-\lambda)p > 1$. This can only happen if $\lambda < 0$, which contradicts loss aversion ($\lambda > 1$). Furthermore, even if $\lambda < 0$, the DM will switch at most once from x_j to $(x_j, p; x_k, 1-p)$ as p increases i.e., P-MON must be satisfied on the set of weak dominance problems, $\mathcal{B}^+_{\sim}$. For strict dominance problems such as $\{x_i, (x_j, p; x_k 1-p)\}$ with $x_i \succ x_j \succ x_k$, both P-MON and X-MON may be violated if $\lambda < 0$.[47] Beyond, it can be shown that P-MON and X-MON may be violated with lotteries containing more than two outcomes.

CASE 2: Quadratic gain-loss utility ($\gamma = 2$) When $\mu(.)$ is quadratic, KR corresponds to the mean-variance case with $\alpha = 1-\lambda$ (see Masatlioglu and Raymond (2016)). Thus, for $\lambda < 1$, KR may violate both X-MON and P-MON, except on $\mathcal{B}^+_{\sim}$. DA may generate violations of both P-MON and X-MON even on $\mathcal{B}^+_{\sim}$.[48]

F.4 Ambiguity aversion and hedging

One could treat a trip x_j as a Savage act $\{f_j(\omega)\}_{\omega \in \Omega}$, assigning different payoffs in different states of the world $\omega \in \Omega$. Below I show how a DM might prefer $pf_j + (1-p)f_k \succ f_j \succ f_k$ if (i) they hold multiple priors $\pi \in \Delta(\Omega)$; (ii) they believe that Nature is adversarial in the sense of Gilboa and Schmeidler (1989) so that

$$f_j \succeq f_k \;\Leftrightarrow\; \min_{\pi \in \Pi} \sum_{\omega \in \Omega} \pi(\omega) f_j(\omega) \geq \min_{\pi \in \Pi} \sum_{\omega \in \Omega} \pi(\omega) f_k(\omega)$$

As an example, assume $\Omega := \{\omega_1, \omega_2\}$ with $\pi := \Pr\{\omega = \omega_1\}$ and there are two trips x_j, x_k, which yield payoffs $(f_j(\omega_1), f_j(\omega_2)) = (2, 5)$ and $(f_k(\omega_1), f_k(\omega_2)) = (4, 1)$. For instance, Ω could capture the weather with ω_1 = "sun" and ω_2 = "rain", and

[47] To see this, note that $x_i \succ (x_j, p; x_k, 1-p) \iff v_i > [(2-\lambda)p - (1-\lambda)p^2](v_j - v_k)$. For $\lambda < 0$, $\Gamma(p) := (2-\lambda)p - (1-\lambda)p^2$ attains its maximum at $p = \frac{2-\lambda}{2-2\lambda}$. Because $\Gamma(p) > 0$ for all p, the RHS expression is increasing in the payoff difference $(v_j - v_k)$, implying a potential violation of X-MON.

[48] In this case, $x_j \succ (x_j, p; x_k, 1-p) \iff v_j > E_{\mathbf{p}}[v] + p(v_j - E_{\mathbf{p}}[v])^2 - \lambda(1-p)(v_k - E_{\mathbf{p}}[v])^2$ i.e., $1 > [p(1-p) - \lambda p^2]\Delta v$. The RHS expression attains its maximum at $p = \frac{1}{2(1+\lambda)} \in (0,1)$ if $\lambda > -\frac{1}{2}$.

each destination differing in the quality of outdoor and indoor activities. If forced to commit to a destination, the DM will choose x_j, which yields a worst possible payoff of 2. Now consider a mixed act $f^p_{j,k} := pf_j + (1-p)f_k$ for some probability $p \in (0,1)$. The optimal p^* solves $2p + 4(1-p) = 5p + (1-p)$, yielding $p^* = \frac{1}{2}$. In other words, the DM will prefer a "mixed" act, which gives minimum payoff guarantees under any state of the world, to a more extreme act. More generally, assume $f_k(\omega_1) > f_j(\omega_1)$, $f_k(\omega_2) < f_j(\omega_2)$ and $f_j(\omega_1) = \min\{f_j(\omega_1), f_j(\omega_2)\} > f_k(\omega_2) = \min\{f_k(\omega_1), f_k(\omega_2)\}$. Let p^* solve $pf_j(\omega_1) + (1-p)f_k(\omega_1) = pf_j(\omega_2) + (1-p)f_k(\omega_2) > f_j(\omega_1)$ i.e.,

$$p^* = \frac{f_k(\omega_1) - f_k(\omega_2)}{f_k(\omega_1) - f_j(\omega_1) + f_j(\omega_2) - f_k(\omega_2)}$$

Clearly, $p^* f_j + (1-p^*)f_k \succ f_j \succ f_k$. The payoff function is piecewise linear. It is increasing in p for all $p < p^*$ and decreasing in p for $p > p^*$. In particular:

- If $p < p^*$, $\min\{f^p_{j,k}(\omega_1), f^p_{j,k}(\omega_2)\} = f^p_{j,k}(\omega_2)$ and $\lim_{p\to 0} f^p_{j,k}(\omega_2) = f_k(\omega_2)$
- If $p > p^*$, $\min\{f^p_{j,k}(\omega_1), f^p_{j,k}(\omega_2)\} = f^p_{j,k}(\omega_1)$ and $\lim_{p\to 1} f^p_{j,k}(\omega_1) = f_j(\omega_1)$.

Let $\tilde{p} \in (0, p^*)$ solve $f^p_{j,k}(\omega_2) = f_j(\omega_1)$. Then for all $p \in [\tilde{p}, 1]$, $\min\{f^p_{j,k}(\omega_1), f^p_{j,k}(\omega_2)\} \geq f_j(\omega_1)$. In words, if the DM prefers the mixed act $f^p_{j,k}$ to the simple act f_j at $p = \tilde{p}$, then they will still (weakly) prefer $f^p_{j,k}$ for all $p > \tilde{p}$ i.e., P-MON must be satisfied.

Trust and hedging It is unlikely that preference for randomization reflects hedging concerns due to respondents mistrusting the experiment. First, self-reported mistrust was low overall. At the end of the survey, respondents were asked to rate their level of trust on a scale from 1 to 5 (1 = "Do not trust at all"; 5 = "Trust completely") regarding 3 aspects of the experiment: (i) "*The trips offered in this study are indeed worth £420.*"; (ii) "*The selection procedure strictly follows the rules specified in the experiment.*"; (iii) "*The lottery draws are indeed random.*" Across all statements, the fraction of respondents who gave a trust rating of 4 or 5 was around 50% or more, and at most 23% gave a rating of 1 or 2 on any statement. In addition, I examined regressions of each trust rating on various measures of preference for randomization and delay. Virtually all coefficients are insignificant, and the few significant ones in fact show a *negative* relationship between mistrust and preference for randomization or delay (see supplemental appendix at https://osf.io/ya7x6/ for details).