Are you an exception to your favorite decision theory? Behavioral decision research is a Linda problem!

Michel Regenwetter

Corresponding author. Departments of Psychology, Political Science, Electrical & Computer Engineering, University of Illinois at Urbana-Champaign

regenwet@illinois.edu. 603, E.Daniel Str., Champaign, IL61820

Maria M. Robinson

Department of Psychology, University of California San Diego

mrobinson@ucsd.edu

Cihang Wang

Department of Economics, University of Illinois at Urbana-Champaign

cwang153@illinois.edu

Abstract

Stylized characteristics, such as loss aversion, risk aversion for gains, risk seeking for losses, overweighting of small probabilities, and diminishing sensitivity permeate both popular science and scholarly treatises about how 'people' make decisions. This note highlights that behavioral decision research is, in effect, a large-scale Linda problem: The likelihood that a given individual satisfies the conjunction of many such stylized characteristics may be vanishingly small. We concentrate on a case study, namely the pervasive oversimplifications surrounding Amos Tversky and Daniel Kahneman's Prospect Theory and Cumulative Prospect Theory. Focussing entirely on evidence from within the original papers, we show that each and every person may be a exception to (Cumulative) Prospect Theory as advertised. Similar problems afflict many other behavioral research paradigms. We call on scholars to relinquish overly simplified characterizations of choice behavior. Telling practitioners and laypersons in stylized fashion how 'people' think promotes conjunction fallacies on a huge scale. Rather than conceptualize individual differences as a mere add-on to a schematic decision theory of central tendencies, decision scholars should recognize heterogeneity as a major theoretical primitive when proposing new theories.

Keywords: Stylized features; scientific conjunction fallacy; theoretical scope

Are you an exception to your favorite decision theory? Behavioral decision research is a Linda problem!

Disclosures and Acknowledgments

M. Regenwetter guided and supervised the project and contributed most of the writing. All authors have read and approved the final manuscript. M. Robinson and C. Wang carried out all computations while M. Robinson was a PhD student in Psychology at Illinois and C. Wang an undergraduate Mathematics/ Economics double major at Illinois. This work was supported financially by National Science Foundation grants SES-14-59699 and SES # 20-49896 to Regenwetter (PI) and Army Research Office MURI Grant W911NF-20-1-0252 (PI: C. Langbort, Co-PIs: M. Başar, M. Regenwetter). ARO and NSF had no other role besides financial support. The authors are not aware of any conflicts of interest. Part of this paper has previously been presented orally at the annual Society for Mathematical Psychology satellite meeting of Psychonomics in 2019, and at the 57th Annual Edwards Bayesian Research Conference in 2020.

Thanks to David Budescu, four referees, Meichai Chen, Brittney Currie, Ido Erev, Yu Huang, Emily Line, Xiaozhi Yang, for comments on earlier drafts. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARO, NSF, the U.S. Government, colleagues, or the authors' home institutions. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

This manuscript was written in LaTeX.

Introduction

Summary statistics are ubiquitous. According to the U.S. census QuickFacts table (https://www.census.gov/quickfacts/fact/table/US/PST045219), the modal race in the U.S. is white, there are slightly more women than men, most people are between 18 and 65 years old, the mean travel time to work is about 27 minutes, the average household size is 2.63, and the median household income is slightly over $60K. Regardless how interesting these statistics may be, one must use them with caution. For example, it would be a foolhardy business strategy, respectively disastrous public policy, for managers and/or policy makers to treat 'consumers' or 'citizens' as though they were, essentially, white women aged between 18 and 65, whose daily commute is about 27 minutes, and who live with one or two other people in a household making altogether slightly more than $60K per year, give or take some noise[1]. The people who either match or closely resemble this 'prototype' presumably make up only a tiny fraction of all U.S. consumers. Regardless of whether we consider consumer, health, manufacturing, service, or other industries, only the most specialized of companies focus on customers with such a narrow profile, and when they do, their goal is more likely to serve a niche market segment, rather than to target the 'prototypical' census respondent. We all know that incorrectly combining summary statistics, which constitutes a logical fallacy of composition, contributes to disinformation and can cause serious policy debacles. Just as it makes little sense to limit a business strategy or public policy to the modal, mean, or median person according to

[1] We use "double quotes" to quote verbatim, to indicate that a word may be used in more than one sense, or to indicate a technical term whose definition we omit. We use 'single quotes' to highlight a vague or ill-defined term.

socio-economic indicators, so should we avoid treating ‘people’ as approximate clones of a mystical modal or median decision maker. In this note, we unpack this simple proposition in the case of Kahneman and Tversky’s hyperinfluential (Cumulative) Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Our goal is not to take a stance on any particular theory like (Cumulative) Prospect Theory. Rather, for most of the paper, we employ (Cumulative) Prospect Theory as a canvas on which to paint a picture that applies to much of behavioral decision research, regardless of paradigm. (Cumulative) Prospect Theory is a choice of great convenience because it is well known to the readers of this journal. Because there are many diverse opinions as to what constitutes ‘genuine’ (Cumulative) Prospect Theory, in the interest of being as concrete and transparent as we can, we focus on the theory as it was stated in its original publication. Our warnings about conjunction fallacies apply to other versions of the theory and other theories as well. (Cumulative) Prospect Theory is also widely used and broadly cherished across numerous disciplines of basic and applied science. We touch on other paradigms and theories in a later section.

Behavioral decision research is a Linda problem!

Successful popular science books, including Ariely (2010); Kahneman (2011); Thaler and Sunstein (2008), have propelled decision theoretic concepts such as reference points, risk aversion, risk seeking, diminishing sensitivity, loss aversion, overweighting of rare events, into popular culture. Scientific works overflow with stylized claims about how ‘people’ make decisions under risk (see, e.g Abdellaoui, L’Haridon, & Paraschiv, 2011;

Ariely, Huber, & Wertenbroch, 2005; Barberis, 2013; Benartzi & Thaler, 1995; Birnbaum, 2008; Birnbaum & Bahra, 2007; Brandstätter, Gigerenzer, & Hertwig, 2006; Camerer & Ho, 1994; de Langhe & Puntoni, 2015; Diederich & Trueblood, 2018; Feess, Müller, & Schumacher, 2016; Fennema & Van Assen, 1998; Gehring & Willoughby, 2002; Grolleau, Kocher, & Sutan, 2016; Harinck, Van Dijk, Van Beest, & Mersmann, 2007; Henderson, 2012; Kermer, Driver-Linn, Wilson, & Gilbert, 2006; Levy, 2016; McGraw, Larsen, Kahneman, & Schkade, 2010; Timmer, 2012; Tom, Fox, Trepel, & Poldrack, 2007; Trepel, Fox, & Poldrack, 2005; Tversky & Kahneman, 1992; Usher & McClelland, 2004; Weingarten, Bhatia, & Mellers, 2019; G. Wu & Gonzalez, 1996), oftentimes relying heavily or even exclusively on the behavior of modal or median decision makers to draw inferences from their own or others' work.

To see the danger of such depictions, we need to look no further than the Linda problem of Tversky and Kahneman (1983) and others. In the Linda problem, put simply, some research participants appear to think that there are more feminist bank tellers who satisfy the characteristics of a fictitious woman called Linda than there are bank tellers who match the same description. This is a mathematical impossibility since there are fewer people who satisfy a combination of "feminist" and "bank teller" than there are people who satisfy just one of these two characteristics. The phenomenon is also often referred to as a *representativeness heuristic* according to which decision makers may violate the rules of probability theory by over-estimating the number of people who satisfy a conjunction of characteristics because it is 'easier to imagine' that they satisfy the combination of features than to entertain realistic nuanced scenarios. Listing "characteristics of human decision making" in stylized fashion, as it is common in works that aim to educate scholars, practitioners, and laypeople about decision making, has the potential to communicate and/or nourish conjunction fallacies about human decision making behavior on a huge scale.

Perhaps the most prominent example of a Linda problem fueling a large scale research program is Kahneman and Tversky (1979). They discussed a collection of "pervasive effects" such as the "Allais paradox," the "certainty effect," the "isolation effect," and others, which they labeled as behavioral "tendencies" that challenge expected utility theory as a descriptive model of decision making under risk.

Table 1

"Pervasive effects" of Kahneman and Tversky (1979) with lower and upper bounds on the proportion of people who exhibit them.

"Pervasive Effect"	Problem Numbers	Prop. V in V vs. W	Prop. X in X vs. Y	Proportion of people exhibiting $\big[[V \text{ AND } X] \text{ OR } [W \text{ AND } Y]\big]$	
				Lower Bound	Upper Bound
Allais Paradox	1 & 2	.82	.83	.65	.99
Allais Paradox	3 & 4	.80	.65	.45	.85
Certainty Effect	5 & 6	.78	.67	.45	.89
Possibility	7 & 8	.86	.73	.59	.87
Reflection Effect	3 & 3'	.80	.92	.72	.88
Reflection Effect	4 & 4'	.65	.58	.23	.77
Reflection Effect	7 & 7'	.86	.92	.78	.94
Reflection Effect	8 & 8'	.73	.70	.43	.97
Isolation Effect	4 & 10	.65	.78	.57	.88
Isolation Effect	11 & 12	.84	.69	.53	.85
Gains vs. Losses	13 & 13'	.82	.70	.52	.88
Prob. weighting	14 & 14'	.72	.83	.55	.89

Note. We treat the sample proportions of Kahneman and Tversky (1979) as fully accurate population proportions. In each pair of problems, we use V to denote the modal choice in one problem and X to denote the modal choice in the other problem.

For example, in Kahneman and Tversky's Problems 1 & 2 on the "Allais paradox," the decision maker had the choice between option A and option B in Problem 1 and the choice between C and D in Problem 2. An expected utility maximizer either prefers both A to B and C to D or prefers both B to A and D to C. We will take Kahneman and Tversky's methods, measurements, and inferences at face value and presume that 82% of the population prefers B to A, whereas 83% of the population prefers C to D. Based on that assumption, how many individuals succumb to the type of Allais paradox that Kahneman and Tversky considered? If those 82% of the population who prefer B to A also all prefer C to D, and those 17% of the population who prefer D to C also all prefer A to B, then 99% of the population succumbs to this Allais paradox. However, if those 18% who prefer A to B also prefer C to D, and those 17% who prefer D to C also prefer B to A, then 100-18-17=65% of the population succumb to this Allais paradox.

In Table 1, we tally Kahneman and Tversky's "pervasive effects" accordingly. For each pair of decision problems, we use V to denote the modal choice in one problem, and X to denote the modal choice in the other problem. The right two columns give lower and upper bounds on the proportion of the population who would show the "pervasive effect," either in the form of preferring both V and X, or in the form of preferring both W and Y.

Kahneman and Tversky (1979) quite unambiguously promoted Prospect Theory as a theory that can handle the conjunction of these behavioral phenomena within individuals. Thinking of this problem as a large scale Linda problem, we can naturally ask how many individuals appear to satisfy that conjunction. Clearly, similar to the original

Linda problem, it cannot be more than the smallest value in the “upper bounds” column. In other words, still taking Kahneman and Tversky’s measurements as true proportions of the population, we note that only at most 77% satisfy the “reflection effect” of Problems 4 & 4’ in the table. This means that at most 77% of the population show the conjunction of these 12 “pervasive effects.” Kahneman and Tversky (1979) mentioned additionally, for some of these behavioral regularities, such as their Allais paradox, that a majority of individuals displayed the phenomenon (within person). Therefore, to be as charitable to Kahneman and Tversky (1979) as we can, let us briefly assume that every phenomenon in the table has an incidence rate matching the upper bound (in the right most column of the table). Under that very lenient assumption, how many people can be exceptions to at least one of these “pervasive effects?” The answer is: Everybody can be an exception! At the opposite extreme, using the lower bound in each row, each individual person could be an exception to three or more of these 12 phenomena.

To summarize what we have found so far, the behavioral phenomena in Kahneman and Tversky (1979) form a large scale Linda problem. While Kahneman and Tversky (1979) unambiguously rely on the conjunction of these phenomena to motivate Prospect Theory, and later Cumulative Prospect Theory, even if we take their own data as perfect measurements of population proportions and even if we permit the most charitable proportions of occurrences of these phenomena based on their numbers, at most 77% of the population satisfies the conjunction of those 12 phenomena. It is mathematically possible, with those given proportions, that nobody satisfies the conjunction. Yet, even if few people may display the conjunction of these properties, Kahneman and Tversky appear to motivate Prospect Theory and Cumulative Prospect Theory by a need to capture the combination of these phenomena within the same individuals.

Cumulative Prospect Theory (CPT)

As a single species, humans must share important invariants underlying decision making. Yet, the sheer amount of variability in behavior also strongly suggests that this invariant or aggregate structure, just like aggregate socio-economic indicators, is overlaid with a wide spectrum of nontrivial individual differences. All too often, these individual differences, rather than forming a theoretical primitive, are relegated to the role of theoretical afterthoughts. They are often tagged on to enhance a schematic decision theory about trends, significant effects, and central tendencies. As we see next, CPT allows for a much more nuanced mind set.

To succeed, a descriptive theory of individual decision making needs to strike a difficult balance between commonalities and differences among decision makers. Tversky and Kahneman's (1992) Nobel-Prize winning Cumulative Prospect Theory is a prominent exemplar of such a theory: Its core mathematical formula might capture a universal invariant, while its free parameters leave the door open to extensive heterogeneity. According to the original version given in Eqs. 5-6 of Tversky and Kahneman (1992), CPT uses a value function v, as well as two probability weighting functions w^+ and w^- for gains and losses, that invoke parameters $\alpha, \beta, \gamma, \delta, \lambda$, such that, writing x for money amount and

p for a probability,

$$v(x) = \begin{cases} x^\alpha & \text{when } x \geq 0 \quad (\text{with } 0 \leq \alpha \leq 1), \\ -\lambda(-x)^\beta & \text{when } x < 0 \quad (\text{with } 0 \leq \beta \leq 1; 0 < \lambda), \end{cases} \tag{1}$$

$$w^+(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}} \quad (\text{with } 0 < \gamma \leq 1), \tag{2}$$

$$w^-(p) = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{\frac{1}{\delta}}} \quad (\text{with } 0 < \delta \leq 1). \tag{3}$$

CPT further combines these equations via a composition rule to generate subjective utilities of uncertain prospects[2]. While the parameter ranges in Equations 1-3 are subject to debate[3], they are popular in the literature because they capture the stylized behavioral characteristics that TK emphasized in their paper.[4]

While the theory was formulated in a fashion that accommodates both invariance and heterogeneity, and even though they carried out some individual subjects studies, Tversky and Kahneman (henceforth TK) chose to rely heavily on summary statistics in reporting and interpreting their findings. They reported (pp. 311-312) median estimated α and β values of 0.88, which they interpreted as showing “diminishing sensitivity,” a median estimated λ of 2.25, which they interpreted as “pronounced loss aversion,” a median estimated γ of 0.61 and a median estimated δ of 0.69, which they interpreted as

[2] This is the version of CPT that we concentrate on. One can apply similar techniques to consider conjunction fallacies in the context of other variants of CPT. Note that problems with conjunction fallacies would be amplified further if we considered generalizations of these formulae, because the theory would allow for even more heterogeneity.

[3] Enlarging the ranges of these parameters would create more opportunity for individual differences and thereby amplify the danger of conjunction fallacies.

[4] Like the original CPT paper, we do not look particularly closely at choice options with more than two outcomes, where the form of the weighting function would differ the most from other models.

"overweighting the probabilities of rare events." We refer to Equations 1 - 3, with $\alpha = \beta = 0.88, \lambda = 2.25, \gamma = 0.61$, and $\delta = 0.69$ as CPT_{MED}. As we mentioned, we neither intend to question nor endorse CPT as a theory. Some people may violate Equations 1 - 3 entirely. Others may satisfy the equations but violate CPT_{MED} by having different parameter values. We will emphasize that, even among those people who satisfy Equations 1 - 3 together with the composition rule that combines them, and even taking Tversky and Kahneman's reported median parameter values as 100% accurate estimates of the population medians, there may, nonetheless, be as much diversity in decision making behavior as there is socio-economic diversity among those summarized by the U.S. Census QuickFacts table.

We advocate that summarizing behavior or a theory through stylized properties, median parameter values, or median responses, jeopardizes our very understanding of that behavior, as well as our understanding of that theory. In our view this practice runs fundamentally counter to good decision science. For example, CPT can jointly model various phenomena that violate classical models like expected utility theory, but it also leaves open the possibility that many individuals obey classical models. Yet, this is not how either proponents or opponents typically characterize CPT. For instance, when TK reported (p. 304) that 46% of Stanford students violated an *independence* property required by expected utility theory, they did not highlight the flip-side that about half of these participants may indeed have satisfied independence. As a consequence of this and similar studies, many scholars presume that 'people' violate the property. Likewise, 'people' allegedly use reference points, 'people' are risk averse for gains and risk seeking for losses, 'people' are subject to diminishing sensitivity, and, 'people' overweight rare events.

While TK reported and acknowledged individual differences, they ultimately relied heavily on strings of modal choices (see, e.g., their discussion of their Tables 1 and 2), as well as on proportions of 'correct predictions' doubly aggregated across participants and decisions (see e.g., their discussion of their Table 4) to promote CPT.[5] For the remainder of this note, we caution against over-interpreting the stylized characterization of decision making associated with the median responses and median parameters in Tversky and Kahneman (1992), as well as other summary statistics. We are concerned that overly abstract summary statistics may exercise a disproportionate and potentially counterproductive influence on management and policy across all aspects of society, by encouraging and rewarding systemic conjunction fallacies.

Exceptions, exceptions, exceptions.

The U.S. census QuickTable provides a very high level, and simple, summary of a society that features great socio-economic diversity. How does this insight translate if we were to treat the summary statistics in Tversky and Kahneman (1992) as a 'QuickFacts table' of human decision making? The initial goal of this project was straightforward: Explore how much heterogeneity in decision making behavior is consistent with Tversky

[5] Yet, like the conjunction of aggregate socio-economic indicators from the QuickFacts table is descriptive of only few Americans, such aggregate measures are uninformative about a theory's empirical fit or scope (see also Regenwetter & Robinson, 2017, who point out inherent logical inconsistencies in such methods and measures). Our goal is not to discuss measures of fit per se, but rather the dangers associated with stylized characterizations of behavior.

and Kahneman's own (1992) reported results. Suppose that everybody satisfies Equations 1-3, as well as the composition rule that combines them. Suppose also that the population median values of $\alpha, \beta, \delta, \gamma, \lambda$ are $0.88, 0.88, 0.69, 0.61$, and 2.25, i.e., suppose that TK have correctly inferred these aggregate measures with perfect accuracy.

Of special importance here will be TK's *loss aversion* study. To establish loss aversion, TK asked participants to provide a dollar amount x such that they would be indifferent between a 50/50 chance of winning a or b and a 50/50 chance of winning either c or x. They used eight different decision problems that differed in their values of a, b, and c and they reported the median observed value of x in each decision problem. In addition to taking the median parameter values of $\alpha, \beta, \delta, \gamma, \lambda$ at face value, suppose that the true population median x values in TK's eight (loss aversion) problems are indeed 61, 101, 202, 280, 112, 301, 149, and 401, respectively (Tversky and Kahneman, 1992, p. 312, Table 6), i.e., suppose again that TK have correctly inferred these population characteristics with perfect accuracy. We know that a few dozen summary statistics in the U.S. census QuickTable abstract away from huge heterogeneity among about 300 million people. If we take Tversky and Kahneman's median parameters and median x values at face value, how heterogeneous can 7 billion human decision makers be? The answer is surprising.

Exceptions to the mathematics of CPT:

We first discovered that our assumptions led to a contradiction! We found that it is impossible, in Problem 7 of TK's loss aversion study, to generate $x \leq 149$ using Equations 1-3. Hence, if the population median x for Problem 7 is indeed 149 (as TK reported), then at least half of the respondents are exceptions to CPT in that they must violate Equations 1-3. Second, we discovered that, in Problem 8, it is impossible to have TK's reported median x of 401 while also having TK's median γ of 0.61. We have reported these and other internal inconsistencies in the Tversky and Kahneman (1992) paper elsewhere (Regenwetter, Robinson, & Wang, n.d.). As we have mentioned earlier, our goal here is not to revisit TK's methodology or quality of evidence, but rather to warn of potential scientific conjunction fallacies. In order to proceed, we drop Problems 7 and 8 of TK's loss aversion study from further consideration because they would lead to mutually incompatible constraints.

A hypothetical scenario:

We now concentrate on a presumed sub-population that satisfies Equations 1-3 (yielding 3 functional and range constraints), whose median parameter values match TK's reported medians (yielding 5 statistical constraints), and whose median x values in Problems 1-6 in the loss aversion study match TK's reported medians (yielding 6 statistical constraints). These 14 constraints permit huge qualitative individual differences: Everybody who satisfies CPT may nonetheless be an exception to one or several of the theory's prominently advertised, stylized, properties. Notice that this is, in fact, a desirable feature of any theory that potentially aims to model all individuals. In other words, our point is again not to label this as a weakness of CPT, but rather to warn against conjunction fallacies in the description and interpretation of CPT by proponents and opponents alike. The same point applies to other theories more broadly.

		H_1	H_2	H_3	H_4	H_5	H_6	H_7	Medians
	α	0.5	0.88	0.1	1	0.1	.95	1	0.88
	β	0.88	0.46	0.99	0.1	1	0.66	1	0.88
	λ	0.093	9.67	4.11	0.681	0.001	9.95	2.25	2.25
	δ	0.4	0.9	0.1	0.69	1	1	0.1	0.69
	γ	0.2	1	0.61	0.1	0.264	0.95	1	0.61
Problem		H_1	H_2	H_3	H_4	H_5	H_6	H_7	Medians
1	x	61	70	0.0002	234	10^{-10}	105	0.21	61
	θ	2.44	2.8	10^{-5}	9.36	10^{-11}	4.2	0.008	2.44
2	x	205	101	0.2	251	10^{-7}	170	0.41	101
	θ	4.1	2.02	0.0042	5.02	10^{-9}	3.4	0.008	2.02
3	x	695	145	202	269	0.0001	276	0.82	202
	θ	6.95	1.45	2.02	2.69	10^{-6}	2.76	0.008	2.02
4	x	1418	179	11192	280	0.009	366	1.23	280
	θ	9.45	1.19	74.6	1.87	0.00006	2.44	0.008	1.87
5	x	225	88	970	72	112	129	50	112
	θ	5.83	1.27	31	0.73	2.07	2.63	0	2.07
6	x	900	215	44321	174	843	301	151	301
	θ	10.0	0.87	589	0.32	9.24	2.01	0.0133	2.01

Table 2

Hypothetical parameter values, x values, and $\theta = |(x-b)/(c-a)|$ values according to H_1 - H_7. All medians, listed in the right most column, match those reported by Tversky and Kahneman (1992).

Table 2 helps us understand the range of possibilities. It shows the 11 statistical constraints in the right most column. We created seven hypothetical decision makers such that the median values of the parameters match the parameters in CPT_{MED} and such that the values of x in Problems 1-6 (loss aversion study) give the median x values reported by Tversky and Kahneman (1992). The CPT parameter values of the seven hypothetical decision makers are given in the top section of the table. In the lower section, the table provides the values of x and of $\theta = |(x-b)/(c-a)|$ implied by H_1 - H_7 for Problems 1-6. Hand-in-hand with this table, Figures 1-2 display the weighting functions and value

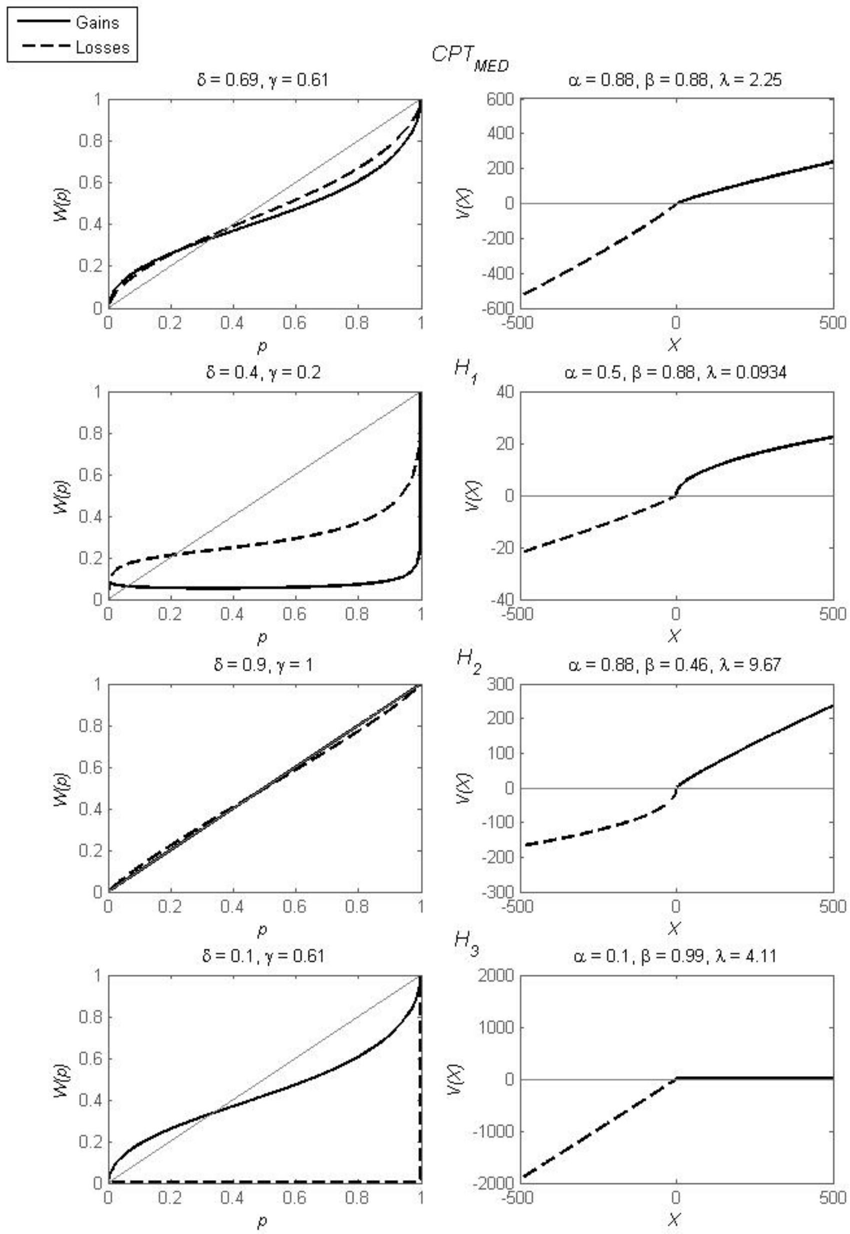

Figure 1
Probability weighting functions w^+, w^- *as functions of* p *(left column) and value function* v *as a function of* x *(right column), for the four different combinations of* $\alpha, \beta, \gamma, \delta, \lambda$, *labeled* CPT_{MED} *as well as* H_1 *-* H_3 *of Table 2.*
Note: The vertical axis for $v(x)$ *is scaled differently for different cases in order to make each shape clearly visible.*

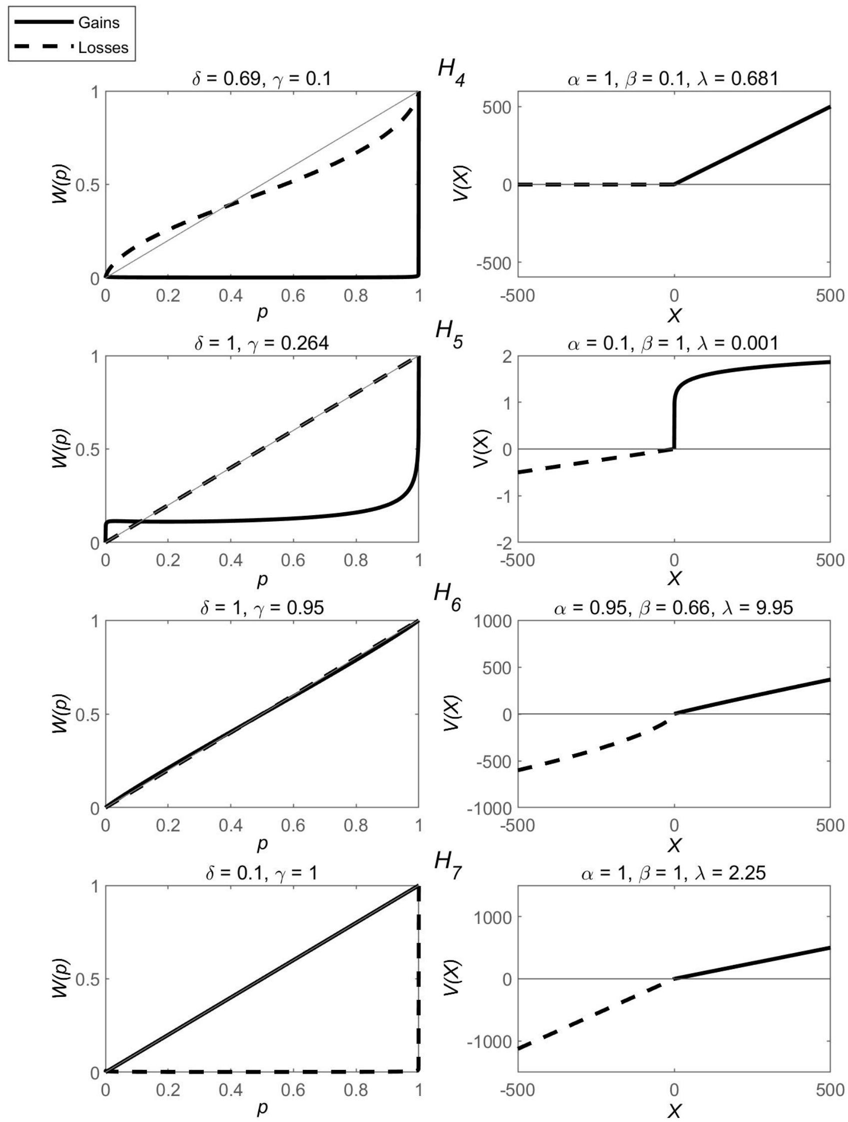

Figure 2

Probability weighting functions w^+, w^- as functions of p (left column) and value function v as a function of x (right column), for the four different combinations of $\alpha, \beta, \gamma, \delta, \lambda$, labeled H_4 - H_7 in Table 2.

Note: The vertical axis for $v(x)$ is scaled differently for different cases in order to make each shape clearly visible.

functions associated with CPT_{MED} and $H_1 - H_7$, respectively.[6]

[6] As a referee noted, while satisfying Equations 2 and 3, some of these weighting functions are not monotononic.

Exceptions to the stylized properties of CPT:

Table 2 gives a particularly good sense of the potential individual differences regarding the presence and degree of loss aversion. Interpreting a median λ value of 2.25 as "pronounced loss aversion" (as Tversky & Kahneman, 1992, do on p. 311) is misleading: The small λ values in H_1, H_4, H_5 imply that arbitrarily close to half the population could be exceptions to "pronounced loss aversion," as defined by, say, $\lambda > 2$, even if everybody satisfied CPT. TK also used the transformed statistic $\theta = |(x-b)/(c-a)|$, whose median is near 2 in Problems 1-6, to suggest that "for even chances to win and lose, a prospect will only be acceptable if the gain is at least twice as large as the loss." We pointed out already that three values of λ in Table 2 are small. In addition, however, the values of θ vary hugely both across decision problems as well as across columns. Even though the median θ is near 2 in every decision problem, it would be profoundly incorrect to characterize H_1-H_7 as though "losses loomed roughly twice as large as gains," as the now popular saying goes. In Table 1, fifteen θ values, out of 42, are smaller than 1, and ten are larger than 5, for example. While each column features one of the six median x and median θ values, every single column varies substantially in its x and θ values.

To discuss some of the other stylized features of CPT, we turn to visualizations. Our Figures 1 and 2 show w^+, w^- (left column), and v (right column), for CPT_{MED} as well as for the seven other combinations of $\alpha, \beta, \gamma, \delta, \lambda$ we used in $H_1 - H_7$. The top of Figure 1 shows the famous value and weighting functions for CPT_{MED}, associated with diminishing sensitivity, loss aversion, risk aversion, overweighting of small probabilities and underweighting of large probabilities. All too often, CPT as a whole is identified with these specific curves. It is fairly hard to find examples of pictures, by anyone describing Cumulative Prospect Theory, other than the famous inverse S-shaped weighting functions, slightly more curved for gains than losses, associated with CPT_{MED}. An important core feature of this pair of graphs is that it embodies a conjunction of phenomena. When scholars or practitioners talk about CPT, they almost universally advertise this graph and thereby they advertise the co-occurrence of the phenomena that it captures jointly. And yet, empirical work pays little attention to issues of co-occurrence.

In our view, the broad preoccupation with CPT_{MED} constitutes a systemic scientific conjunction fallacy: Different parameter values generate a very broad variety of value and weighting functions. For large γ, δ, small probabilities are heavily overweighted in a broad range of probabilities, whereas, for small γ, δ, only a small range of small probabilities is overweighted. For α near 0 the value function quickly approaches 1, and for β near zero it quickly approaches λ. At the other extreme, for α and β near 1, value of money approximates an identity function for gains, respectively λ times identity for losses. Each of H_1-H_7 contradicts the popular stylized characterization of how 'people' make decisions according to CPT_{MED}, in at least some way. For instance, H_1 strongly underweights a much broader range of probabilities and it shows little loss aversion. H_2 is nearly calibrated in how it interprets probabilities and it shows no loss aversion. H_3 strongly underweights nearly all probabilities for losses, has huge diminishing sensitivity for gains (risk aversion), but not for losses and, furthermore, is extremely loss averse. H_4 shows nearly the opposite pattern of behavior from H_3. In contrast, H_5 weights probabilities for gains linearly, underweights a broad range of probabilities for losses, shows large diminishing sensitivity for gains, and no loss aversion in the range of X values we display. On the other hand, H_6 is nearly calibrated in how it weights all probabilities, whereas H_7 weights probabilities for gains linearly but underweights all probabilities for losses. Neither of the latter hypothetical cases show diminishing sensitivity for gains and both are loss averse. Several of these figures also make it very salient how β and λ work in tandem to determine whether or not, and for what range of X values, a person is loss averse. Every one of H_1-H_7, while satisfying CPT, is an exception to at least one of CPT's prominent qualitative features, yet, collectively, they perfectly match TK's median parameter values and median responses in Problems 1-6 of their loss aversion study. In sum, a collection of stylized features of a theory, when bundled together, may neither represent nor approximate even a single individual.

THE FOURFOLD PATTERN: Even understanding the joint occurrence of a collection of, say, four key qualitative features, is both a challenge and a valuable goal. Consider, for instance, the famous fourfold pattern of risk attitudes, namely "risk aversion for gains and risk seeking for losses of high probability; risk seeking for gains and risk aversion for losses of low probability" (Tversky and Kahneman, 1992, abstract). CPT and the fourfold pattern deserve a much higher level of nuance, both theoretically and empirically, than they are usually given (for some notable exceptions, see, e.g., Erev, Ert, Plonsky, Cohen, & Cohen, 2017; Murphy & ten Brincke, 2018).

Starting with theoretical nuance, as formulated in Equations 1-3, depending on the stimuli and the parameter values, CPT need not imply a fourfold pattern. Consider the stimuli that TK used in their own fourfold pattern study. Among their 8 Gain lotteries with low probability of winning, our hypothetical decision maker H_3 prefers the lottery over the sure amount in 3 stimuli (37.5% risk seeking); among their 17 Gain lotteries with high probability of winning, H_3 never prefers the lottery (0% risk seeking); among both their 8 Loss gambles with small probability of losing as well as their 17 Loss gambles with large probability of losing, H_3 prefers the lottery in every instance (100% risk seeking in both cases), a striking violation of the fourfold pattern. In a different paper (Regenwetter et al., n.d.), we have shown that Equations 1-3 permit almost any conceivable proportion of risk seeking choices in the 8 low-probability Gain prospects and the 8 low-probability Loss prospects that Tversky and Kahneman used in their own fourfold pattern study. Moving to empirical nuance, the percentages of risk seeking choices displayed by Tversky and Kahneman's Subject 9 (Tversky and Kahneman, 1992, p.308, Table 4) were 16%, 0%, 80% and 100%, respectively. The empirical data for this decision maker are much more similar to hypothetical decision maker H_3 than they are to a fourfold pattern. Not only does CPT not necessarily imply a fourfold pattern, Tversky and Kahneman's own data also show substantial individual differences in displayed choice behavior, with the empirical fourfold pattern only emerging as a stylized feature at a high level of aggregation. Claiming that 'people' display a fourfold pattern, in our view, constitutes another systemic scientific conjunction fallacy.

Ultimately, the stylized fourfold pattern is a huge abstraction across stimuli and people. We briefly consider the impact, just of individual differences alone, on the fourfold pattern and why even a modest goal in moving beyond stylized theory is a big goal. Regardless of how many people do or do not satisfy CPT, and taking the proportions in Tversky and Kahneman's (1992) Table 4 at face value, suppose that 88% of decision makers are 'consistently' risk averse for gains of high probability, 87% are 'consistently' risk seeking for losses of high probability, 78% are 'consistently' risk seeking for gains of low probability, and 80% are 'consistently' risk averse for losses of low probability. Under these assumptions, how many people would satisfy the heavily advertised fourfold pattern? Could we conclude that most people do? Under the above assumptions, anywhere between 33% and 78% of the population fall into that stylized mold, whereas anywhere between 22% and 67% are exceptions to the stylized feature of a fourfold pattern. If the four component events were mutually independent, the joint probability that a randomly sampled individual shows a fourfold pattern would be $.88 \times .87 \times .78 \times .80 = .48$. In that case, fewer than half of the population would satisfy the 'behavioral regularity' of a fourfold pattern. If each of the four stylized component properties of the fourfold pattern were to hold for 75% of the population, then anywhere between 0% and 75% of the population would satisfy the fourfold pattern. In this latter example, and in the 'best case scenario' that indeed 75% did, then a person would satisfy one of the four properties if and only if they satisfied the entire fourfold pattern. In that case, a quarter of the population would *violate all* four component properties of the fourfold pattern. In our view, it would be worthwhile to also study those two billion decision makers closely.

Keeping track of heterogeneity

How can behavioral decision research move beyond systemic scientific conjunction fallacies? Again, thinking about socio-economic indicators can provide guidance. The QuickFacts table of the U.S. census provides summary statistics of what are, in reality, jointly distributed random variables. For instance, race, gender, age, commuting time, household size, and income are, in actuality, jointly distributed. Understanding the socio-economic make-up of the U.S. population, beyond simple summary statistics, hinges on better understanding important features of the joint distribution of these indicators. To be successful, companies aim to understand some of the relevant co-occurrences and synergies of socio-economic features that characterize their customers and clients. Another useful case study to learn from is climate change. The rise in average temperatures, what is commonly referred to as "global warming," seems entirely innocuous by itself. On the surface, that stylized fact does not seem to have major implications for business strategy or public policy. Yet, the heterogeneity that goes with this stylized fact is far less innocuous: (Co-)occurrence of extreme weather of every imaginable kind. Different disasters, often occur even in close temporal or geographic proximity. Drought in the West coincides with flooding in the East, freezing in the South coincides with polar melting in the North. Average precipitation or average temperatures obstruct these phenomena. Dismissing such phenomena as "errors" or "noise" around an average would, in our view, be counter-productive. Indeed, Climate Science is somewhat successful in getting out the message: The outlook of ever more extreme and variable weather has led many industries and governments to take huge strategic initiatives.

In our view, the decision sciences deserve a similar broadening and unpacking of perspectives. Stylized facts about hypothetical constructs such as risk attitudes, attitudes to gains and losses, the use, non-use, or location of reference points, the reliance on compensatory strategies (like CPT) or noncompensatory strategies (such as some simple heuristics) while easy to understand, and while offering descriptive convenience in stylized form, may prevent managers and policy makers from seeing a fuller picture. As a case in point, consider the range of decision behaviors associated with the Covid-19 pandemic. Business leaders, politicians, and policy makers alike both display and face enormous differences and disagreements in risk attitudes, in reference point setting, selective attention, framing, and in the types of compensatory or heuristic decision strategies that they seem to either advertise or decry. The pandemic is a particularly striking example where profound individual differences in decision making, ranging from very simple heuristics to elaborate compensatory calculus, from extremely risk seeking to extremely risk averse, are on prominent display, are heavily discussed on a national scale, and have enormous societal impact. In our view, any pandemic policy response that presumed that 'people' essentially satisfy the conjunction of behavioral phenomena in Kahneman and Tversky (1979) or are essentially clones of CPT_{MED}, give or take some noise, would be entire hopeless and bound to fail (instead, see Erev, Plonsky, & Roth, 2020, for a study that models a heterogeneous population in which some people panic and others are complacent). This concern is not specific to (Cumulative) Prospect Theory. As we have furthermore seen, CPT can actually accommodate much heterogeneity. The concern applies most strongly to the broad range of theories that fail to treat individual differences as a theoretical primitive and that, instead, merely aim to characterize trends or central tendencies.

We are not at all the first to ring the alarm bells. There is already extensive work in the literature on CPT that aims to keep track of various types of heterogeneity. Broadly speaking, this work falls into two categories. One category operates at the level of behavioral phenomena (or mathematical axioms in the theory domain) where it keeps track of individual differences or other types of heterogeneity. The other category considers Equations 1-3 or variants thereof where it keeps track of heterogeneity typically in terms of parameters or functional forms.

The first category includes work on axiom testing that attends to individual differences (e.g., Birnbaum, 2004, 2007; Birnbaum & Bahra, 2007, 2012; Birnbaum & Navarrete, 1998; Birnbaum, Patton, & Lott, 1999; Brooks, Peters, & Zank, 2014; Brooks & Zank, 2005; Chung, von Winterfeldt, & Luce, 1994; Ho, Regenwetter, Niederée, & Heyer, 2005; Tsai & Böckenholt, 2006). Some work in this category also studies how a given behavioral phenomenon can vary across stimuli (e.g., Conlisk, 1989) and how behavior varies with experience (e.g., Erev et al., 2020).

The second category of approaches looks through a theoretical lens, and implicitly or explicitly treats the parameters of CPT as jointly distributed random variables. In this approach, some papers have classified participants by the shape of their utility and/or probability weighting functions (see, e.g., Abdellaoui, 2000; Bleichrodt & Pinto, 2000). Similarly, some papers carried out individual subject analyses of Equations 1-3 or variants thereof. They typically listed or enumerated the number of people who satisfied a given combination of properties, such as, say, how many had concave utility with overweighting of small probabilities, (e.g. Birnbaum & Chavez, 1997; Fan, Budescu, & Diecidue, 2019; Gonzalez & Wu, 1999; Stott, 2006; S.-W. Wu, Delgado, & Maloney, 2009). Another collection of works modeled heterogeneity with hierarchical Bayesian approaches and estimated joint distributions of parameters (for prominent examples, see Kellen, Pachur, & Hertwig, 2016; Murphy & ten Brincke, 2018; Nilsson, Rieskamp, & Wagenmakers, 2011; Pachur, Schulte-Mecklenberg, Murphy, & Hertwig, 2018; Scheibehenne & Pachur, 2015). Since different hierarchical Bayesian papers report different (posterior) distributions on the CPT parameters, future work still needs to figure out how to consolidate these. This category of work also attends to different functional forms of CPT (e.g., Blavatskyy & Pogrebna, 2010; Stott, 2006).

Beyond Cumulative Prospect Theory

Stylized characterizations of decision making behavior are not at all limited to discussions surrounding (Cumulative) Prospect Theory or even risky choice. Nor are we alone in raising concerns about the scientific conjunction fallacies that go hand in hand with stylized theory across the entire spectrum of behavioral decision research. In this section, we touch on a few examples that highlight the scope of the problem.

Birnbaum (2008) discussed numerous "paradoxes of risky choice," relying heavily (but not exclusively) on comparing modal choices with choices predicted by a CPT competitor model called TAX. For instance, Birnbaum's Table 1 considered modal choice on 10 different choice tasks. For every one of these problems separately, most people chose as predicted by "Prior TAX," namely 62%, 74%, 73%, 94%, 57%, 65%, 66%, 62%, 67%, and 67%. How plausible is "Prior TAX," which Birnbaum (2008) advertises as modeling the conjunction of those 10 features, as a model of an individual decision maker? Assuming that Birnbaum's percentages are completely accurate population percentages, how many people satisfy the "Prior TAX" preference pattern among these 10 problems? At best 57% do. How many people could be exceptions to that version of TAX? Everybody can be an exception to at least one prediction. It is even easy to imagine a population distribution in which everybody is an exception to three or more of these 10 phenomena. We do not mean to promote this idea, but if one were to assume, for simplicity, that the 10 phenomena were independent events, then the probability that a randomly sampled individual satisfied the combination of phenomena would be 0.021: Everyone but two percent of the population would violate "Prior TAX" in at least some way. Note that Birnbaum (2008) did not rely exclusively on patterns of modal choices and that the paper did emphasize individual differences. Furthermore, like the general CPT, the general TAX model has parameters that can vary. Yet, papers like Birnbaum (2008) could still do far more to explicitly warn readers of possible conjunction fallacies in the interpretation of their findings. In fact, the problem centrally affects the very design of Birnbaum's study: Just like we pointed out for

CPT_{MED} in the first half of the paper, we would have to emphasize that treating "Prior TAX" as a proxy for TAX has a conjunction fallacy baked in. More generally, research study "recipes" for generating interesting phenomena (usually to challenge a target theory), such as those underlying Birnbaum (2008) and Tversky and Kahneman (1992), are designed to provide evidence separately for separate phenomena, often with limited attention to the co-occurrence of such phenomena that would help prevent scientific conjunction fallacies.

Moving slightly beyond risky choice, consider the famous phenomenon of "ambiguity aversion," and the "Ellsberg's paradox" (Ellsberg, 1961), which is often depicted in stylized fashion (see, e.g., Camerer & Kunreuther, 1989). Some work has shown this phenomenon to be highly dependent on both individuals and stimuli (see, e.g., Budescu, Kuhn, Kramer, & Johnson, 2002; Kocher, Lahno, & Trautmann, 2018). The same applies for the Allais paradox (see, e.g., Conlisk, 1989). Similarly, context effects, such as the "compromise effect," the "similarity effect," and "asymmetric dominance," while well documented separately, appear to be rare within the same individual (Huber, Payne, & Puto, 2014; Huber & Puto, 1983; Liew, Howe, & Little, 2016; Spektor, Kellen, & Hotaling, 2018; Trueblood, Brown, & Heathcote, 2015).

Related to all of these, and similarly to Birnbaum (2008), a prominent study by Erev et al. (2017) considered 14 phenomena, including replications of prior studies regarding these phenomena. Again, such phenomena may vary extensively across people, stimuli, and other factors. In their Appendix E, Erev et al. (2017) reported how many individuals displayed each phenomenon, as well as the correlations among individual tendencies to exhibit those phenomena. Taking their methodology and stimuli at face value, and treating their estimates as completely accurate reflections of population proportions, 33% of the population shows their “Allais paradox,” 59% of people show their “reflection effect,” 66% show their “reversed reflection effect,” 54% show “overweighting of rare events,” 58% display “underweighting of rare events,” 67% are loss averse, 60% show their “magnitude effect,” 61% are risk averse, 63% are ambiguity averse, whereas 56%, 52%, 54%, 91%, respectively 74% show a series of five other phenomena. Trusting these as fully accurate population proportions, at most 33% of the population can show the conjunction of the 14 phenomena, since only 33% show Erev et al.’s version of the Allais paradox. At worst, nobody shows the conjunction of these phenomena. Many of Erev et al.’s (2017) correlations among the individual tendencies to display these phenomena were close to zero. We do not mean to promote this idea, but if one were to assume, for simplicity, that the 14 phenomena were independent events, then the probability that a randomly sampled individual satisfies the combination of phenomena would be 0.00066. By that calculus, hardly anyone is described by a theory that ‘people’ satisfy the conjunction of such phenomena.

Recall our recommendation to think conceptually of the CPT parameters as jointly distributed random variables, and our earlier review of papers that took such approaches. Similarly, a major point of Erev et al. (2017) was to model behavioral phenomena jointly, rather than separately (as did Kahneman & Tversky, 1979). They also specified joint distributions of parameters $(\sigma_i, \kappa_i, \beta_i, \theta_i, \gamma_i, \varphi_i)$ in their BEAST model (p. 387) so as to model and capture individual differences. Yet, in our view, in this and similar papers across behavioral decision research, there is still much room to attend to scientific conjunction fallacies and prevent them. For instance, even though Erev et al. (2017) reported both the proportions of, and correlations among, individuals who showed each of the 14 behavioral phenomena under consideration, that information was reported in an appendix. Like many other papers in behavioral decision research, the paper prominently listed a collection of phenomena of interest without explicitly warning readers that it is logically invalid to infer that they hold jointly in (many) individuals.

Behavioral decision research is not the only domain that has to grapple with this problem. In macro-economics, there is much ongoing work that evaluates the potential pitfalls of representative agent models, according to which the choices of many individuals can be summarized with a single 'representative decision-maker' (e.g., Benninga & Mayshar, 2000; Chan & Kogan, 2002; Hara, 2006; Hara, Huang, & Kuzmics, 2007; Harrison & Rutström, 2008; Kirman, 1992; Munshi, 2004; Wang, 1996). For instance, Hara et al. (2007) considered heterogeneous individuals with both convex or concave risk tolerances. They proved analytically that a representative agent's risk tolerance is nonetheless strictly convex. The authors emphasized how dramatically this limits the representative agent model: " . . . this implies that the representative consumer may well be very different from any individual in the economy, his utility function may not even be in the same class as every individual's" (p. 653). Using intertemporal choice as an example paradigm, and in a similar vein, Chen, Regenwetter, and Davis-Stober (2021) warned that econometric models such as logit and probit models may tell nothing about anyone's individual preference. These efforts to keep track of heterogeneity are an important step towards recognizing conjunction fallacies regarding economic behavior.

Marketing research has developed powerful approaches to capture systematic differences between groups of people. Most notably, segmentation methods partition a population of interest according to various characteristics, such as demographics (Smith, 1956). In the context of decision theories, segmentation methods may allow to keep track of systematic differences in decision making between subpopulations of various kinds (see also Böckenholt & Dillon, 1997, for an example). However, even there, some scholars (e.g., Johnson, 2006) have warned against overly stylized descriptions of behavior because segmentation analysis or classification of people by "types" may overlook important individual differences. Related to marketing, behavioral decision research has explored ways to capture individual differences by modeling decision characteristics as a function of age, culture, education, gender, or various (other) socio-economic features (see, e.g. Booij, van Praag, & van de Kuilen, 2010; Harrison & Rutström, 2008). The core idea of these paradigms is to provide a coarse summary by lumping together individuals who are similar on some key dimensions of decision-making. We would reiterate Johnson's (2006) warning that these approaches can also be overly stylized and we reinforce our earlier caution against conjunction fallacies here as well.

A final related phenomenon is Simpson's paradox (Blyth, 1972; Curley & Browne, 2001; Simpson, 1951) which has important implications in many disciplines. It refers to situations where a relationship between two variables observed at the level of individuals or subgroups, disappears, or reverses when considering the same data at the aggregate. For a review and tutorial see Kievit, Frankenhuis, Waldorp, and Borsboom (2013).

Solutions

Perhaps a major driver of stylized decision theory, even among quantitatively savvy scholars, is the intuitive fear that genuine accounts of heterogeneity would lead to unfalsifiable theories and data overfitting. A common heuristic belief is that theories with a large number of parameters can 'fit anything.' While such rules of thumb might be reasonable in some circumstances, parsimony need not equal frugality of parameters. There are widespread misunderstandings about what constitutes testable theories. These misconceptions form a major obstacle to advanced behavioral decision research with heterogeneity as a core primitive.

We can model commonalities and differences using axiomatizations of decision theories, combined with probabilistic mixtures. Specifically, starting from a collection of behavioral properties, one can study mathematically what it means for everybody to satisfy their conjunction (see also Davis-Stober & Regenwetter, 2019, for related discussions). For instance, the combination of the three axioms called "transitivity," "completeness," and "asymmetry," in conjunction, define *strict linear orders*. Consider a master set of 10 choice options. A *binary relation* is a set of ordered pairs of elements of the master set. There are $2^{10\times 9} = 1.2 \times 10^{27}$ distinct binary relations. Of these $\sim 1,200,000,000,000,000,000,000,000,000$ relations, just $10! = 3,628,800$, are strict linear orders. At the same time, if there are 3.6 million different ways to satisfy the conjunction of "transitivity," "asymmetry," and "completeness," for 10 choice alternatives, then there is potential for a huge amount of heterogeneity, even when everyone is transitive. On the surface, despite the reduction by 21 orders of magnitude above, this may seem unparsimonious, maybe even vacuous. To understand the parsimony of strict linear orders, it is useful to note that this heterogeneity is dwarfed by the amount of heterogeneity among preferences that violate "transitivity," while also jointly satisfying "asymmetry" and "completeness." For 10 options, there are $2^{\binom{10}{2}} - 3,628,800 = 3.5 \times 10^{13}$ such preferences, seven orders of magnitude more than strict linear orders. A probability distribution over 3,628,800 strict linear orders has 3,628,799 degrees of freedom. If one presumes/forces "asymmetry" and "completeness" in observed choices via a two-alternatives forced-choice paradigm, the 45 binary choice probabilities induced by all possible distributions over 3.6 million strict linear orders form an extremely parsimonious convex polytope that only occupies 1.15 in 10^{10} of the space of all such probabilities (Regenwetter, Davis-Stober, Smeulders, Fields, & Wang, 2021). At the same time, 45 Bernoullis/Binomials generate 45 degrees of freedom in the data, compared to 3.6 million degrees of freedom in the model. Yet, $\frac{999{,}999{,}999}{10{,}000{,}000{,}000}$ of all possible parameter values of these 45 Bernoullis/Binomials are incompatible with *any* probability mixture of 3.6 million strict linear orders. The conceptual, mathematical, methodological, and statistical questions associated with this sort of approach are manifold. We do not discuss them here (but see, e.g., Birnbaum, 2011; Davis-Stober, 2009; Heck & Davis-Stober, 2019; Hertwig & Pleskac, 2018; Regenwetter, 2020; Regenwetter & Cavagnaro, 2019; Regenwetter, Dana, Davis-Stober, & Guo, 2011; Regenwetter & Davis-Stober, 2017; Regenwetter et al., 2021, 2014; Regenwetter & Robinson, 2017, 2019; Smeulders, Davis-Stober, Regenwetter, & Spieksma, 2017; Zwilling et al., 2019, for pointers and discussions).

Over the past decade, some progress has been made with this type of approach. The most extensive paradigm has been the study of mixtures of strict linear orders (Birnbaum et al., 1999; Brown, Davis-Stober, & Regenwetter, 2015; Cavagnaro & Davis-Stober, 2014; Dai, 2017; Davis-Stober, Park, Brown, & Regenwetter, 2016; Fiorini, 2004; McCausland, Davis-Stober, Marley, Park, & Brown, 2020; Regenwetter et al., 2018; Regenwetter, Dana, & Davis-Stober, 2010, 2011; Regenwetter, Dana, Davis-Stober, & Guo, 2011). Similarly, Regenwetter and Davis-Stober (2012) considered all possible conjunctions of the axioms of "negative transitivity" and "asymmetry" on five choice alternatives, permitting mixtures of the resulting *strict weak orders* within individuals, as well as comparing individuals with each other. Similarly, individual preference states can also be heuristics in a heuristic toolbox.

Besides considering all possible conjunctions for a set of axioms and all possible probability distributions over those preference patterns, one can instead take a theory like CPT, consider the preference patterns associated with different combinations of parameter values[7], and consider all possible probability distributions over those preference patterns. Guo and Regenwetter (2014) took this approach for the "PRAM" theory of Loomes (2010).

[7] Guo and Regenwetter (2014), Regenwetter and Robinson (2017), Regenwetter et al. (2014), as well as Zwilling et al. (2019) approximated the collection of possible preference patterns in a given theory by running a discrete and fine-grained grid-search of the theory's parameter values.

Regenwetter et al. (2014) and Zwilling et al. (2019) did the same for two different versions of Cumulative Prospect Theory. Likewise, Regenwetter and Robinson (2017) employed the same approach to model mixtures of the preference patterns associated with various models related to CPT. These *random preference models* treat binary choice probabilities as marginals of probability distributions over preference patterns. For a related approach, see Birnbaum's extensive research program on "true and error models" (e.g., Birnbaum & Bahra, 2007, 2012; Birnbaum & Quispe-Torreblanca, 2018). For a related recent (parametric) idea, see Bhatia and Loomes (2017), who highlighted the importance of distinguishing between, what they labeled "preference noise" and "response noise." Similarly, hierarchical Bayesian models like the ones we reviewed for CPT, which typically model individual differences and response noise jointly through a hierarchy of parametric distributions, can be applied to a broad range of theories and research paradigms (see, e.g., Lee, 2006; Lee & Newell, 2011; van Ravenzwaaij, Moore, Lee, & Newell, 2014).

There are also situations, in which modal choice is informative about individuals. An obvious case are data from a single individual who provides multiple observations for various pairwise choices who has a single underlying preference pattern and makes probabilistic errors in responses, e.g., via a "weak utility model" (Block & Marschak, 1960; Luce & Suppes, 1965). For data pooled across people, modal choice is informative when the individuals in question share a common latent (unanimous) preference, such as a collective weak utility model. For more details, see also the tutorials and reviews of Regenwetter and Davis-Stober (2017); Regenwetter et al. (2021, 2014); Zwilling et al. (2019).

Conclusion and Discussion

SCIENTIFIC CONJUNCTION FALLACIES IN DECISION RESEARCH: This paper points out the elephant in the room. As Kirman (1992) put it rather poetically, three decades ago, "…it is clear that the 'representative' agent deserves a decent burial, as an approach to economics analysis that is not only primitive, but fundamentally erroneous" (p.119). There is no doubt that both TK and many scholars in the field are aware of individual differences. However, in our view, the field vastly underestimates the qualitative nature of behavioral heterogeneity. While we certainly agree that papers need to summarize their findings, we also believe that authors should throttle back on abstract summary statistics. They should avoid over-interpreting trends in their data. They should include explicit disclaimers about the limitations of their summary statistics. Scholars should systematically cross-check their own conclusions for scientific conjunction fallacies, they should avoid language that suggests or promotes such logic errors, and they should explicitly warn readers of potential fallacies of sweeping generalization and of fallacies of composition (Regenwetter & Robinson, 2017). As many have advocated before us, scholars should steer clear of statistics that readers readily mis- or over-interpret.

For instance, Tversky and Kahneman should have, at the very least, added inter-quartile ranges or box-and-whisker plots to their CPT_{MED} median parameter estimates. The preoccupation with stylized aggregate features, and how 'people' choose, does a disservice to applied decision science, management science, and policy making, as well as to numerous other disciplines across basic and applied science. When we summarize findings with aggregated measures or aggregated parameter values, when we treat variability as mere noise, we perpetuate profound misconceptions of the genuine heterogeneity in behavior. These common practices also misrepresent (Cumulative) Prospect Theory's and other theories' genuine scope. Our examples show how strong these differences can be, even among those people who perfectly satisfy (Cumulative) Prospect Theory. This insight has important implications for researchers, managers, and policy makers. No business would bet huge stakes on treating everyone as an approximate clone of a middle-aged caucasian woman living in a median-sized household with a median income. In most of the world, very few real-estate developers can build homes, offices, or businesses that require weather conditions to remain right at annual mean all year long. The automotive industry designs vehicles to be resilient to a broad range of driving conditions, even without measuring the joint distributions of all relevant parameters. Agronomists intently study the sensitivity of their crops to deviations from average climate conditions.

In the same vein, even taking Kahneman and Tversky (1979) and Tversky and Kahneman (1992) at face value, one should not bet on finding large numbers of people consistently matching the stylized mold of (Cumulative) Prospect Theory along which the two theories are routinely advertised by both proponents and opponents. We have also seen that similar challenges with logical conjunctions apply across a broad range of behavioral decision research paradigms. Managers and policy makers should scrutinize the sensitivity of business plans and policies to violations of stylized decision theory. They should invest intellectual capacity into enhancing the resilience of management and policy solutions to violations of stylized theory. If behavioral decision theory is to make itself useful for applied decision science, then it must graduate from over-simplifications. Any theory of decision making, CPT or other, if applied too narrowly, may end up properly representing or serving *nobody. Everybody* may well be an exception to either an entire theory (e.g., violate Equations 1-3 altogether), or, while satisfying the theory, be an exception to some of the advertised stylized properties when the latter are not actually universally implied by the theory in question. In that sense, even taking Kahneman and Tversky's own (1979, 1992) seminal findings fully at face value, *everyone* may well be an exception to (Cumulative) Prospect Theory. The same may apply to the reader's own personal favorite decision theory, whatever that might be.

Cost-Benefit Analysis: The difficulties, costs, and enormous benefits of measuring the joint distribution of socio-economic measures are the reason why countries like the United States run a national census and other large-scale studies. The importance of understanding climate change has led many governments and businesses around the globe to invest huge amounts of resources into climate science, climate resilience, and climate preparedness. Like economics and climate, decision making affects virtually every aspect of human activity. While it is undoubtedly difficult and expensive to model, measure, and understand heterogeneity of decision making across mankind, decision science should invest intellectual capital into at least five key areas: 1) Complement and enhance business strategies and government policies with careful sensitivity analyses to gauge their susceptibility to violations of stylized theory. 2) Based on lessons learned in sensitivity analyses, develop ways to render business strategies and government policies resilient to violations of stylized theory. 3) Develop strategies to move beyond stylized theory, e.g., investigate much more deliberately and vigorously when and where which decision theoretic constructs apply and how they vary, co-occur, and co-vary across 7 billion people. Studying the co-occurrence of heuristics, biases, or reasoning errors is hugely important and useful. For instance, a referee has pointed out that a leading cause of accidents is the joint occurrence, hence compounding, of multiple human errors. 4) Even if this is difficult or seems redundant, scholars should warn practitioners against conjunction fallacies and communicate scientific information using language that eschews such fallacies. 5) Even though this encounters hurdles with curriculum design, business school, public policy school, and other faculty should thoroughly train their students to recognize and combat conjunction fallacies. In particular, they should carefully avoid these fallacies in their own decision classes by steering clear of stylized depictions about how 'people' behave.

Psychology has embarked on a number of remarkable initiatives to make the discipline more relevant and to enhance the external validity of its claims. Many of these are either anti-fraud or methodological innovations, such as pre-registration, meta-analysis, advanced stimulus and study design, replication, prediction tournaments, and others. While these efforts bring great benefits, they can also carry very serious potential costs: Unfortunately, none of these initiatives automatically guards against logical errors in scientific reasoning, such as the systemic conjunction fallacies that plague behavioral decision research. On the contrary, nearly all of these efforts have the inherent potential to repeat, reinforce, and perpetuate conjunction fallacies about how ‘people’ make decisions.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, *46*, 1497-1512.

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science*, *57*, 1879-1895.

Ariely, D. (2010). *Predictably irrational: The hidden forces that shape our decisions.* Harper Collins.

Ariely, D., Huber, J., & Wertenbroch, K. (2005). When do losses loom larger than gains? *Journal of Marketing Research*, *XLII*, 134-138.

Barberis, N. (2013). The psychology of tail events: Progress and challenges. *American Economic Review*, *103*, 611-616.

Benartzi, S., & Thaler, R. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, *110*, 73-92.

Benninga, S., & Mayshar, J. (2000). Heterogeneity and option pricing. *Review of Derivatives Research*, *4*(1), 7–27.

Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review*, *124*, 678-687.

Birnbaum, M. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes*, *95*, 40–65.

Birnbaum, M. (2007). Tests of branch splitting and branch-splitting independence in Allais paradoxes with positive and mixed consequences. *Organizational Behavior and Human Decision Processes*, *102*, 154-173.

Birnbaum, M. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501.

Birnbaum, M. (2011). Testing mixture models of transitive preference. Comment on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review*, *118*, 675–683.

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

Birnbaum, M., & Bahra, J. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making*, *7*, 402–426.

Birnbaum, M., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, *71*, 161-194.

Birnbaum, M., & Navarrete, J. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty*, *17*, 49–78.

Birnbaum, M., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes*, *77*, 44-83.

Birnbaum, M., & Quispe-Torreblanca, E. (2018). TEMAP2.R: True and error model analysis program in R. *Judgment and Decision Making*, *13*, 428-440.

Blavatskyy, P., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, *25*, 963–986.

Bleichrodt, H., & Pinto, J. (2000). A parameter-free elicitation of the probability weighting fundation in medical decision analysis. *Management Science*, *46*, 1485-1496.

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (p. 97-132). Stanford: Stanford University Press.

Blyth, C. R. (1972). On simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, *67*(338), 364-366.

Böckenholt, U., & Dillon, W. (1997). Some new methods for an old problem: Modeling preference changes and competitive market structures in pretest market data. *Journal of Marketing Research*, *34*, 130-142.

Booij, A., van Praag, B., & van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, *68*, 115-148.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The Priority Heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409-432.

Brooks, P., Peters, S., & Zank, H. (2014). Risk behavior for gain, loss, and mixed prospects. *Theory and Decision*, *77*, 153–182.

Brooks, P., & Zank, H. (2005). Loss averse behavior. *Journal of Risk and UncertaintyUncertainty*, *31*, 301–325.

Brown, N., Davis-Stober, C., & Regenwetter, M. (2015). Commentary: "Neural signatures of intransitive preferences". *Frontiers in Human Neuroscience*.

Budescu, D., Kuhn, K., Kramer, K., & Johnson, T. (2002). Modeling certatiny equivalents for imprecise gambles. *Organizational Behavior and Human Decision Processes*, *88*, 748-768.

Camerer, C., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearlity in probability. *Journal of Risk and Uncertainty*, *8*, 167-196.

Camerer, C., & Kunreuther, H. (1989). Experimental markets for insurance. *Journal of Risk and Uncertainty*, *2*, 265-299.

Cavagnaro, D. R., & Davis-Stober, C. P. (2014). Transitive in our preferences, but transitive in different ways: An analysis of choice variability. *Decision*, *1*(2), 102.

Chan, Y. L., & Kogan, L. (2002). Catching up with the joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy*, *110*(6), 1255–1285.

Chen, M., Regenwetter, M., & Davis-Stober, C. (2021). Collective choice may tell nothing about anyone's individual preferences. *Decision Analysis*, *18*, 1-24.

Chung, N.-K., von Winterfeldt, D., & Luce, R. (1994). An experimental test of event commutativity in decision making under uncertainty. *Psychological Science*, *5*, 394-400.

Conlisk, J. (1989). Three variants of the allais example. *The American Economic Review*, *79*, 392-407.

Curley, S., & Browne, G. (2001). Normative and descriptive analyses of simpson's paradox in decision making. *Organizational Behavior and Human Decision Processes*, *84*, 308-333.

Dai, J. (2017). Are intertemporal preferences transitive? A Bayesian analysis of repeated individual intertemporal choices. *Decision*, *4*(1), 1-24.

Davis-Stober, C. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology*, *53*, 1-13.

Davis-Stober, C., Park, S., Brown, N., & Regenwetter, M. (2016). Reported violations of rationality may be aggregation artifacts. *Proceedings of the National Academy of Sciences of the United States of America*.

Davis-Stober, C., & Regenwetter, M. (2019). The 'paradox' of converging evidence. *Psychological Review*, *126*, 865-879.

de Langhe, B., & Puntoni, S. (2015). Bang for the buck: Gain-loss ratio as a driver of judgment and choice. *Management Science*, *61*, 1137–1163.

Diederich, A., & Trueblood, J. (2018). A dynamic dual process model of risky decision making. *Psychological Review*, *125*, 270-292.

Ellsberg, D. (1961). Risk, ambiguity, and the Savave axioms. *The Quarterly Journal of Economics*, *75*, 643-669.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review*, *124*, 369-409.

Erev, I., Plonsky, O., & Roth, Y. (2020). Complacency, panic, and the value of gentle rules enforcement in addressing pandemics. *Nature Human Behavior*, *4*, 1095-1097.

Fan, Y., Budescu, D., & Diecidue, E. (2019). Decisions with compound lotteries. *Decision*, *6*, 109-133.

Feess, E., Müller, H., & Schumacher, C. (2016). Estimating risk preferences of bettors with different bet sizes. *European Journal of Operational Research*, *249*, 1102-1112.

Fennema, H., & Van Assen, M. (1998). Measuring the utility of losses by means of the tradeoff method. *Journal of Risk and Uncertainty*, *17*, 277-295.

Fiorini, S. (2004). A short proof of a theorem of Falmagne. *Journal of Mathematical Psychology*, *48*, 80-82.

Gehring, W., & Willoughby, A. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*, 2279-2282.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*, 129-166.

Grolleau, G., Kocher, M., & Sutan, A. (2016). Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science*, *62*, 3428-3438.

Guo, Y., & Regenwetter, M. (2014). Quantitative tests of the Preceived Relative Argument Model: Comment on Loomes (2010). *Psychological Review*, *121*, 696-705.

Hara, C. (2006). Heterogeneous risk attitudes in a continuous-time model. *The Japanese Economic Review*, *57*, 377–405.

Hara, C., Huang, J., & Kuzmics, C. (2007). Representative consumer's risk aversion and efficient risk-sharing rules. *Journal of Economic Theory*, *137*, 652–672.

Harinck, F., Van Dijk, E., Van Beest, I., & Mersmann, P. (2007). When gains loom larger than losses: Reversed loss aversion for small amounts of money. *Psychological Science*, *18*, 1099-1105.

Harrison, G. W., & Rutström, E. E. (2008). Expected utility theory and prospect theory: one wedding and a decent funeral. *Experimental Economics*, *12*, 133-158.

Heck, D., & Davis-Stober, C. (2019). Multinomial models with linear inequality constraints: Overview and improvements of computational methods for Bayesian inference. *Journal of Mathematical Psychology*, *91*, 70 - 87.

Henderson, V. (2012). Prospect theory, liquidation, and the disposition effect. *Management Science*, *58*, 445-460.

Hertwig, R., & Pleskac, T. (2018). The construct–behavior gap and the description–experience gap: Comment on Regenwetter and Robinson (2017). *Psychological Review*, *125*, 844-849.

Ho, M.-H., Regenwetter, M., Niederée, R., & Heyer, D. (2005). An alternative perspective on Von Winterfeldt et al.'s (1997) test of consequence monontonicity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 365-373.

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect.

Journal of Marketing Research, *51*, 520–525.

Huber, J., & Puto, C. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research*, 31–44.

Johnson, E. (2006). Things that go bump in the mind: How Behavioral Economics could invigorate Marketing. *Journal of Marketing Research*, *43*, 337-340.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Strauss and Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition*, *157*, 126-138.

Kermer, D., Driver-Linn, E., Wilson, T., & Gilbert, D. (2006). Loss aversion is an affective forecasting error. *Psychological Science*, *17*, 649-653.

Kievit, R., Frankenhuis, W., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*, 513.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives*, *6*, 117–136.

Kocher, M., Lahno, A., & Trautmann, S. (2018). Ambiguity aversion is not universal. *European Economic Review*, *101*, 268-283.

Lee, M. (2006). A hierarchical Bayesian model of human decision making on an optimcal stopping problem. *Cognitive Science*, *30*(1-26).

Lee, M., & Newell, B. R. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judgment and Decision Making*, *6*, 832–842.

Levy, H. (2016). Aging population, retirement, and risk taking. *Management Science*, *62*, 1415-1430.

Liew, S. X., Howe, P., & Little, D. (2016, 03). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, *23*.

Loomes, G. (2010). Modeling choice and valuation in decision experiments. *Psychological Review*, *117*, 902-924.

Luce, R. D., & Suppes, P. (1965). Preference, utility and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. III, p. 249-410). New York: Wiley.

McCausland, W., Davis-Stober, C., Marley, A., Park, S., & Brown, N. (2020). Testing the random utility hypothesis directly. *The Economic Journal*, *130*, 183-307.

McGraw, A., Larsen, J., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science*, *21*, 1438-1445.

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the Indian green revolution. *Journal of Development Economics*, *73*, 185–213.

Murphy, R., & ten Brincke, R. (2018). Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates. *Management Science*, *64*, 308-326.

Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for Cumulative Prospect theory. *Journal of Mathematical Psychology*, *55*, 84 - 93.

Pachur, T., Schulte-Mecklenberg, M., Murphy, R., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General*, *147*, 147-169.

Regenwetter, M. (2020). "With sufficient increases in x, more people will engage in the target behavior". *Journal of Mathematical Psychology*, *99*, 102457.

Regenwetter, M., & Cavagnaro, D. (2019). Tutorial on removing the shackles of regression analysis: How to stay true to your theory of binary response probabilities. *Psychological Methods*, *24*, 135-152.

Regenwetter, M., Cavagnaro, D., Popova, A., Guo, Y., Zwilling, C., Lim, S., & Stevens, J. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision*, *5*, 63-94.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Quantitative Psychology and Measurement*.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, *118*, 42-56.

Regenwetter, M., Dana, J., Davis-Stober, C. P., & Guo, Y. (2011). Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). *Psychological Review*, *118*, 684-688.

Regenwetter, M., & Davis-Stober, C. (2017). The role of independence and stationarity in probabilistic models of binary choice: A summary and review. *Journal of Behavioral Decision Making*, *31*, 100-114.

Regenwetter, M., Davis-Stober, C., Smeulders, B., Fields, B., & Wang, C. (2021). (Ir)rationality of animal choice? A guide to testing transitivity. *Quarterly Journal of Biology*. (Manuscript in press)

Regenwetter, M., & Davis-Stober, C. P. (2012). Behavioral variability of choices versus structural inconsistency of preferences. *Psychological Review*, *119*, 408-416.

Regenwetter, M., Davis-Stober, C. P., Lim, S. H., Guo, Y., Popova, A., Zwilling, C., ... Messner, W. (2014). QTest: Quantitative testing of theories of binary choice. *Decision*, *1*, 2-34.

Regenwetter, M., & Robinson, M. (2017). The construct-behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review*, *124*, 533-550.

Regenwetter, M., & Robinson, M. (2019). The construct-behavior gap revisited: Reply to Hertwig and Pleskac (2018). *Psychological Review*, *126*, 451-454.

Regenwetter, M., Robinson, M., & Wang, C. (n.d.). *Four internal inconsistencies in Tversky and Kahneman's (1992) Cumulative Prospect Theory paper: A case study in ambiguous theoretical scope.* (Unpublished Manuscript)

Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to access the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, *22*, 391-407.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *13*(2), 238-241.

Smeulders, B., Davis-Stober, C., Regenwetter, M., & Spieksma, F. (2017). Testing probabilistic models of choice using column generation. *Computers & Operations Research*, *95*, 32-43.

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, *21*, 3–8.

Spektor, M., Kellen, D., & Hotaling, J. (2018). When the good looks bad: An experimental exploration of the repulsion effect. *Psychological Science*, *29*(8), 1309-1320.

Stott, H. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*, 101-130.

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Timmer, C. P. (2012). Behavioral dimensions of food security. *Proceedings of the National Academy of Sciences of the USA*, *109*, 12315–12320.

Tom, S., Fox, C., Trepel, C., & Poldrack, R. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515-518.

Trepel, C., Fox, C., & Poldrack, R. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research*, *23*, 34-50.

Trueblood, J., Brown, S., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to tsetsos, chater, and usher (2015). *Psychological*

Review, *122*, 848–853.

Tsai, R.-C., & Böckenholt, U. (2006). Modelling intransitive preferences: A random-effects approach. *Journal of Mathematical Psychology*, *50*, 1-14.

Tversky, A., & Kahneman, D. (1983). Extensionsional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Usher, M., & McClelland, J. (2004). Loss aversion and inhibition in dynamical models of multiattribute choice. *Psychological Review*, *111*, 757-769.

van Ravenzwaaij, D., Moore, C. P., Lee, M. D., & Newell, B. R. (2014). A hierarchical Bayesian modeling approach to searching and stopping in multi-attribute judgment. *Cognitive Science*, *38*, 1384–1405.

Wang, J. (1996). The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics*, *41*, 75–110.

Weingarten, E., Bhatia, S., & Mellers, B. A. (2019). Multiple goals as reference points: One failure makes everything else feel worse. *Management Science*, *65*, 3337-3352.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, *42*, 1676-1690.

Wu, S.-W., Delgado, M., & Maloney, L. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences of the USA*, *106*, 6088-6093.

Zwilling, C., Cavagnaro, D., Regenwetter, M., Lim, S., Fields, B., & Zhang, Y. (2019). QTest 2.1: Quantitative testing of theories of binary choice using Bayesian inference. *Journal of Mathematical Psychology*, *91*, 176-194.