HARVARD | BUSINESS | SCHOOL

DOCTORAL PROGRAMS |

DISSERTATION ACCEPTANCE CERTIFICATE

(To be placed in First Library Copy)

The undersigned, approved by the Chair of the Doctoral Programs, have examined a dissertation entitled

The Continuum of Choice: Essays on How Consumer Decisions are Made, Changed, and Perceived

presented by **Katherine Neva Barasz**

candidate for the degree of Doctor of Business Administration and hereby certify that it is worthy of acceptance.

Signature .....
Michael I. Norton, *Chair*

Signature .....
Leslie K. John

Signature .....
John T. Gourville

Date 3/28/16

WYSS | SOLDIERS FIELD | BOSTON, MA 02163 | Ph 617.495.6101 | Fx 617.496.2203 | doctoralprograms@hbs.edu | GEORGE F. BAKER FOUNDATION

The Continuum of Choice:

Essays on How Consumer Decisions are Made, Changed, and Perceived

A dissertation presented

by

Katherine N. Barasz

to

The Marketing Unit at Harvard Business School

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Marketing

Harvard University

Cambridge, Massachusetts

March 2016

Dissertation Advisor: Professor Michael Norton Katherine N. Barasz

The Continuum of Choice:

Essays on How Consumer Decisions are Made, Changed, and Perceived

ABSTRACT

This research investigates the *continuum of choice*—unseen, unanticipated causes and consequences of consumer decisions. Three essays investigate hidden factors that influence the choices we make, subtle ways to affect choice at the moment of execution, and the overlooked signals that our choices convey (correctly or incorrectly) about us to others. Essay one investigates the perverse tendency to hope for the worst: when faced with a difficult decision (e.g., whether or not to have surgery), people can paradoxically feel subjectively *better* with—and even actively prefer—objectively *worse* but certain news (e.g., "95% chance of a disease") over objectively *better* but more uncertain news (e.g., "50% chance of a disease"). This, in turn, has the potential to meaningfully change people's subsequent choices and preferences in unexpected ways. Essay two examines a subtle intervention to change people's decisions about engagement levels: arbitrarily grouping discrete tasks or items together as part of an apparent

"set" motivates people to reach perceived completion points—or finish a *pseudo-set*—even in the absence of extrinsic incentives. Essay three explores the judgments people make after observing others' choices; specifically, upon learning of someone's choice of one option, people erroneously believe that person must dislike dissimilar options, leading to a pervasive and systematic prediction error.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Without a doubt, I wouldn't have survived long enough to write a dissertation—let alone need an acknowledgements section—were it not for the invaluable advice, support, input, and friendship of so many people over so many years. This has been a group effort, and there are many people to whom I am deeply indebted and forever grateful:

My committee, Mike, Leslie, and John, for being all-around good and inspiring humans, for asking the right questions and pushing me to think harder, for helping me gain the confidence to stand on my own two feet, and for enduring way more than their fair share of pie charts.

My HBS colleagues—or more fittingly, dear friends—without whom I would have smiled much less often, including (but certainly not limited to): GiNorton/N.E.R.D.lab members; the DBA cohort of 2010; Marketing faculty and students, past and present; and two people who put up with me in exceptionally close quarters—my former officemate and favorite Italian, Silvia, and my current officemate and whiteboarding partner-in-crime, Tami.

My husband, Aaron, who has become in-house counsel advising me (with much-appreciated candor) on the attractiveness (or not) of new research ideas, and who has propped me back up and cheered me on countless times as I've approached this finish line.

My two vivacious children, Owen and Eliza, who accompanied me (in utero) to more than half of my doctoral courses, and without whom I would've gotten much more sleep, devoted many more hours to research, and lacked precious perspective on my life.

And last, but most certainly not least, my family and friends, none of whom vocalized their concerns when I quit my job six years ago to pursue a PhD they didn't know was a thing, and all of whom have smiled and nodded with convincing enthusiasm as I've tried to explain what exactly it is that I do.

INTRODUCTION

The course of any single day can be mapped out by the sequence of choices we have made—choices about what to do with our time, where to go and when, with whom to interact, what to consume and what to expend. Some choices we consider carefully (e.g., choices around career and family, politics and religion), others we enact mindlessly (e.g., choices of clothing and food, magazines and music), but ranking just behind eating and breathing, our most fundamental task as humans is making choices.

Choices could reasonably be viewed as discrete moments in time—the instant of execution when we said yes, or turned left, or pressed "send," or selected answer C; but as behavioral research has pushed us to recognize, we can also view our choices as more integrated, complex processes, both the culmination of inputs and commencement of effects. As such, any single choice is comprised not only of the completed act of decision, but also of its upstream antecedents (e.g., the heuristics and decision rules that guided the choice) and its downstream consequences (e.g., the direct and indirect repercussions of a choice).

In my dissertation, I broadly investigate this *continuum of choice*, examining factors that influence the choices we make, ways to affect choice at the moment of execution, and signals our choices convey about us (both correctly or incorrectly) to others. As such, my research is broadly situated within three streams: (1) Antecedents of Choice, (2) Choice Interventions, and (3) Consequences of Choice. Below are three essays, each corresponding to one of these three distinct streams, but linked by the common objective of uncovering consequential insights into the way consumer choices are made, changed, and perceived.

Hoping for the Worst: When and Why People Prefer Bad News

KATE BARASZ

SERENA HAGERTY

Abstract

Three studies investigate when and why people prefer bad news. When facing a consequential decision (e.g., whether to have surgery), people report greater positive affect after receiving worse but relatively certain news (e.g., "you have a 95% chance of appendicitis") than better but relatively uncertain news (e.g., "you have a 50% chance of appendicitis," Study 1). People even seek out bad news, preferring an objectively worse injury that definitely requires invasive treatment over an objectively less severe injury where the right treatment decision is less certain (Study 2). Such bad news delivers greater certainty by reducing decisional conflict, in turn increasing positive affect (Study 3). Although worse news often begets worse objective outcomes, people may paradoxically hope for the worst.

Imagine you are suffering from what could be appendicitis; the only treatment is surgery to remove your appendix. If not appendicitis, it is just a false alarm. Your doctor performs tests to help determine the diagnosis, and based on the results, you must decide whether or not to undergo surgery. What news do you hope to hear: that your chances of having appendicitis are low—or high?

It seems reasonable to assume most people would prefer better news—that the chance of appendicitis is low, and correspondingly, chances of false alarm are high. However, when asked which diagnosis they would rather receive—a 50% ("better") or 95% ("worse") chance of appendicitis—most participants ($N = 95$) had the opposite intuition: 78% said they would actually *prefer* a 95% chance of appendicitis, deliberately choosing a worse diagnosis over a better one (results in Supplemental Material). Why would people prefer—even seek out—bad news? This research investigates when and why people paradoxically hope for the worst. We suggest that while negative news may beget negative outcomes, it can have the advantage of minimizing uncertainty around a consequential decision, conferring subjective emotional benefits that may outweigh objectively worse news.

Much research demonstrates that people desire certainty (e.g., Camerer & Weber, 1992; Curley, Yates, & Abrams, 1986; Ellsberg, 1961; Frisch & Barron, 1988; Fox & Tversky, 1995). For instance, in their work on the *uncertainty effect*, Gneezy, List, and Wu (2006) demonstrated the monetary premium of certainty: people may value worse-but-certain options (e.g., 100% chance of receiving a $50 gift card) *higher* than better-but-more-uncertain ones (e.g., lottery for a 50/50 chance of a $50 or $100 gift card). Simonsohn (2009) found evidence that this surprising valuation is driven by a "literal distaste for uncertainty"—that uncertainty, in and of itself, is aversive.

Uncertainty can be especially aversive when there is a choice to be made. Even mundane decisions can become subjectively difficult when options possess uncertain trade-offs (e.g., Dhar 1997), but already-difficult decisions can be made unbearable (e.g., Botti, Orfali, & Iyengar, 2009; Kahn & Luce, 2003; Sah, Elias, & Ariely, 2013; Schwartz, 2004). As a result, people may adopt alternate coping and processing strategies (Luce, Bettman, and Payne, 1997) or avoid decisions altogether (Dhar, 1997; Luce, 1998; Tversky & Shafir, 1992). Therefore, motivation to reduce uncertainty may be strongest when people face non-trivial, consequential decisions.

We argue that these two dynamics—people's general aversion to uncertainty and the burden of uncertainty on decision-makers—can lead people to hope for the worst: that faced with a tradeoff between the negativity of uncertainty and the negativity of news, people may choose to reduce the former more than the latter. We propose that when facing a consequential decision (e.g., whether to have surgery), certainty—even certainty of a bad outcome—can ease decision-making and increase positive affect. Consequently, people may ironically feel relatively better with—and even actively prefer to receive—relatively worse news.

In documenting a novel risk preference, our findings contribute broadly to research on decisions under uncertainty and uncertainty aversion. We also contribute to research on the uncertainty effect specifically (Gneezy et al., 2006; Simonsohn, 2009). While Gneezy et al. (2006) noted their effect was detectable only between-subjects (participants provided "correct" valuations when options were presented concurrently), we show that similar results can emerge within-subjects: when facing subjectively difficult decisions, people may prefer worse-but-certain options to better-but-more-uncertain ones—even when both are simultaneously available. Further, we investigate uncertainty aversion in the domain of losses; whereas Gneezy et al.

(2006) use lotteries with only upside, we use situations in which people stand a greater chance of physical injury or harm when worse options are preferred.

In addition, we contribute to a richer understanding of the decision making process for consequential choices. Previous research has identified the role of factors like choice autonomy (Botti et al. 2009) and trade-offs among important goals (Luce, 1998; Luce, 2005; Luce et al., 1997); we add an understanding of the *type* of information—better or worse news—that can help or hinder the making of consequential decisions.

Three studies investigate when and why people prefer bad news. Study 1 manipulates the certainty of an unfavorable diagnosis to assess the impact on affective response. Study 2 examines which diagnoses people would prefer to receive, and investigates the moderating role of facing a decision. Using moderation and mediation, Study 3 examines how the ease or difficulty of making a decision influences affective response—and can perversely motivate people to hope for the worst.

Study 1: Feel Better with the Worst

Study 1 offers an initial, between-subjects demonstration of hoping for the worst. When faced with a consequential decision—whether to surgically remove one's appendix—people paradoxically feel better after receiving worse news.

Pretest

In a separate pretest, participants ($N = 47$; 53% female; $M_{age} = 32.2$, $SD = 9.2$) assessed which news was worse. They were asked, "Which diagnosis conveys worse news?" and were given the choices of "You have a 50% chance of appendicitis" and "You have a 95% chance of appendicitis." Almost all participants (91.5%, 95% CI = [.84, .99]) selected the latter diagnosis,

providing support for our account that having higher chances of a bad thing constitutes worse news.

Method

Participants. Participants (N = 202; 50% female; M_{age} = 33.0 years, SD = 9.6) completed an online survey via Amazon's Mechanical Turk and were paid a flat rate for participation.

Procedure. Study 1 used a between-subjects design with two conditions: a *better news* condition (i.e., 50% chance of appendicitis) versus a *worse news* condition (i.e., 95% chance of appendicitis). All participants read a vignette about possibly having appendicitis and needing to decide whether or not to undergo an appendectomy:

> Imagine you are suffering from what could be appendicitis, which could be fatal if left untreated. The only treatment is surgical removal of the appendix. You visit your doctor. He is not completely sure that you have appendicitis—it can be difficult to diagnose definitively—but suggests that you consider having your appendix out regardless, just to be safe.
>
> This is a difficult decision for you. You are worried about the pain of surgery and recovery, the expense of the procedure, and the time you'll have to take off from work. But you also do not want to risk having complications from appendicitis.
>
> Your doctor then performs some tests to help confirm the diagnosis. Based on the results, he tells you there is a [50% / 95%] chance you have appendicitis. If it is NOT appendicitis, your doctor thinks it's a false alarm and you will recover without intervention. Based on this information, you will have to make the final decision about whether or not to have the surgery.

Participants were then asked three questions assessing their affective response. They indicated the extent to which they would feel happy, relieved, and anxious (reverse-coded) on a 7-point Likert scale (1 = *not at all* to 7 = *very*); we averaged these three items to create a composite measure of affective response (α = .77). Finally, on a separate page, participants indicated whether they would choose to have their appendix removed (*yes* / *no*).

This and all subsequent experiments concluded with basic demographic questions. We set the desired number of participants at the outset of the experiments, targeting recruitment of at least 100 participants per cell; we did not analyze the data until the final number of participants was reached. No data were excluded and we report all measures and conditions.

Results

Unsurprisingly, more participants chose surgery in the worse news condition (94.1%, 95% CI = [.89, .99]) than in the better news condition (69.0%, 95% CI = [.60, .78]; $\chi^2(1, N = 202) = 21.29, p < .001$, Cramer's V = .33). However, these participants also felt better: despite the increased chance of appendicitis and the unpleasant realities of surgery, participants in the worse news condition reported greater positive affect ($M = 2.60$, $SD = 1.44$, 95% CI = [2.32, 2.88]) than participants in the better news condition ($M = 2.13$, $SD = 1.13$, 95% CI = [1.91, 2.35]; $t(200) = -2.57, p = .01, d = .36$). Further, while participants in the better news condition were more divided on whether or not to have surgery, those who ultimately chose surgery felt the same as those who did not ($M_{surgery} = 2.20$, $SD = 1.23$, 95% CI = [1.75, 2.65] vs. $M_{no\text{-}surgery} = 2.10$, $SD = 1.09$, 95% CI = [1.83, 2.36]; $t(98) = -.44, p = .66, d = .09$), suggesting that this group's lower overall positive affect was not merely driven by people choosing surgery under uncertain circumstances. (There was also no difference in affective response between those who did or did not choose surgery in the worse news condition; $p = .11$).

Discussion

To be sure, participants were not thrilled about possibly having appendicitis; however, participants in the worse news condition felt *less* bad, suggesting that when bad news confers certainty about the "right" choice to make, the negative impact of the bad news can be mitigated. The next studies explore when and why this occurs.

Study 2: Choose the Worst

Having shown that worse news can increase positive affect, Study 2 explored which diagnoses people would prefer to receive. In other words, do people actively and deliberately hope for the worst?

We suggest that "hoping for the worst" only occurs when people face difficult decisions, as worse-but-certain news can help identify the "right" choice to make. To test this, Study 2 varied whether participants did or did not have to make a decision based on news they received. All participants imagined having a shoulder tendon tear—the size of which was yet unknown—and then indicated the tear length they would prefer to have. In the *decision* conditions, surgery was required above a specific threshold length, which varied between-subjects as 2-, 3-, or 4-centimeters; below that threshold, participants could opt for surgery but the decision was theirs to make. In the *no decision* conditions, threshold lengths were discussed but no consequent decision was required.

We predicted that participants in the decision conditions would prefer larger, "above-threshold" tears that would require surgery and minimize uncertainty, and accordingly, that tears would increase with threshold size: participants would prefer tears larger than 2-centimeters when the surgery threshold was 2-centimeters, larger than 3-centimeters when the threshold was three, and larger than 4-centimeter when the threshold was four. In contrast, we predicted that participants in the no decision conditions—facing no choice uncertainty—would hope for better news (i.e., minimal injuries) across all thresholds.

Method

Participants. Participants (N = 548; 42% female; M_{age} = 32.5 years, SD = 11.0) completed an online survey via Amazon's Mechanical Turk and were paid a flat rate for participation.

Procedure. Study 2 used a 2 (decision: *no decision* vs. *decision*) by 3 (threshold: *2-centimeter*, *3-centimeter*, or *4-centimeter*) between-subjects design. All participants first read the following: "Imagine you have been experiencing pain in your shoulder, so you go to the doctor to determine the cause of the pain. Your doctor is fairly certain that the pain is the result of a tendon tear; the question is how big of a tear it is. A longer tear is a worse injury." Participants in the *no decision* conditions further read: "Tendon tears are typically 1 to 5 centimeters in length. If a tear is larger than [2 / 3 / 4] centimeters, it is considered a 'Mathysen tear.'" Participants in the *decision* conditions read: "Tendon tears are typically 1 to 5 centimeters in length. If a tear is larger than [2 / 3 / 4] centimeters, it is considered a 'Mathysen tear,' and surgery is always the medically indicated treatment. Therefore, if your tear is [2 / 3 / 4] centimeters or longer, your medical treatment is certain and you will have no decision to make; you will need surgery. If your tear is less than [2 / 3 / 4] centimeters, you will have to personally decide whether or not to have surgery, which many people consider a difficult decision. Surgery can be painful, but guarantees proper healing in the long-run." (We created the term "Mathysen tear" to serve as an identical anchor in both conditions.) At the end of both vignettes, we told participants to assume all costs were covered by insurance.

As our dependent measure, we asked participants: "What tear length would you prefer to have in this situation?" Preferred tear length was indicated on a slider bar from 1 to 5 centimeters, recorded to one decimal place. To take the most conservative measurement, the

slider's default position was the far left side of the bar (i.e., 1 centimeter), such that indicating a non-minimal tear length required a volitional action.

Results

We first analyzed average preferred tear length. A 2 (decision: *no decision* vs. *decision*) by 3 (threshold: *2-centimeter*, *3-centimeter*, or *4-centimeter*) ANOVA revealed significant differences in preferred tear length by condition. There was a main effect of decision, such that participants in the decision conditions hoped for larger tears (M = 1.67 centimeters, SE = .05, 95% CI = [1.57, 1.77]) than participants in the no decision conditions (M = 1.28, SE = .05, 95% CI = [1.18, 1.38]; $F(1, 542) = 28.25$, $p < .001$, $\eta^2 = .05$). There was also a main effect of threshold, such that preferred tear length increased as the size of the threshold increased (M_{2cm} = 1.34, SE = .06, 95% CI = [1.22, 1.47]; M_{3cm} = 1.45, SE = .06, 95% CI = [1.32, 1.58]; M_{4cm} = 1.63, SE = .06, 95% CI = [1.50, 1.75]; $F(2, 542) = 5.09$, $p = .006$, $\eta^2 = .018$). Critically, these main effects were qualified by a significant interaction ($F(2, 542) = 4.01$, $p = .019$, $\eta^2 = .015$): as the threshold size increased, preferred tear length increased specifically among those in the decision conditions. As a follow-up test, a one-way ANOVA revealed that—among the decision conditions—there were significant differences in preferred tear length as a function of threshold size ($F(2, 272) = 6.04$, $p = .003$, $\eta^2 = .04$; see Fig. 1); however, the same analysis for no decision conditions revealed non-significant differences ($F(2, 270) = 1.15$, $p = .32$, $\eta^2 = .008$). In other words, the average preferred tear changed as a function of threshold size for those in the decision conditions, but not in the no decision conditions.

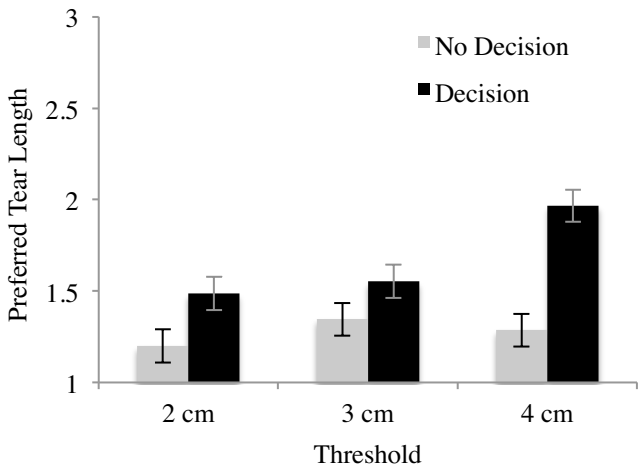

Figure 1: Average preferred tear length by condition (Error bars represent standard errors; Study 2)

For the decision conditions, we wanted to confirm that preferred tear lengths were not larger *arbitrarily*—instead, we expected that participants chose preferred tear lengths deliberately, specifically selecting diagnoses that required surgery and reduced decision uncertainty. To test this, we created a dichotomous measure for preferred tear length based on whether it was above or below the threshold: a length greater than or equal to the threshold for surgery ("above-threshold") was coded as 1, while a length less than the threshold ("below-threshold") value was coded as 0. (As an example, a preferred tear length of 2.5 would be coded as 1 for someone in the 2-centimeter condition and 0 for someone in the 3-centimeter condition, as 2.5 is above and below the thresholds, respectively.) Relative to no decision conditions, more participants in the decision conditions chose "above-threshold" tears across all threshold sizes (Fig. 2): when the threshold was 2-centimeters (*Decision* = 31.1%, 95% CI = [.21, .41] vs. *No Decision* = 7.8%, 95% CI = [.02, .13]; $\chi^2(1, N = 180) = 15.64, p < .001$, Cramer's V = .30), when the threshold was 3-centimeters (*Decision* = 19.1%, 95% CI = [.11, .27] vs. *No Decision* = 4.4%, 95% CI = [.00, .09]; $\chi^2(1, N = 180) = 9.44, p = .002$, Cramer's V = .23), and when the

threshold was 4-centimeters (*Decision* = 18.8%, 95% CI = [.11, .27] vs. *No Decision* = 2.2%, 95% CI = [-.01, .05]; $\chi^2(1, N = 188) = 13.58, p < .001$, Cramer's V = .27).

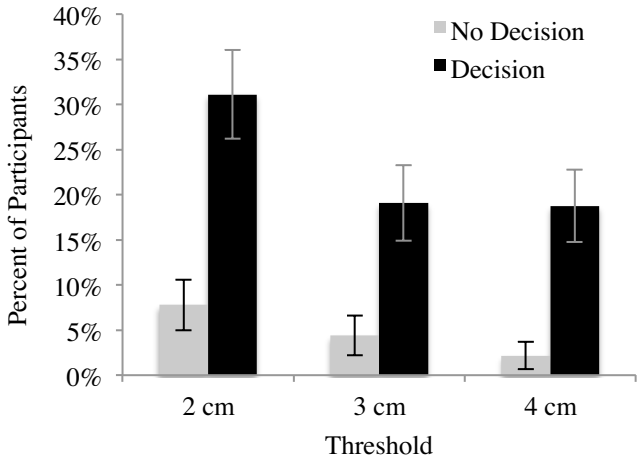

Figure 2: Participants who preferred "above-threshold" tears (Error bars represent standard errors; Study 2)

Discussion

Participants did not indiscriminately prefer bad news—only when facing a difficult decision were they motivated to hope for the worst. All participants had identical reference points—a tear range of 1-5 centimeters and a specific anchor ("Mathysen tear") of 2, 3, or 4 centimeters—but only participants in the decision condition faced a choice about surgery, prompting many to perversely desire a worse injury. Further, among the decision conditions' participants, the severity of the desired injuries varied with threshold size; many chose a tear that allowed the diagnosis—rather than themselves—to make the choice.

Study 3: Certain with the Worst

While worse news may beget worse outcomes, it can also deliver greater certainty, which can lower people's decisional conflict, and in turn, increase positive affect. Therefore, bad news should be most beneficial when people are especially conflicted (e.g., no clear "right choice"). In

contrast, when people feel an easy or obvious choice exists, bad news should not confer incremental benefit.

To test this, Study 3 used moderation and varied both diagnosis (*better news* or *worse news*) and disease—appendicitis (like Study 1) or appendiceal cancer. We expected to replicate Study 1's results for participants in the appendicitis condition: people would feel better with worse news. However, when confronting cancer, people strongly prefer action to inaction, and are uniquely compelled to choose aggressive treatments (Ubel, 2012). Therefore, we predicted that participants in the appendiceal cancer condition would find the surgery decision relatively easy to make—regardless of the valence of the news—and differences in affective response would diminish.

Study 3 also measured decisional conflict—uncertainty surrounding the correct action to take (O'Connor, 1995)—as a mediator. We anticipated that differences in affective response would be explained via decisional conflict, such that higher decisional conflict would decrease positive affect, particularly in the appendicitis condition.

Method

Participants. Participants ($N = 447$; 44% female; $M_{age} = 33.2$ years, $SD = 10.6$) completed an online survey via Amazon's Mechanical Turk and were paid a flat rate for participation.

Procedure. Study 3 used a 2 (diagnosis: *better news* vs. *worse news*) by 2 (decision context: *appendicitis* vs. *appendiceal cancer*) between-subjects design. All participants read the medical scenario from in Study 1, with either 50% (*better news*) or 95% (*worse news*) chance of either appendicitis or appendiceal cancer.

As in Study 1, participants indicated the extent to which they would feel happy, relieved, and anxious (reverse-coded) on a 7-point Likert scale (1 = *not at all* to 7 = *very*). We averaged these three items to create a composite measure of affective response ($\alpha = .77$).

Finally, we administered the 16-item Decisional Conflict Scale (O'Connor, 1995). The scale measures various factors contributing to decisional conflict (e.g., *I feel sure about what to choose*; *I am clear about the best choice for me*; *The decision is easy for me to make*). Participants responded to each statement using a 7-point Likert scale (1 = *strongly disagree* to 7 = *strongly agree*); for ease of interpretation, we reverse-coded responses such that higher scores reflect greater decisional conflict.

Results

Affective response. We ran a 2 (diagnosis: *better news* vs. *worse news*) by 2 (decision context: *appendicitis* vs. *appendiceal cancer*) ANOVA to analyze the results. The results indicate a main effect of diagnosis ($F(1, 443) = 16.39, p < .001, \eta^2 = .04$), but no main effect of disease ($F(1, 443) = .50, p = .48, \eta^2 = .00$). Importantly, the interaction was significant ($F(1, 443) = 3.94, p = .048, \eta^2 = .009$). Specifically, when facing cancer—a disease with a known bias for surgical intervention (Ubel, 2012)—there was no significant difference in affective response between participants receiving better news ($M = 2.65$, $SD = 1.37$, 95% CI = [2.40, 2.91]) and worse news ($M = 2.93$, $SD = 1.55$, 95% CI = [2.63, 3.22]; $t(221) = -1.40, p = .16, d = .19$). However, replicating Study 1's results, participants receiving worse news about appendicitis reported significantly greater positive affect ($M = 3.10$, $SD = 1.46$, 95% CI = [2.82, 3.37]) than those receiving better news ($M = 2.30$, $SD = 1.21$, 95% CI = [2.07, 2.52]; $t(222) = -4.46, p < .001, d = .60$; see Fig. 3).

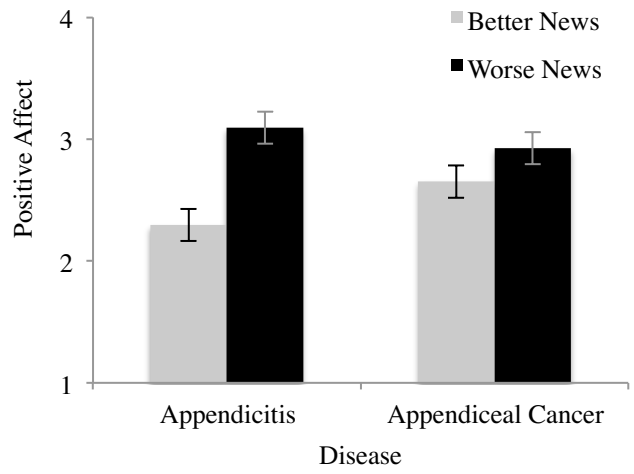

Figure 3: Affective response as a function of diagnosis and disease (Error bars represent standard errors; Study 3)

Decisional conflict. We ran the same 2x2 ANOVA using decisional conflict as the dependent measure. There was a main effect of both diagnosis ($F(1, 443) = 45.55, p < .001, \eta^2 = .09$) and disease ($F(1, 443) = .50, p = .48, \eta^2 = .00$); the latter supports our account that, on average, people experience less decisional conflict with cancer ($M = 2.45$, $SD = 1.15$, 95% CI = [2.30, 2.60]) than with appendicitis ($M = 2.68$, $SD = 1.24$, 95% CI = [2.52, 2.84]; $t(445) = 2.03, p = .04, d = .19$). The interaction was again significant ($F(1, 443) = 14.66, p < .001, \eta^2 = .03$): when facing appendiceal cancer, participants in the better news condition had higher decisional conflict ($M = 2.60$, $SD = 1.19$, 95% CI = [2.38, 2.83]) than those in the worse news condition ($M = 2.29$, $SD = 1.09$, 95% CI = [2.09, 2.50]; $t(221) = 2.03, p = .04, d = .27$), but this difference was especially pronounced for participants facing appendicitis ($M_{better\text{-}news} = 3.24$, $SD = 1.25$, 95% CI = [3.00, 3.47] vs. $M_{worse\text{-}news} = 2.12$, $SD = .93$, 95% CI = [1.94, 2.29]; $t(222) = 7.61, p < .001, d = 1.02$; see Fig. 4).

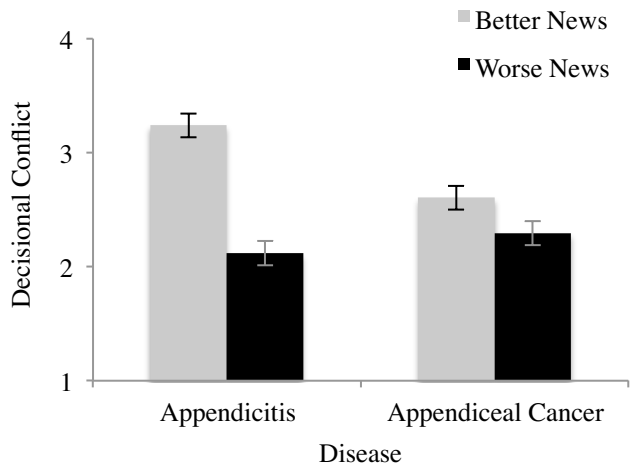

Figure 4: Decisional conflict as a function of diagnosis and disease (Error bars represent standard errors; Study 3)

Mediated moderation analysis. We conducted a mediated moderation analysis to simultaneously test moderation by disease (appendicitis vs. appendiceal cancer) and mediation by decisional conflict (Edwards and Lambert, 2007). As reported above, disease moderated both the dependent variable (affective response) and the mediator (decisional conflict). Further, when affective response was regressed on diagnosis, disease, their two-way interaction, and decisional conflict, the mediator was significant (B = -.26, SE = .06, 95% CI = [-.37, -.14]), and the effect of the interaction between diagnosis and disease was reduced to non-significance (from B = -.53, SE = .27, 95% CI = [-1.05, .00], p = .048 to B = -.32, SE = .27, 95% CI = [-.85, .21], p = .24). A 5,000 sample bootstrap analysis showed that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero (95% CI = [-.37, -.10]), suggesting a significant indirect effect (Preacher & Hayes, 2004). Finally, the effect size of the mediator was significantly higher in the appendicitis condition (.29) than in the appendiceal cancer condition (.08), supporting our account that decisional conflict is especially impactful in situations where the "right choice" is unclear.

Discussion

Using moderation and mediation, Study 3 provided further evidence of the main effect and its drivers. First, we replicated Study 1: worse news increased positive affect more than better news. Second, we showed that—even holding the decision constant (i.e., whether to surgically remove one's appendix)—the benefit of worse news is mitigated when the decision feels easier to make (i.e., as a cancer patient rather than an appendicitis patient). Finally, Study 3 demonstrated that positive affect increases as decisional conflict decreases.

General Discussion

Three studies demonstrated when and why people paradoxically hope for the worst. Study 1 investigated the relationship between news and affect: people felt better receiving worse news. Using moderation, study 2 examined preferences for bad news: many people actively preferred worse diagnoses, but only when a decision had to be made. Using moderation and mediation, Study 3 showed that bad news confers the most benefit when the "right choice" is not obvious, and demonstrated the inverse relationship between decisional conflict and positive affect.

Our findings make several theoretical contributions. First, we add broadly to research on uncertainty aversion, showing the lengths to which people will go—or the harm they will willingly incur—to attain certainty. We document a novel and consequential risk preference: that faced with a tradeoff between the negativity of uncertainty and the negativity of news or risk, people may prefer to reduce the former more than the latter. We also contribute specifically to research on the uncertainty effect—i.e., that worse-but-certain options may be valued higher than better-but-uncertain ones (Gneezy et al., 2006; Simonsohn, 2009)—by identifying ways in which

the effect may have broader applicability than originally identified. While Gneezy et al. (2006) found that the uncertainty effect disappeared in a within-subjects context, we find evidence that this is not always the case: when there is a consequential choice to be made, people will deliberately forgo better-but-uncertain options in favor of worse-but-certain ones. Further, this happens in a loss-frame, when a preference for a worse-but-certain option necessarily entails greater downside.

We also contribute to an understanding of how people respond to difficult, consequential choices. Previous research has shown that consequential choices are emotionally burdensome for any decision-maker (Botti et al., 2009; Kahn & Luce, 2003), but our results suggest that they may be even *more* difficult for decision-makers who have received objectively better news. For example, Botti et al. (2009) showed that when facing the gut-wrenching decision of whether to take their premature infants off life support, parents making the choice themselves felt significantly worse than parents for whom the decision was made (i.e., by a doctor). Applying our findings to this paradigm, we might expect that—among the parents making the decision themselves—those whose newborns had a better but more uncertain prognosis (e.g., very slim chance of survival) may have, paradoxically, felt worse than those whose newborns had the worst prognosis possible (e.g., no chance of survival).

In addition to theoretical contributions, our results also have significant practical implications. First, while people in our studies seemed to recognize their aversion to uncertainty—actively choosing worse-but-certain options over better-but-uncertain ones in within-subjects designs—the tendency to "hope for the worst" may be less obvious to observers. We asked a separate group of participants ($N = 98$) to predict who feels worse: Friend A with a 50% chance of appendicitis or Friend B with a 95% chance of appendicitis, both of whom faced

a surgery decision. Studies 1 and 3 provide strong evidence that Friend A feels worse, but most participants (71%, 95% CI = [.62, .81]) incorrectly predicted Friend B—the one receiving worse news—felt worst (results in Supplemental Material). In other words, observers seemed to attend to the valence of the news while overlooking the resulting decisional conflict that would ensue. This misprediction may have profound interpersonal consequences: without recognizing the negative impact of better-but-uncertain news, people in support roles may miss the opportunity to counsel or console those who need it most. For example, patients choosing prophylactic treatment to minimize *risk* of disease (better-but-uncertain diagnosis) may suffer greater decisional and emotional conflict than patients choosing therapeutic treatment for a *known* disease (worse-but-certain diagnosis)—a circumstance that may go unrecognized by doctors, friends, and family.

"Hoping for the worst" may also introduce unanticipated perverse incentives. Study 2 showed that people desire worse injuries; might this perverse preference also provoke perverse action? If a tendon tear were on the cusp of *certainly* needing surgery, might a patient be tempted to exacerbate the injury through excessive activity, and in so doing, avoid making the surgery decision himself? More broadly, might people forgo opportunities to improve their diagnosis if improvement meant swapping worse-but-more-certain news for better-but-more-uncertain news? Anecdotally, we know of a nurse, seven-months pregnant with her second child and wrestling with the decision of whether to have another cesarean section or attempt natural delivery; when she heard the fetus was breech—bad news necessitating a C-section—she admitted feeling relieved and hoping the baby would remain in the wrong position, allowing her to evade the delivery choice altogether. In this case, hoping for the worst might make a person feel better *and* actively forego risk reduction or improvement opportunities—e.g., declining procedures that

attempt to flip the baby into the correct (and less risky) head-down position. Our results help elucidate the circumstances in which these unexpected, perverse consequences of better-but-uncertain news may arise.

Boundary conditions undoubtedly exist. For instance, if there were no action that could be taken—or decision to be made—based on the bad news (Dawson, Savitsky, & Dunning, 2006), we would not expect people to hope for the worst. Further, with particularly grave diagnoses (e.g., a 50% versus 95% chance of terminal brain cancer), patients receiving better news may feel better overall, encoding uncertainty as hope. However, with specific regard to making a decision—particularly if the decision involves a grim trade-off (e.g., brain surgery with grave risks)—there still may be relief in receiving worse-but-certain news.

Finally, our studies used medical contexts, both because these decisions are especially consequential and because this domain clearly illustrates why hoping for the worst is perverse (i.e., wishing for worse bodily injury). However, people undoubtedly "hope for the worst" in other domains, as well. For instance, when deciding whether to fix or sell a car after a serious collision, an owner may feel better learning his car is totaled (worse-but-certain news) than that it is eligible for repair (better-but-uncertain news). Or consider a romantic relationship in which one partner suspects the other of being unfaithful and must decide whether to stay or leave; definitively knowing about an infidelity (worse-but-certain news) may feel better than simply believing infidelity was possible (better-but-uncertain news). In short, across an array of situations in a variety of domains, an aversion to uncertainty—coupled with the stress of making difficult decisions—may encourage people to paradoxically hope for the worst.

Pseudo-Set Framing

KATE BARASZ

LESLIE K. JOHN

MICHAEL I. NORTON

Abstract

Five studies demonstrate that arbitrarily grouping items or tasks together as part of an apparent "set" motivates people to reach perceived completion points: to finish a *pseudo-set*. We use both verbal (e.g., "there are four tasks in every set") and visual cues (e.g., a four-slice pie chart that "fills in" as tasks are completed) to show that *pseudo-set framing* affects task completion patterns. Specifically, pseudo-set framing alters perceptions of completeness, making intermediate progress (e.g., "3 of 4 tasks done") seem less complete; in turn, this feeling of incompleteness motivates people to persist until the pseudo-set has been fulfilled. Pseudo-set framing changes giving behavior (Studies 1 and 5), effort (Studies 2 and 3), and gambling (Study 4). The effects persist in the absence of any reward, when a cost must be incurred, and even when participants are explicitly informed that set size is completely arbitrary.

Suppose you are asked to participate in a charitable service project by writing greeting cards to isolated senior citizens. You can complete as many or as few cards as you would like. How many would you write? And how would you determine when to stop?

People frequently encounter tasks with no obvious stopping point, begging the question: "How much is enough?" How many times should we donate to a cause? How many friends should we invite to join a new service or try a new product? How many tedious questions should we voluntarily complete as lab participants? Contexts that lack explicit targets also lack signals for satisfactory endpoints. Fundraisers, organizations, and researchers often prefer more donations, referrals, and data, but lacking a desirable stopping point—and absent incentives for completion—people may feel their efforts are "complete enough" and quit prematurely. In other cases, the preference is for people to complete only a certain number of tasks: think of parents who want their child to eat only six pieces of Halloween candy at a sitting, or a gym that hopes customers come at least twice a week (so they use the gym enough to renew their membership) but not more (so the gym is not overcrowded). How can people be encouraged to do the "right" number of tasks—no more, and no less?

We suggest a solution to this problem: presenting tasks as constituent parts of a cohesive "set." Instead of simply being asked to write "as many cards as you want" for lonely seniors, imagine learning that the cards will be arbitrarily batched in sets of four, or having your card-writing progress tracked as part of a four-slice pie chart that "fills in" as you go. Even without any explicit reason for the size of these sets, we suggest these arbitrary "sets" make people more likely to complete exactly four cards. (Likewise, pie charts with five or six slices make people more likely to complete exactly five or six cards). We call this *pseudo-set framing*: grouping discrete items or tasks together as part of a cohesive set via verbal descriptions and visual cues.

We show that without providing an explicit target or goal, pseudo-sets signal how many tasks or items people "should" do if they want to reach a satisfying endpoint—even when completion entails additional effort with no additional reward.

We suggest pseudo-set framing is effective because pseudo-sets convey distinct reference points; such reference points have been repeatedly shown to predictably shift behavior (Heath, Larrick, & Wu, 1999; Kahneman & Tversky, 1979). Indeed, a large body of research demonstrates that judgments and choices are sensitive to the manner in which items are partitioned, assorted and categorized (e.g., Fox, Ratner, & Lieb, 2005; Kahn & Wansink, 2004; Read & Loewenstein, 1995). In particular, sets are perceived differently, convey distinct meaning, and elicit specific behaviors (Campbell, 1958; Dasgupta, Banaji, & Abselson, 1999; Evers, Inbar, & Zeelenberg, 2014; Hamilton & Sherman, 1996; Koffka, 1922; Köhler, 1947). For instance, people report a greater willingness to donate to cohesive sets of victims (e.g., six disadvantaged children from one family) than to the equivalent number of discrete victims (e.g., six individual children; Smith, Faro, & Burson, 2013). Similarly, qualitative research suggests collectible sets, like baseball cards or coins, signal the "right" number of goods to purchase, motivating a "collect them all" mentality (Belk, 1994; Danet & Katriel, 1994; McIntosh & Schmeichel, 2004; Stewart, 1993).

Sets such as baseball cards, however, are "real" sets, in that a) they contain varied members or items, which b) comprise a meaningful whole, such that c) there is often a reward to completing the set: baseball cards each depict a different player, the complete set comprises all players in the league, and complete sets are often priced higher than the sum of their constituent parts. (And per the above example, complete families may lack a higher price, but have variety and inherent meaning.) We suggest that despite lacking *all* of these features—variety, meaning,

and reward for completion—pseudo-sets continue to function as reference points, such that people feel compelled to complete the entire "set."

Much like round numbers, which people intuitively incorporate into judgments and behavior (Pope & Simonsohn, 2011; Rosch, 1975), we suggest that pseudo-sets convey reference points designating "completeness." However, unlike other naturally occurring reference points that are often fixed and inflexible (e.g., to be meaningful reference points, round numbers must, in fact, be exactly round), pseudo-sets are malleable and can make any arbitrary set size meaningful and motivating. By reconfiguring discrete tasks into a cohesive set—of any arbitrary number—we expected people would be more likely to expend effort or resources to fulfill the set. Further, we predicted that when viewing one's progress relative to a pseudo-set, intermediate effort (e.g., "3 of 4 tasks completed") would feel less complete relative to a control (e.g., "3 tasks completed"); in turn, this feeling of incompleteness would drive people to persist until pseudo-sets are fulfilled.

Five studies demonstrate pseudo-set framing in a variety of contexts—from donations to gambling—while offering support for our proposed mechanism. To demonstrate that pseudo-sets exert their impact by conveying distinct and malleable reference points, we assess three behaviors that prior research has shown are sensitive to reference points. First, people work harder to reach reference points (Heath et al., 1999; Kahneman & Tversky, 1979); therefore, we examine whether pseudo-sets change effort levels (Studies 1-3). Second, people become more risk-seeking if they have not yet reached a reference point (Heath et al., 1999; Kahneman & Tversky, 1979; Payne, Laughhunn, & Crum, 1980); accordingly, we predicted an increase in risk-taking under pseudo-set framing (Study 4). Third, different reference points lead people to perceive identical levels of performance differently (Heath et al., 1999; Medvec & Savitsky,

1997; Pope & Simonsohn, 2011); hence, even holding absolute progress constant, pseudo-sets should change perceptions of that progress, and in turn, change completion patterns (Study 5).

Study 1: Reframing a Service Project

Study 1 investigates whether pseudo-sets affect people's attrition points in an ongoing task. Participants wrote holiday cards to seniors in nursing homes; they could write as many (or few) as they wished. In the control condition, participants simply saw a count of their completed cards; in the pseudo-set condition, participants learned there were four cards in every "batch"—a number we selected arbitrarily and never further explained—and viewed their progress framed accordingly (e.g., "You have completed 75% of one batch!"). We expected participants in the pseudo-set condition would be more likely to complete cards in multiples of four, since four cards comprised the "pseudo-set" reference point. We included exploratory items assessing possible mechanisms underlying pseudo-set framing: perceived impact, meaningfulness, and enjoyment.

Method

Participants. Participants (N = 192; 54% female; M_{age} = 23.5 years, SD = 4.7) completed the study at a laboratory at a U.S. university as part of a multi-study session conducted a few weeks before the December holidays. All participants were randomly assigned to a desk that was separated by a partition, such that participants could not readily observe others seated nearby.

Procedure. Study 1 was a two-condition, between-subjects design. All participants learned they would contribute to a nonprofit's holiday card-writing drive, which "facilitates the sending of holiday cards to seniors in nursing homes who are particularly isolated." On every desk, there were 12 identical, non-denominational, blank holiday greeting cards, one pen, and

large collection envelopes in which completed greeting cards were placed. All participants were asked to complete at least one card and were given a 10-line greeting to handwrite inside, ensuring that all cards were identical and there was no variety in the card-writing experience. The actual writing task was exactly the same for all participants, and there were no incentives for completing cards.

Before starting the card-writing task, participants in both conditions read: "Our plan is to collect as many cards as possible, and send roughly 200 cards to each nursing home. Each senior will receive one card." Pseudo-set participants received one additional line of instructions: "Within each nursing home package, we are batching cards in sets of four." After completing the single required card, participants could write more cards if they chose to. If they wanted to complete more cards than had initially been provided on their desk, an experimenter would bring more, creating an unlimited series of tasks with no definite endpoint.

Participants tracked their progress by clicking a button on their computer after finishing each card; this was the point at which we manipulated the framing of participants' card-writing progress. On their computer screens, control participants received a simple count of how many cards they had completed: "Thanks! You have completed ONE card. Would you like to write another? (Yes/No)." If they answered "No," they were taken to the end of the section. If they answered "Yes," they completed another card and then received an updated progress report: "Thanks! You have completed TWO cards. Would you like to write another?" This loop continued for as long as participants agreed to write cards.

Pseudo-set participants' progress was framed in terms of the pseudo-set: "Thanks! You have completed 25% of one batch of cards. Would you like to write another? (Yes/No)." We also included a visual indication of progress: a pie chart with four component parts representing one

set. A "slice" filled in for every card participants completed (see Fig. 5). Once a full set was completed, an additional pie chart appeared to the right with four new empty slices; note that in both conditions, participants were aware of the total number of cards they had completed.

To show that pseudo-set framing affects completion behaviors relative to baseline, our dependent measure was whether or not participants completed "full sets"—exact multiples of four cards.

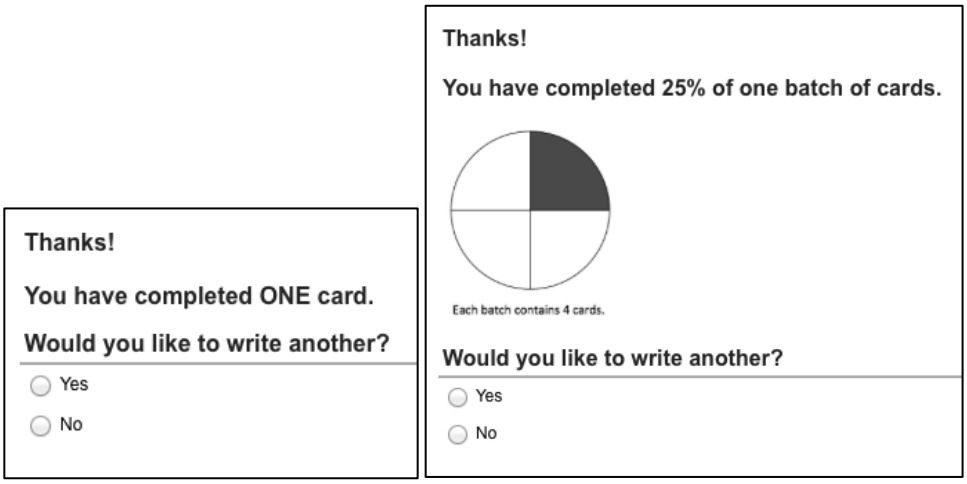

Figure 5: Progress display for control (left) and pseudo-set (right) conditions (Study 1)

At the end of the study, participants completed six exploratory items. Participants rated how meaningful they thought each card would be to the recipient (1 = *not at all meaningful* to 7 = *very meaningful*), and how much they agreed (1 = *strongly disagree* to 7 = *strongly agree*) with four prosocial impact statements (adapted from Grant et al., 2007): *I feel that my donation makes a positive difference in other people's lives; I am very aware of the ways in which my donation is benefiting others; I am very conscious of the positive impact that my donation has on others; My donation really makes others' lives better*. We averaged these four items to create a composite measure of impact (α = .96). Finally, we asked participants how much they enjoyed the service project (1 = *not at all* to 7 = *very much*).

This and all subsequent experiments concluded with demographic questions. We set the desired number of participants at the outset of the experiments and did not analyze the data until that number was reached. No data were excluded and we report all conditions and measures.

Results

Our primary dependent measure was dichotomous: whether or not participants completed a multiple of four cards. The introduction of pseudo-sets significantly changed participants' likelihood to complete full sets of cards (i.e., multiples of four). In the control condition, just 4.3% (95% CI = [.00, .08]) of control participants completed a multiple of four cards; in the pseudo-set condition, 39.8% (95% CI = [.30, .50]) of participants completed a multiple of four cards ($\chi^2(1, N = 192) = 34.87, p < .001$, Cramer's V = .43). These results demonstrate that pseudo-set participants' completion of exact sets was non-random. A two-sample Kolmogorov-Smirnov test confirmed that the two distributions were distinct ($Z = 1.87, p = .002$); Figure 6 shows a clear spike at four cards completed in the pseudo-set condition, which is not present in the control condition.

No participants asked for more cards beyond the 12 that were initially provided, and research assistants checked completed cards to make sure the appropriate greeting had been written.

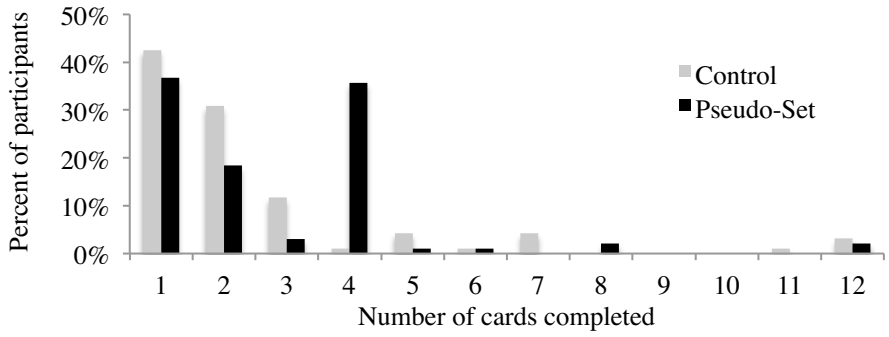

Figure 6: Number of cards completed by condition (Study 1)

Pseudo-set framing did not affect participants' perceptions of impact: participants saw the cards they wrote as equivalently meaningful ($M_{control}$ = 5.1, SD = 1.49, 95% CI = [4.83, 5.44] vs. $M_{pseudo\text{-}set}$ = 4.8, SD = 1.71, 95% CI = [4.41, 5.10]; $t(190) = 1.65$, $p = .10$, $d = .24$) and impactful ($M_{control}$ = 4.9, SD = 1.51, 95% CI = [4.61, 5.23] vs. $M_{pseudo\text{-}set}$ = 4.7, SD = 1.50, 95% CI = [4.37, 4.97]; $t(190) = 1.17$, $p = .24$, $d = .00$), suggesting that a heightened sense of personal impact does not explain our results. Control participants rated the task as more enjoyable relative to pseudo-set participants ($M_{control}$ = 5.4, SD = 1.51, 95% CI = [5.11, 5.72] vs. $M_{pseudo\text{-}set}$ = 4.8, SD = 1.70, 95% CI = [4.46, 5.15]; $t(190) = 2.62$, $p = .01$, $d = .37$), and number of cards completed was more highly correlated with enjoyment for control participants ($r = .30$, 95% CI = [.17, .41], $p = .004$) than pseudo-set participants ($r = .19$, 95% CI = [.00, .39], $p = .07$), suggesting that something other than mere enjoyment drove pseudo-set participants to continue writing cards. We more fully explore our proposed mechanism underlying pseudo-set framing—feelings of completeness—in the studies that follow.

Study 2: Putting the "Pseudo" In "Pseudo-Set"

While the size of the pseudo-sets we used in Study 1 (batches of four cards) was left intentionally vague and unexplained, it is possible participants made some inference of meaning—perhaps that it was somehow easier for the nursing home to receive cards in batches of four. In Study 2, we clearly told participants that set size is arbitrary and without meaning, showing them two different possible set sizes and randomly—and explicitly—assigning them to one of the two. We anticipated that even when participants knew their given set size was determined arbitrarily, their behavior would continue to be shaped by a desire to complete pseudo-sets.

Method

Participants. Participants (N = 362; 62% male; M_{age} = 32.6 years, SD = 9.7) completed an online survey via Amazon's Mechanical Turk and were paid a flat rate for participation.

Procedure. Study 2 was a three-condition, between-subjects design. All participants were informed that they would be solving CAPTCHA verifications (i.e., visual perception tests requiring people to decode obscure words and numbers). The instructions specified: "There are an unlimited number of CAPTCHA tasks, but you may solve as many or as few as you'd like; there is no penalty for skipping to the end. You will be paid a flat rate of $0.30 no matter what."

To unambiguously communicate that pseudo-sets are arbitrary and meaningless, the instructions further explained: "All of the CAPTCHAs are numbered. However, to help track your progress, some of you will see graphics that depict how many you have solved. This will be determined randomly and has no bearing on your compensation. If you see a progress graphic, one slice will fill after every 5 CAPTCHAs that are solved. The two progress indicators are as follows." Below these instructions, participants saw two pie charts: one with four component parts, and one with seven component parts.

After reading the instructions, participants advanced to the next screen and learned how their progress would be tracked. Participants in the *control* condition read: "You will not see a progress graphic, but you will be able to see which CAPTCHA number you are completing on each screen." Participants in the *four-unit* and *seven-unit* pseudo-set conditions saw the appropriate progress graphic and read: "Your progress will be tracked with the below graphic. For every 5 CAPTCHAs you complete, one of these slices will fill." The implication was that four-unit participants needed to complete 20 CAPTCHAs to fulfill a set, while seven-unit participants needed to complete 35 CAPTCHAs to fulfill a set. All participants had an option to

"Skip to the end of the section" at the bottom of every page and could complete as many or few CAPTCHAs as they wanted. They could not advance until the CAPTCHA had been correctly solved.

Despite knowing the pseudo-set graphic was randomly assigned and unrelated to their compensation, we anticipated participants would complete full sets at non-chance levels. Therefore, we measured how many participants completed exact multiples of 20 CAPTCHAs (a complete set for four-unit participants) and how many completed exact multiples of 35 (a complete set for seven-unit participants). We expected four-unit participants would be more likely to complete multiples of 20 tasks than participants in the control or seven-unit pseudo set conditions; likewise, seven-unit participants would be more likely to complete multiples of 35 tasks than participants in the other two conditions.

Results

Our primary dependent measure was dichotomous: whether or not participants completed a full pseudo-set, as specified by the graphic they encountered. Despite full knowledge of the graphic's meaninglessness, pseudo-set participants still completed full sets at non-chance levels. Participants in the four-unit pseudo-set condition (19.8%, 95% CI = [.13, .27]) were significantly more likely to complete exact multiples of 20 tasks than participants in the seven-unit pseudo-set condition (4.2%, 95% CI = [.01, .08]; $\chi^2(1, N=240) = 13.80, p < .001$, Cramer's V = .24) or control condition (3.3%, 95% CI = [.00, .06]; $\chi^2(1, N=243) = 16.33, p < .001$, Cramer's V = .26; overall chi-square $\chi^2(2, N=362) = 25.27, p < .001$, Cramer's V = .26; see Fig. 7). Likewise, participants in the seven-unit pseudo-set condition (11.8%; 95% CI = [.06, .18]) were significantly more likely to complete exact multiples of 35 tasks; no participants completed exact multiples of 35 in the four-unit pseudo-set condition (0.0%; $\chi^2(1, N=240) = 15.12, p < .001$,

Cramer's V = .25) or control condition (0.0%, $\chi^2(1, N=241) = 15.24, p < .001$, Cramer's V = .25; overall chi-square $\chi^2(2, N=362) = 29.74, p < .001$, Cramer's V = .29).

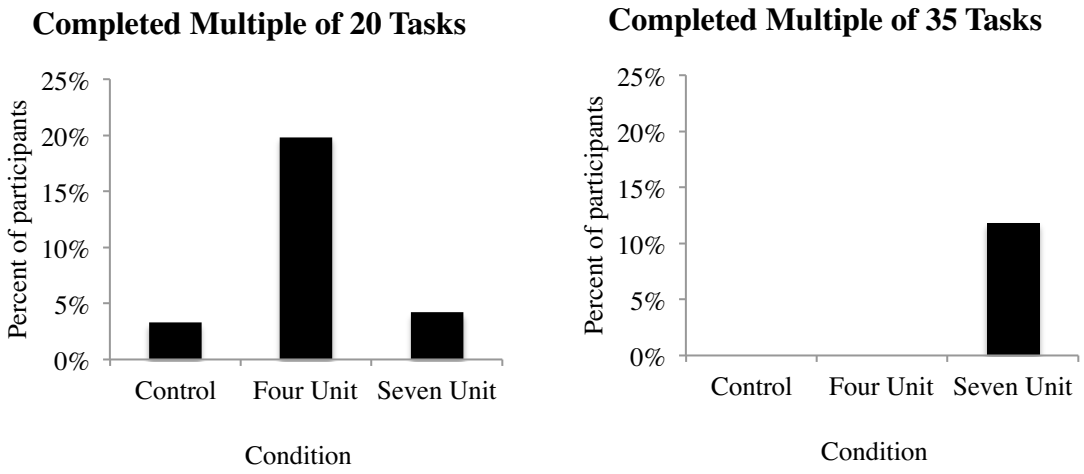

Figure 7: Percent of participants completing full pseudo-sets of 20 (left) or 35 (right) tasks (Study 2)

Study 3: Progress versus Completion

In Studies 1 and 2, pseudo-set participants had reference points that control participants did not (i.e., four cards, 20 or 35 CAPTCHAs), providing them with more information about their progress. In Study 3, we added a third condition—*visual control*—that provided visual evidence of progress but did not appear as a cohesive set: five pie wedges not arrayed in a circle (see Fig. 8). This condition allowed us to control for the amount and type of information displayed (Amir & Ariely, 2008; Cheema & Bagchi, 2011; Locke & Latham, 1990), and provided reference points identical to those in the pseudo-set: the maximum number of tasks, with progress depicted after every 5 tasks were completed. Despite having identical *numeric* reference points, we expected pseudo-sets participants would be more likely to complete the full set—again, even though completion came with no monetary reward.

Method

Participants. Participants (N = 331; 54% male; M_{age} = 52.1 years, SD = 12.3) completed an online survey via Amazon's Mechanical Turk and were paid a flat rate for participation.

Procedure. Study 3 was a three-condition, between-subjects design. All participants were told they would see 25 dot-counting tasks and that they could do as many or as few tasks as they wanted. Each dot-counting task was presented on its own screen, and a "Skip to the End" button appeared at the bottom of every page. There were between 20 and 60 dots on every screen and participants had to enter the total number in a text box. The task number was displayed next to each question, so participants always knew how far into the task they had advanced.

To manipulate framing, we randomly assigned participants to one of three conditions: control, visual control, and pseudo-set. In the control condition, participants only saw the task number they were currently completing; thus, the tasks appeared as 25 repeating questions, unrelated and discrete. In the visual control condition, participants saw the task number and a graphic with five discrete "pie slices"; they were told that one circle would fill after every five questions (or one-fifth of the total questions) were answered. In the pseudo-set condition, participants saw the task number and a pie chart graphic with five component slices; they were told that one slice would fill after every ten questions answered (see Fig. 8).

We expected pseudo-set participants to be more likely to see the 25 tasks as part of a set, and be more motivated for full completion relative to the other two conditions. Therefore, our dependent measure was the percent of participants who voluntarily completed all 25 dot-counting tasks—the full "set."

At the end of the study, participants completed three follow-up measures that assessed task perceptions: *Would you like to continue doing more dot-counting questions?* (*yes*/*no*); *How much did you enjoy the dot-counting task?* (1 = *not at all* to 7 = *very much*); and, *How satisfied*

are you with your effort and performance on the dot-counting task? (1 = *not at all satisfied* to 7 = *very satisfied*).

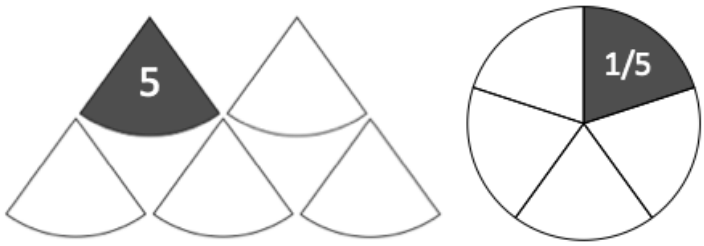

Figure 8: Progress display for visual control (left) and pseudo-set (right) conditions (Study 3)

Results

Our primary measure was the percentage of participants who completed the task set; a chi-square test revealed a significant main effect of condition on completion ($\chi^2(2, N = 331) = 13.47, p = .001$, Cramer's V = .20). Specifically, significantly more participants in the pseudo-set condition completed all 25 dot-counting questions (40.4%; 95% CI = [.31, .50]) than did participants in the visual control (26.8%; 95% CI = [.18, .35]; $\chi^2(1, N = 221) = 4.58, p = .03$, Cramer's V = .14) or control conditions (18.2%; 95% CI = [.11, .26]; $\chi^2 (1, N = 219) = 13.03, p < .001$, Cramer's V = .24). The control and visual control conditions did not differ significantly ($p = .13$, Cramer's V = .10).

Unlike Study 1, in which pseudo-set participants reported enjoying the task less, there was no effect of condition on ratings of task enjoyment ($M_{control} = 2.40, SD = 1.52$, 95% CI = [2.11, 2.69]; $M_{visual} = 2.64, SD = 1.62$, 95% CI = [2.34, 2.95]; $M_{pseudo\text{-}set} = 2.68, SD = 1.73$, 95% CI = [2.35, 3.01]; $F(2, 328) = .96, p = .39, \eta^2 = .00$). There was also no difference in satisfaction with progress and performance ($M_{control} = 4.7, SD = 1.88$, 95% CI = [4.34, 5.05]; $M_{visual} = 4.8, SD = 1.85$, 95% CI = [4.40, 5.10]; $M_{pseudo\text{-}set} = 4.7, SD = 1.65$, 95% CI = [4.41, 5.04]; $F(2, 328) = .10$,

$p = .97, \eta^2 = .00$). There was a marginally significant effect on participants' willingness to continue completing dot-counting tasks ($\chi^2(2, N=331) = 5.59, p = .06$, Cramer's V = .13): pseudo-set participants were somewhat more likely to say "yes" to counting more dots (12.8%, 95% CI = [.06, .19]) than those in the visual control condition (5.4%, 95% CI = [.01, .10]; $\chi^2(1, N=221) = 3.76, p = .05$, Cramer's V = .13) or control condition (5.5%, 95% CI = [.01, .10]; $\chi^2(1, N=219) = 3.60, p = .06$, Cramer's V = .13).

Study 4: The Cost of Completion

Prior research suggests that if a reference point has not yet been reached, people will become more risk-seeking to attain it (Heath et al., 1999; Payne et al., 1980). Accordingly, if pseudo-sets convey distinct references points, they should change the amount of risk people are willing to assume. Study 4 investigates the effect of pseudo-sets on risk-taking—specifically examining whether participants would risk a monetary cost for completion.

Method

Participants. Participants ($N = 201$; 58% male; $M_{age} = 52.5$ years, $SD = 11.0$) completed an online survey via Amazon's Mechanical Turk. Participants were paid a flat rate of $0.25 for their participation and received a bonus payment of up to $0.25 based on the gambles' outcomes.

Procedure. Study 3 used a two-condition, between-subjects design. Participants learned they had the chance to win up to $0.25, effectively doubling their participation compensation. Participants were first told that they would see four gambles and have the chance to accept or decline each one; if they declined, they would cash out of the study and forfeit additional gambles.

Each gamble was displayed on a separate screen:

Gamble 1: You have a 90% chance of winning $0.05 and a 10% chance of winning $0.00
Gamble 2: You have a 75% chance of winning $0.10 and a 25% chance of winning $0.00
Gamble 3: You have a 50% chance of winning $0.05 and a 50% chance of winning $0.00
Gamble 4: You have a 50% chance of winning $0.05 and a 50% chance of losing $0.10

After seeing the odds and stakes of each gamble, participants chose between two options: "Yes, I would like to take the gamble" or "No, I would like to cash out." Participants who selected "Yes" advanced to a new screen to learn whether they had won or lost the gamble; participants who selected "No" were taken to the end of the study.

The first three gambles had a positive expected value ($0.045, $0.075, and $0.025, respectively) with no chance of incurring a loss. In contrast, the fourth gamble—the one in which we were primarily interested because it offered the opportunity to "complete the set"—had a negative expected value (-$0.025) and the chance of losing $0.10. To ensure that all participants faced Gamble 4 with the same accrued bonus amount, we rigged Gambles 1 through 3 so that participants won each round; thus, all participants had accrued $0.20 out of $0.25 when making a decision about the risky fourth gamble.

Participants saw their accrued bonuses both numerically and graphically; in the visual control condition, participants saw a collection of nickels (with blank circles depicting unfulfilled winnings), while participants in the pseudo-set condition saw their bonus as fifths of a quarter (see Fig. 9). (Having established that the visual control condition provides the more conservative comparison, we omitted the non-visual control.) These graphics were meaningless: all participants were paid electronically and knew that they could cash out at any point and receive the amount displayed numerically on their screens. Our dependent measure was the percent of participants who accepted all four gambles.

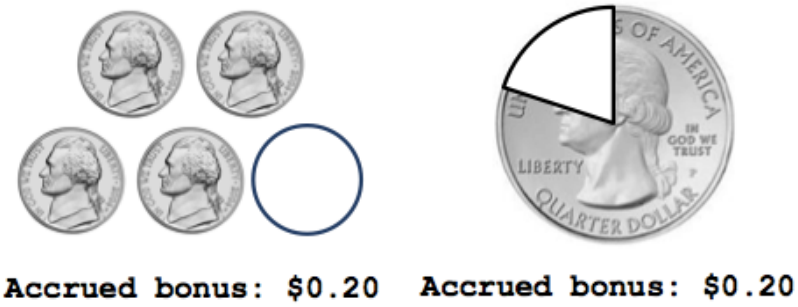

Figure 9: Bonus accrual display for visual control (left) and pseudo-set (right) conditions, immediately after Gamble #3 (Study 4)

Results

Our dependent variable was the percentage of participants in each condition who "completed the set" of gambles and accepted all four. As predicted, significantly more participants in the pseudo-set condition accepted all four gambles (29.4%, 95% CI = [.20, .38]) than in the visual control condition (16.2%, 95% CI = [.09, .24]; $\chi^2(1, N=201) = 4.99, p = .03$, Cramer's V = .16). In other words, pseudo-set framing made participants more willing to take one last (bad) gamble in an effort to complete an apparent set.

Study 5: Perceived Completeness

Studies 3-4 demonstrated that pseudo-sets are distinct relative to other reference points, signaling where people should stop if they want to reach a "complete" endpoint. We suggest pseudo-set framing also makes intermediate points appear less complete and satisfying, and this feeling of "incompleteness" in turn drives completion of pseudo-sets. Therefore, Study 5 examines the role of perceived completeness in motivating set completion.

In addition, while Studies 1-4 investigated the effect of pseudo-set framing on decisions to persist in ongoing tasks, Study 5 examines whether pseudo-set framing affects one-time

choices. We used a donation context in which donation amounts are typically presented in ongoing monetary intervals ($5, $10, and so on) with no clearly desirable "completion" point. We reframed these intervals to appear as part of a set and measured the impact on donation choices, expecting that the reframing would encourage more participants to complete the "donation set" by contributing the maximum amount.

Lastly, Study 5 instantiates pseudo-sets verbally—without reliance on visuals—by arbitrarily describing items as either individual units or as parts of a set.

Method

Participants. Participants (N = 302; 62% male; M_{age} = 31.3 years, SD = 10.1) completed an online survey via Amazon's Mechanical Turk and were paid a flat rate for participation.

Procedure. All participants made a decision about how much to hypothetically donate among a repeated set of options, and evaluated the perceived completeness of the midscale donation option. We counterbalanced the order in which participants saw the donation decision and perceived completeness questions; there was no order effect so we collapse across order.

For the donation decision, all participants read: "Imagine that you've been asked to donate money to a local school's book fundraiser. The school is in the process of updating its existing textbook inventory for the entire fourth grade." Participants could donate between $0 and $25 in multiples of $5.

Participants were randomly assigned to one of two conditions: control and pseudo-set. In the control condition, we framed donation units as discrete individual textbooks, designed to mirror typical donation scales. Control participants read:

> Each textbook costs $5 and you can donate up to 5 for a total of $25. How many would you be willing to donate?
> - 1 textbook ($5)
> - 2 textbooks ($10)

- 3 textbooks ($15)
- 4 textbooks ($20)
- 5 textbooks ($25)
- I would not donate any money

In the pseudo-set condition, participants were given the option to donate the same number of textbooks for the same corresponding dollar amounts, but textbooks were described as component parts of a set. Pseudo-set participants read:

> Each set of textbooks costs $25, and there are 5 books in every set. How many would you be willing to donate?
>
> - 20% of one set ($5)
> - 40% of one set ($10)
> - 60% of one set ($15)
> - 80% of one set ($20)
> - One whole set ($25)
> - I would not donate any money

Our dependent measure was the percent of participants who selected the maximum donation value, or completion of the full "set."

For the perceived completeness question, participants evaluated the perceived completeness of the midscale donation value ($15): "How complete does this donation choice feel to you?" As above, control participants saw the midscale donation value as "3 textbooks ($15)", while pseudo-set participants saw the midscale donation value as "60% of one set ($15)." Completeness was measured on a sliding scale with endpoints labeled "*Feels very incomplete*" and *"Feels very complete*." Actual scale values were hidden, but responses were recorded as a number between 0 and 100.

Next, all participants rated the perceived efficacy and impact of their donations. Participants indicated how effective they thought the textbooks would be (1 = *not at all effective* to 7 = *very effective*) and how much they agreed (1 = *strongly disagree* to 7 = *strongly agree*) with the four prosocial impact statements used in Study 1. As before, we averaged these four

items to create a composite measure of impact (α = .94). Finally, to demonstrate that pseudo-set framing changes perceived completeness at all intermediate, non-complete points, we asked participants to evaluate the perceived completeness of the other non-endpoint donation values (i.e., $5, $10, and $20), using the same scale described above. (We did not test the scale endpoint values of $0 or $25, anticipating floor and ceiling effects).

Results

Our dependent measure was the percentage of participants who were willing to donate the maximum amount—or the "complete set." The introduction of pseudo-sets significantly changed participants' likelihood to donate the maximum amount of $25. In the control, 21.7% (95% CI = [.15, .28]) of participants chose to donate the maximum amount, while in the pseudo-set condition, 38.0% (95% CI = [.30, .46]) donated the maximum amount ($\chi^2(1, N = 302) = 9.58$, $p = .002$, Cramer's V = .18). A two-sample Kolmogorov-Smirnov test confirmed that the two distributions were distinct ($Z = 1.42$, $p = .04$; see Fig. 10).

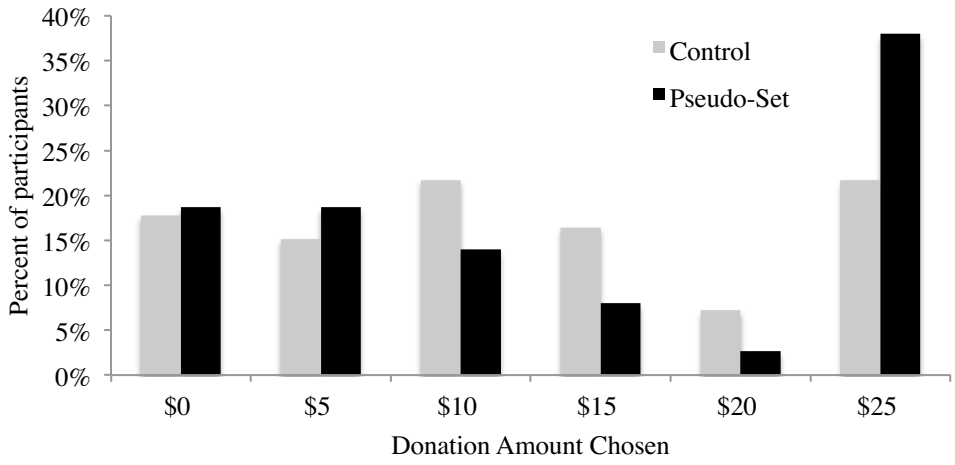

Figure 10: Donation amount chosen by condition (Study 5)

We expected pseudo-set participants to perceive the midscale donation value as significantly less complete than control participants. As predicted, when asked to assess the same

midscale donation value ($15), pseudo-set participants perceived it as significantly less complete ($M = 57.58$, $SD = 25.98$, 95% CI = [53.39, 61.77]) than control participants ($M = 71.09$, $SD = 20.43$, 95% CI = [67.81, 74.36]; $t(300) = 5.03$, $p < .001$, $d = .58$).

Further, we expected that perceived completeness would mediate the decision about whether or not to donate the maximum amount; the less complete participants viewed an intermediate donation value, the more likely they would be to increase their contributions to the maximum, "complete" amount. The effect of condition on selection of maximum donation was reduced (from $B = .79$, $p < .001$, to $B = .63$, $p = .02$) when perceived completeness was included in the model. A bootstrap analysis showed that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero [.04 to .37], suggesting a significant indirect effect (Preacher & Hayes, 2004).

As in Study 1, participants again viewed the donation as being equivalently effective ($M_{control} = 5.5$, $SD = 1.40$, 95% CI = [5.23, 5.68] vs. $M_{pseudo\text{-}set} = 5.5$, $SD = 1.23$, 95% CI = [5.35, 5.72]; $t(300) = -.54$, $p = .59$, $d = .06$) and impactful ($M_{control} = 5.2$, $SD = 1.42$, 95% CI = [4.97, 5.42] vs. $M_{pseudo\text{-}set} = 5.2$, $SD = 1.26$, 95% CI = [4.95, 5.36]; $t(300) = .26$, $p = .79$, $d = .03$). Thus, supporting the notion that the pseudo-set framing was truly arbitrary, participants in the pseudo-set condition were no more likely to think that donating five books would be impactful, relative to those in the control condition.

Pseudo-set participants perceived all intermediate donation amounts as being less complete than control participants; while the difference was only marginally significant at the $5 donation level ($M_{control} = 37.2$, $SD = 32.84$, 95% CI = [31.97, 42.49] vs. $M_{pseudo\text{-}set} = 30.7$, $SD = 28.05$, 95% CI = [26.16, 35.21]; $t(300) = 1.86$, $p = .06$, $d = .21$), there were significant differences at the $10 ($M_{control} = 48.5$, $SD = 27.03$, 95% CI = [44.19 52.85] vs. $M_{pseudo\text{-}set} = 39.2$,

SD = 21.95, 95% CI = [35.69, 42.77]; $t(300) = 3.28$, $p = .001$, $d = .38$) and $20 ($M_{control} = 82.5$, *SD* = 20.89, 95% CI = [79.19, 85.89] vs. $M_{pseudo\text{-}set} = 75.3$, *SD* = 22.76, 95% CI = [71.67, 79.01]; $t(300) = 2.86$, $p = .004$, $d = .33$) donation levels, suggesting that pseudo-set framing changed perceived completeness across all intermediate scale values.

General Discussion

Five studies reveal the impact of pseudo-set framing on behavior. The presence of pseudo-sets—arbitrary and devoid of meaning—altered the number of greeting cards written (Study 1), tasks completed (Studies 2 and 3), gambles taken (Study 4), and the amount of money donated (Studies 5). Pseudo-sets serve as compelling and distinct reference points above and beyond other reference points (Studies 3 and 4). Finally, pseudo-set framing changes perceptions of completeness at intermediate points, driving people to reach a satisfying endpoint: the completion of the pseudo-set (Study 5).

We contribute to research on reference points by offering a novel case that implicitly activates a "mere goal"—one not accompanied by external or discontinuous rewards (Heath et al., 1999). We suggest the critical aspect of pseudo-set framing—theoretically, and also in practice—is its complete arbitrariness. Unlike other research on sets in where sets are inherently meaningful (a family; Smith et al., 2013), contain variety (baseball cards; Stewart, 1993) and/or come with a reward (completing a loyalty card; Kivetz, Urminsky, & Zheng 2006; Nunes & Drèze, 2006), pseudo-sets are effective in the absence of these factors. As a result, the size of pseudo-sets is extremely flexible: the parent hoping to encourage six pieces of candy eaten or the gym hoping for four visits per month can easily change the size of these sets to change the relevant behavior.

While we have demonstrated the effectiveness of pseudo-set framing across a variety of contexts, pseudo-sets are likely to be bounded in their effectiveness. For example, if pseudo-set completion is too difficult or simply impossible, pseudo-set framing may lose its appeal; we generally chose tasks participants could complete given a moderate amount of effort. Relatedly, if the number of tasks required to fill a "pie piece" is prohibitively high, people may decide not to engage at all to avoid anticipated dissatisfaction with partial completion. Future research should investigate the boundary conditions—including the type and number of tasks required to make "progress"—under which pseudo-set framing is most effective and impactful. Still, our results show pseudo-set framing is effective across a wide variety of tasks, increasing effort—even at mind-numbing tasks—and promoting both prosocial behavior and gambling.

The Role of (Dis)similarity in (Mis)predicting Others' Preferences

KATE BARASZ

TAMI KIM

LESLIE K. JOHN

Abstract

Consumers readily indicate liking options that appear dissimilar—for example, enjoying both rustic lake vacations and chic city vacations, or liking both scholarly documentary films and action-packed thrillers. However, when predicting other consumers' tastes for the same items, people believe that a preference for one precludes enjoyment of the dissimilar other. Five studies show that people sensibly expect others to like similar products, but erroneously expect others to *dislike dissimilar* ones (Studies 1 and 2). While people readily select dissimilar items for themselves (particularly if the dissimilar item is of higher quality than a similar one), they fail to predict this choice for others (Studies 3 and 4)—even when monetary rewards are at stake (Study 3). The tendency to infer dislike from dissimilarity is driven by a belief that others have a narrow and homogeneous range of preferences (Study 5).

Suppose you encounter a consumer who is choosing between two products, Widget A and Widget B. After considering both options, the consumer decides on Widget A. Knowing of this choice, what would you predict about the consumer's preferences for forgone Widget B?

Choices reveal information about people's preferences—both for options that are chosen, as well as those that are forgone. When we observe others make a choice, we consistently assume they like the option they chose (Miller and Nelson 2002), but make more nuanced inferences about their preferences for unchosen options. As this paper demonstrates, people make different predictions about a forgone option depending on whether it is similar—or dissimilar—to a chosen option; simply put, people expect others to *like* similar options and to *dislike* dissimilar ones. Consequently, Widget B will be perceived as a well-liked substitute when it is relatively similar to Widget A, but believed to be disliked and actively rejected when it is relatively dissimilar to Widget A.

However, as we show, this intuition is often incorrect because it fails to account for the range and diversity of others' preferences: people predict others dislike dissimilar options despite recognizing that they, themselves, simultaneously like dissimilar things. For example, people readily indicate enjoying dissimilar vacation destinations (e.g., lake and city) and dissimilar movies (e.g., documentaries and thrillers), but predict that—for others—a preference for one precludes enjoyment of the other.

Five experiments demonstrate that people sensibly expect others to like similar products, but erroneously expect others to *dislike dissimilar* ones (Studies 1 and 2). While people readily select dissimilar items for themselves (particularly if the dissimilar item is of higher quality than a similar one), they fail to predict this choice for others (Studies 3 and 4)—even when monetary

rewards are at stake (Study 3). We show that the tendency to infer dislike from dissimilarity is driven by a belief that others have a narrow and homogeneous range of preferences (Study 5).

THEORETICAL BACKGROUND

Predicting Others' Preferences

The ability to accurately predict others' preferences is important across a number of contexts. We predict others' preferences when we buy gifts or make joint household decisions. Agents, such as realtors and money managers, are tasked with curating choice sets based on their clients' implicit and explicit desires. In times of medical crisis, physicians and next-of-kin are called on to make life-or-death decisions based on what they infer the patient's preferences to be. And of course, predicting customers' tastes is an essential marketing task: whether it is selecting the right merchandise to stock or promotions to run, businesses face the perennial task of making accurate predictions about their clientele.

Very often, such predictions must be made with little or no explicit information about others' actual preferences, leaving people to rely on previously observed behaviors: how has the other person chosen in the past? Consequently, previous choice information becomes influential in preference prediction. Therefore, when faced with the above question—"what would you predict about the consumer's preferences for Widget B?"—we suggest that people engage in a discernible evaluative process, such that they: 1) attend to available information about the consumer's related choices (in this case, to the consumer's selection of Widget A); 2) evaluate the relative similarity of Widgets A and B; and 3) conclude that Widget B is liked if similar, but disliked if dissimilar (Figure 11).

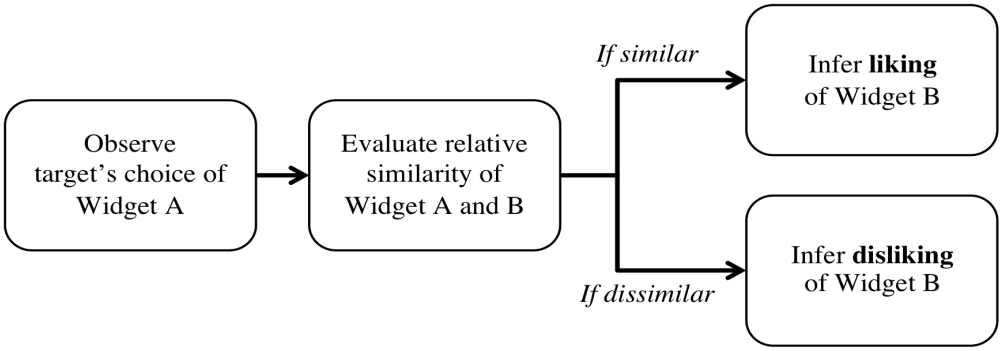

Figure 11: Conceptual model for predicting a target's preferences for unchosen Widget

To elucidate, consider the first node in this evaluative process: observing a target's choice of Widget A. People readily observe one another's choices—the cars they drive, the food they eat, the clothes they wear—and unfailingly conclude that others *like* the options they choose: if a consumer chooses Widget A, the consumer must like Widget A. This intuition is robust and unyielding, and persists even when the chosen option is objectively undesirable (e.g., Jones and Harris 1967; Miller and Nelson 2002). In essence, observers use others' known choices as meaningful benchmarks against which unchosen options are compared.

This observation of choice is a necessary precondition for the proposed evaluative process; it provides a focal item against which to compare unchosen options. Without knowing that Widget A was chosen, observers would have no reason to use similarity as a basis for their predictions, and may simply project their own preferences (i.e., "I like Widget B; I predict he will like Widget B"), or evaluate Widget B on its own merits (i.e., "Widget B is good; I predict he will like Widget B"). However, we suggest that a known choice triggers the process outlined in Figure 11, such that people evaluate Widget B by comparing it to Widget A. Consequently, even if observers would find Widget B appealing and likeable in its own right—and would, in

isolation, predict that it is liked—the known choice of Widget A can significantly sway their predictions.

Central to our account, this comparison process—outlined in the second and third nodes of Figure 11—systematically shapes the specific predictions people make about others' preferences. Research has shown that when making comparative judgments, the starting point—or "subject"—of the comparison carries special weight (Houston, Sherman and Baker 1989); people evaluate new options relative to the subject (or, in our model, to the known choice of Widget A) and pay particular attention to which attributes the new option lacks or shares. Therefore, the chosen option's features become a kind of *de facto* checklist; people note which attributes Widget B has in common with Widget A, and base their evaluations of Widget B on the degree of overlapping features (Houston and Sherman 1995; Houston, Sherman and Baker 1989).

These overlapping or common features also affect perceived (dis)similarity; items appear more similar as a function of their shared features, and less similar as a function of their distinctive or unique features (Tversky 1977). Therefore, with respect to observers' predictions, an unchosen option that shares many features—and thus possesses many of the attributes on the desired "checklist"—will be perceived as similar and presumed to be liked, while an unchosen option that shares few features will be perceived as dissimilar and presumed to be disliked.

The surprising aspect of these hypotheses is not that people expect others to like similar items (Lefkoff-Hagius and Mason 1993); indeed, this appears to be a sensible and correct inference (Linden, Smith, and York 2003). The more interesting question is: why do people predict that others must *dislike* dissimilar items? Even if a dissimilar option lacks attributes on the known choice's "checklist"—and therefore does not possess the qualities people assume

others value—it may still have features that are objectively desirable. But, as we argue, dissimilar items are not simply seen as neutral alternatives; people instead predict that they are actively rejected and disliked. Therefore, we suggest a more fundamental misperception drives this inference of active dislike: a tendency to underestimate the range and diversity of other people's tastes. In other words, people believe others' preferences to be more homogeneous or one-dimensional than they, in fact, are.

Researchers have documented a related tendency among groups, such that people infer out-group members are more alike than members of in-groups (e.g., Judd, Ryan and Park 1991; Linville, Fischer, and Salovey 1989; Park and Hastie 1987; Park and Judd 1990; Quattrone and Jones 1980). As a result of this phenomenon, the "out-group homogeneity effect," people perceive members of an out-group to lack interpersonal diversity and be "tightly bunched around the group central tendency" (Judd et al. 1991, p. 367). As an example, a woman might believe that "all men are alike" and, correspondingly, believe that "all men are very adventurous"—or that men are narrowly distributed or "tightly bunched" around the trait of adventurousness (Rubin and Badea 2012).

Where out-group homogeneity research has shown that people see out-groups as more *inter*personally homogeneous (e.g., "all men are alike"), we suggest people also perceive other *individuals* are more *intra*personally homogeneous (e.g., "all of that person's preferences are alike"). Particularly once a target's choice is known, people (mistakenly) believe all of the other person's preferences to be narrowly clustered around that single choice. For example, if an observer learns that someone chose to vacation at a lake destination, he may infer that person only likes outdoorsy vacations. We argue that this inference of homogeneity, in turn, drives the

inference of dislike: narrower ranges of preferences mean that more items fall outside the realm of "what is liked," and thus have a higher probability of being *dis*liked.

Although research has documented the persistent belief that out-group members are overly similar, it does *not* purport that this belief is always mistaken or erroneous (Judd et al. 1991). In the same vein, our investigation looks at the persistent belief that others' preferences are similar and narrow; we do *not* suggest that the belief is always incorrect. Undoubtedly, there are situations in which dissimilar may actually be indicative of dislike—for instance, someone who chose an environmentally-friendly Prius may actually hate gas-guzzling Hummers—but we document instances in which this belief is both wrong and costly (Studies 3 and 4).

Similarity and Stimuli Selection

Because perceived (dis)similarity is central to our investigation, it is critical to articulate what qualifies as "similar" or "dissimilar." Tversky (1977) proposed that similarity is a function of the features that are shared by two objects, and the features that are unique to one but not the other. Since then, a vast body of subsequent research—spanning cognitive, social, and consumer psychology—has highlighted the ways in which similarity judgments can be more complex and malleable (for a review, see Medin, Goldstone and Gentner 1993); for example, similarity may vary as a function of context and option set (Dhar and Glazer 1996; Tversky and Gati 1982), time horizon (Day and Bartels 2008), characteristics and attributes considered (Houston and Sherman 1995; Lefkoff-Hagius and Mason 1993), and mode of comparison (Dhar, Nowlis and Sherman 1999; Houston, Sherman and Baker 1989).

Taking these complexities into account, we carefully selected our experimental stimuli to represent similar and dissimilar options. We used the methodology employed by past research on

similarity (e.g., Boldry, Gaertner and Quinn 2007; Dhar and Glazer 1996; Lefkoff-Hagius and Mason 1993) and pretested our stimuli to gather global assessments of similarity (e.g., "On a scale from 1 to 7, how similar do you think these two items are?"). In addition, we directly manipulated the perceived similarity of two imaginary products (Study 2) based on Tverksy's (1977) feature-matching definition, ensuring that our results were causally related to perceived similarity, and not merely an artifact of the specific stimuli we chose.

OVERVIEW OF STUDIES

Five studies document when and why people infer dislike from dissimilarity. Studies 1 and 2 show the basic effect, demonstrating that people make predictions based on (dis)similarity. Studies 3 and 4 document the error in these predictions by showing that people do indeed enjoy dissimilar options for themselves—and readily choose dissimilar items that are of higher quality—but fail to predict that others do, too. Finally, Study 5 provides evidence of the explanation underlying the prediction error: the tendency to infer dislike from dissimilarity is driven by the belief that others have a narrow and homogeneous range of preferences.

Across the five studies, we measured our dependent variable in three different—but complementary—ways to ensure the robustness of our results. We asked participants to make predictions about whether a target *dislikes/is neutral about/likes* a given option (Studies 1 and 2), how a target would rate an option on a 7-point Likert scale (Studies 2, 3, and 5), and how a target would choose between two options (Studies 3 and 4). Further, we asked participants to make these predictions about options that were forgone both explicitly (e.g., consumer was deciding between Widget A and B and chose A; make a prediction about Widget B) and implicitly (e.g., consumer chose Widget A; make a prediction about Widget B). Regardless of how the question

was posed, and irrespective of whether the option was explicitly or implicitly forgone, the studies provide converging evidence that people infer others must dislike dissimilar things.

STUDY 1

By our account, when predicting someone's preferences for unchosen Widget B, observers evaluate how similar Widget B is to chosen Widget A, and predict that Widget B is liked if similar and disliked if dissimilar (Figure 11). As discussed above, one necessary precondition is that a target's previous choice is known—for instance, that observers know the target selected Widget A—such that there is a focal item against which to compare unchosen options. In other words, without knowing Widget A had been chosen, observers will have no reason to use similarity as a basis for their predictions, and may predict that both Widget A and B are liked—even if they are dissimilar.

Study 1 demonstrated this dynamic by comparing a condition in which a target's previous choice was *known* to a condition in which a target's previous choice was *unknown*. We used a real-world forum for the sharing and observing of people's choices—Facebook status updates—and showed participants a post from a hypothetical friend, Joe Smith. All participants had to predict whether Joe liked or disliked two unchosen vacation options (city and mountains); in one condition, participants had no information about Joe's vacation preferences, while in another, participants learned that Joe had previously chosen to vacation at a lake. We examined how participants' predictions of Joe's preferences changed as a function of both (dis)similarity, and whether or not the previous vacation choice was known.

Procedure

Pretest. The stimuli in Study 1 were three vacation destinations: lake, mountains, and city. In the known choice condition, Joe Smith selected the lake; therefore, we selected mountains and city to appear similar and dissimilar, respectively, to that known choice. In a pretest on Mechanical Turk (N = 89), we asked participants to rate the extent to which different vacation options were similar to each other (1 = *not at all similar* to 7 = *very similar*). The results confirmed that lake and mountains were perceived as significantly more similar (M = 3.31, SD = 1.72) than lake and city (M = 1.69, SD = .98; t(88) = 8.73, $p < .001$).

Design. Study 1 was a two-condition, between-subjects design. Online participants (N = 205, 59% male; M_{age} = 31.1, SD = 10.0) were either given (*known choice* condition) or not given (*unknown choice* condition) information on another person's vacation choice. To start, all participants saw the Facebook status update of a hypothetical friend, Joe Smith. In the unknown choice condition, the status update simply read: "Just booked a vacation!" In the known choice condition, the status update read: "Just booked a vacation! Headed to a lake."

Prediction task. All participants predicted Joe Smith's preferences for two additional vacation destinations: mountain (similar) and city (dissimilar). Participants indicated their prediction by completing two sentences, "Joe ___ city vacations" and "Joe ___ mountain vacations," with one of three options: *dislikes*, *is neutral about*, or *likes*.

This and all subsequent experiments concluded with basic demographic questions. We set the desired number of participants at the outset of the experiment, targeting recruitment of at least 100 participants per cell; for our lab sessions, slightly fewer participants showed up than were expected. We did not analyze the data until the final number of participants was reached. No data were excluded and we report all measures and conditions. All stimuli are included in the web appendix.

Results and Discussion

Prediction task. The known choice—"Headed to a lake"—had a significant effect on participants' predictions (Figures 12A and 12B). We conducted a repeated measures logistic regression using prediction as the dependent measure; because we were most interested in the propensity to predict "dislike," we dummy coded predictions as 1 = *dislike* and 0 = *neutral* or *like*. There was no main effect of condition (B = -2.26, SE = 4.53, p = NS), but there was a main effect of destination (B = -6.93, SE = 1.44, $p < .001$), and importantly, also a significant interaction (B = 4.52, SE = 1.71, $p = .008$). Follow-up tests revealed that when people predicted Joe's preferences for the similar destination, mountain, "dislike" predictions did not differ as a function of condition (B = -.01, SE = .93, p = NS); in both conditions, only 2.9% of participants predicted that Joe dislikes mountains. However, when predicting Joe's preferences for the dissimilar destination, city, people were significantly more likely to predict "dislike" when they knew about Joe's previous choice of lake (B = 1.51, SE = .39, $p < .001$); at baseline, only 9.8% of participants predicted that Joe dislikes cities, but 33.0% predicted that Joe dislikes cities upon learning of his dissimilar choice of lake.

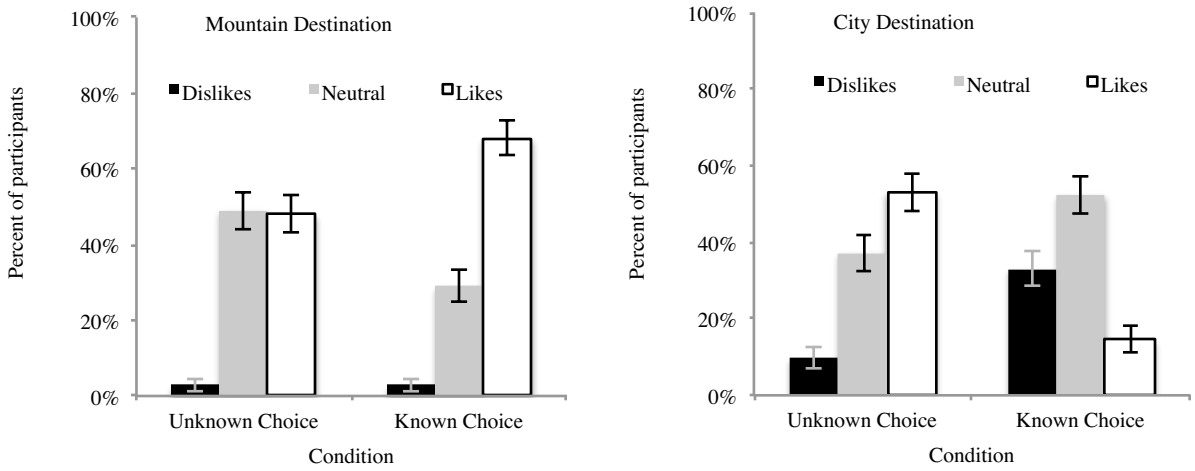

Figure 12A and 12B: Percent of participants predicting each preference for city destination (left) or mountain destination (right) (Clustered bars sum to 100%; error bars depict standard errors; Study 1)

Study 1 provides evidence that observers infer that—relative to a known choice—similar options are liked, while dissimilar options are disliked. In addition, the results also show that observers make significantly different predictions about others' preferences when a previous choice in that category is known. As a result, an option—like the city vacation—may appear perfectly likeable in one context (unknown choice) but unlikeable in another (known choice). This supports our account that a target's known choice is a necessary precondition for this inferential process, and demonstrates that observers form their predictions about unchosen options relative to known choices. (Having established this point, our remaining studies examine observers' predictions when previous choices are always known.)

STUDY 2

Study 2 replicated and extended Study 1 in two key ways. First, we ruled out the possibility that the effect was merely an artifact of the vacation stimuli employed in Study 1. To do so, we employed imaginary products, Widget A and Widget B, and directly manipulated their

perceived similarity based on Tversky's (1977) feature-matching definition of similarity. Second, we included additional dependent measures to provide converging evidence of the effect.

Procedure

Design. Study 2 was a two-condition, between-subjects design. Online participants (N = 297, 67% male; M_{age} = 30.7, SD = 9.7) read a scenario about a consumer, Jane, who was choosing between two products. In one condition, the two products were framed to appear relatively similar (*similar* condition); in the other condition, they were framed to appear relatively dissimilar (*dissimilar* condition).

All participants encountered the same basic scenario: "Suppose that Jane is choosing between two different products, Widget A and Widget B. Both Widgets can be described by 5 key attributes: price, size, shape, function, and quality." In the similar condition, participants learned: "Widget A and Widget B share 4 of 5 attributes. In other words, they have 4 attributes in common." In the dissimilar condition, participants learned: "Widget A and Widget B share 1 of 5 attributes. In other words, they have 1 attribute in common." In both conditions, participants were informed of Jane's choice: "Jane chooses Widget A."

Prediction task. Participants made three predictions about Jane's attitudes toward Widget B, the forgone option. First, participants predicted how much Jane likes or dislikes Widget B on a 7-point Likert scale (1 = *does not like at all* to 7 = *likes very much*). Next, as in Study 1, participants completed the sentence, "Jane ___ Widget B," by selecting one of three options: *dislikes*, *is neutral about*, or *likes*. Last, participants estimated how likely Jane would be to buy Widget B if Widget A were no longer available (1 = *not at all likely* to 7 = *very likely*).

Manipulation check. At the end of the survey, participants assessed the perceived (dis)similarity of Widgets A and B (1 = *very dissimilar* to 7 = *very similar*).

Results and Discussion

Manipulation check. As intended, participants perceived Widgets A and B to be less similar in the dissimilar condition (M = 2.88, SD = 1.47) than in the similar condition (M = 5.82, SD = 0.80; $t(295) = 21.44$, $p < .001$).

Prediction task. While Widget B was forgone in both conditions—rejected by Jane in favor of Widget A—participants made different predictions about Jane's preferences depending on Widget B's perceived (dis)similarity to Widget A. For the first measure, participants in the dissimilar condition predicted Jane liked Widget B significantly less (M = 3.25, SD = 1.23) than participants in the similar condition (M = 4.70, SD = 1.17; $t(295) = 10.37$, $p < .001$).

For the second measure, participants predicted whether Jane *dislikes*, *is neutral about*, or *likes* Widget B. The overall chi-square test revealed a significant main effect of condition on preference prediction ($\chi^2(2) = 78.492$, $p < .001$; Figure 13). Most participants in the dissimilar condition (61.5%) predicted that Jane dislikes Widget B; in contrast, few (14.8%) participants in the similar condition predicted that Jane dislikes Widget B ($\chi^2(1) = 68.76$, $p < .001$).

Finally, when asked to imagine that Widget A was unavailable, participants in the dissimilar condition thought Jane would be less likely to buy Widget B (M = 4.11, SD = 1.57) than participants in the similar condition (M = 5.81, SD = 1.04; $t(295) = 11.02$, $p < .001$), suggesting that different beliefs about Jane's preferences also translated into different beliefs about Jane's future behavior.

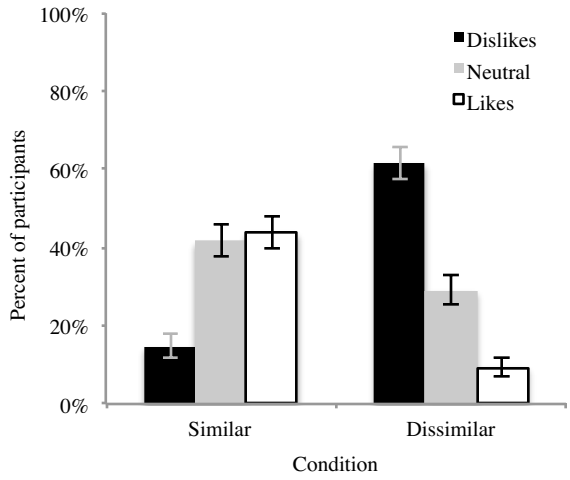

Figure 13: Percent of participants predicting each preference for Widget B (Clustered bars sum to 100%; error bars depict standard errors; Study 2)

All participants were given the same information about Jane's preferences: she had selected Widget A over Widget B. However, as expected, participants made significantly different judgments about forgone Widget B depending on its relationship to chosen Widget A. Therefore, Study 2 makes two key contributions. First, it rules out the possibility that Study 1's results were simply an artifact of the specific stimuli we used, instead showing that the effect emerges even in a minimalist setting using a clean and crisp manipulation of similarity. Second, by documenting the effect using three different dependent measures—the sentence completion task from Study 1, a scale measure of liking, and a prediction of future behavior—Study 2 provides converging evidence that people use dissimilarity to predict dislike.

So far, we have shown that observers make different predictions when they know (versus do not know) what someone else previously chose (Study 1), and more importantly, that they evaluate new options based on the relative (dis)similarity to that known choice (Studies 1 and 2). In Studies 3 and 4, we tested whether these inferences are accurate.

STUDY 3

Studies 3 and 4 tested the accuracy and strength of the belief that others dislike dissimilar things. To test the accuracy of this belief, both studies collected both actual and predicted preferences: participants either indicated their own preference (*self* condition), or—as in Studies 1 and 2—predicted someone else's preference (*observer* condition). Comparing actual and predicted preferences provided a measure of accuracy.

To test the strength of this belief, both studies asked participants to make a trade-off between similarity and quality. Specifically, participants had to choose (either for themselves or someone else) between a three-star movie in a preferred genre and a five-star movie in a dissimilar genre. In addition, Study 3 also used an incentive compatible design with a monetary bonus awarded for predictive accuracy. We anticipated that participants would choose the higher rated but dissimilar movie for themselves, but—despite the quality trade-off and monetary incentive—would predict others to make the opposite choice.

Procedure

Design. Study 3 was a two-condition, between-subjects design. Lab participants (N = 178; 52% male; M_{age} = 22.6, SD = 3.8) completed a study at a northeastern U.S. university lab in exchange for $20 and the possibility to earn a $2 bonus. Upon entering the lab, participants were randomly placed in dyads; one partner in the dyad reported his/her own preferences (*self*), while the other predicted his/her partner's preferences (*observer*). Dyads were seated on opposite sides of the room and remained anonymous to one another; their only interaction was through paper exchange via an experimenter.

Establishing a known choice. To begin, participants in the self condition indicated on a piece of paper which of two movies—a five-star thriller and a five-star documentary—they would prefer to watch. An experimenter collected the papers and delivered them to the partners who had been assigned the role of observer. Thus observers saw both the question and how their partner had answered. Both participants in the dyad also entered the preferred movie choice into the computer.

Ratings task. Next, we measured participants' actual and predicted ratings. All participants provided ratings of both genres (thrillers and documentaries) on a 7-point Likert scale (1 = *not at all* to 7 = *very much*). Specifically, participants in the self condition were asked: "Generally speaking, how much do you like each genre?" Participants in the observer condition were asked: "Generally speaking, how much do you think your partner likes each genre?"

Trade-off task. Next, both partners were presented with a trade-off choice task. Participants in the self condition were told: "Suppose that your first-choice movie is no longer available. Which movie would you select instead?" These participants had the option of a lower quality, three-star version of their preferred genre (as established in the known choice task) or a higher quality, five-star version of the unchosen, dissimilar option. Specifically, had a participant preferred thrillers, her trade-off choice would be between a three-star thriller and a five-star documentary; had the participant preferred documentaries, her trade-off choice would be between a three-star documentary and a five-star thriller.

In the observer condition, each participant predicted his/her partner's choice from the same two options (i.e., a three-star version of the preferred genre or a five-star version of the dissimilar, unchosen genre). Observers were told: "Suppose that your partner's first-choice movie is no longer available. Making a decision on your partner's behalf, which movie would

you select instead?" In addition, they were told: "If your answer matches your partner's selection, you will receive a $2 bonus upon checkout."

Preference strength task. As one final measure, participants in the self condition indicated how much they liked the two trade-off options (three-star version of preferred genre and five-star dissimilar option) after they had made their trade-off choice. These participants rated each option on a 7-point Likert scale (1 = *not at all* to 7 = *very much*). This measure was designed to elucidate the magnitude of the prediction error, determining whether "mistaking dissimilar for dislike" results in choosing something on someone's behalf that is slightly—or considerably—less preferred.

Results and Discussion

Ratings task. A 2 (between-subjects, role: *self* versus *observer*) x 2 (within-subjects, option: *preferred* versus *dissimilar* option) mixed ANOVA revealed two main effects. Unsurprisingly, regardless of role, the preferred option (as indicated in the known choice task) was rated significantly higher than the unchosen, dissimilar option ($M_{preferred}$ = 5.46, SE = .07; $M_{dissimilar}$ = 3.47, SE = .10; $F(1, 176) = 439.08$, $p < .001$). There was also a main effect of role (M_{self} = 4.83, SE = .10; $M_{observer}$ = 4.10, SE = .10; $F(1, 176) = 27.20$, $p < .001$).

Critically, these main effects were qualified by a significant interaction ($F(1, 176) = 23.56$, $p < .001$). Specifically, when rating the preferred option, the difference between self and observer ratings was small in magnitude and marginal in significance (M_{self} = 5.60, SD = 1.00; $M_{observer}$ = 5.33, SD = .89; $t(176) = 1.91$, $p = .06$). However, observers' predicted ratings for the unchosen, dissimilar option were substantially and significantly lower than their partners' actual ratings ($M_{observer}$ = 2.88, SD = 1.01; M_{self} = 4.07, SD = 1.51; $t(176) = 6.18$, $p < .001$). Assessed

another way, self participants reported a smaller difference in ratings between their preferred and unchosen options (M = 1.53, SD = 1.37) than observer participants predicted (M = 2.45, SD = 1.16; t(176) = -4.85, $p < .001$). In addition, observer ratings for the dissimilar option were significantly below the scale midpoint (t(88) = -10.50, $p < .001$), while self ratings were not (t(88) = .42, $p = .68$).

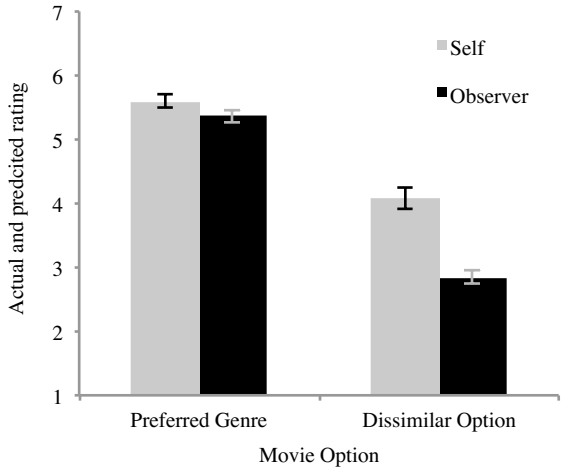

Figure 14: Self versus observer rating for movie genres (Error bars depict standard errors; Study 3)

Trade-off task. All participants were asked to choose—either for themselves or on behalf of their partner—between a three-star version of the preferred genre and a five-star version of the dissimilar, unchosen genre. While most (68.5%) participants chose the higher quality, dissimilar option for themselves, fewer than half (39.3%) of those in the observer condition predicted that choice ($\chi^2(1) = 15.29, p < .001$)—despite having been incentivized for accuracy.

Preference strength task. Participants in the self condition not only selected the higher quality dissimilar option in the trade-off choice task; they also strongly preferred the higher quality, dissimilar option (M = 4.83, SD = 1.59) to the lower quality version of the preferred genre (M = 3.66, SD = 1.28; t(88) = 5.37, $p < .001$). Additionally, self participants rated the

higher quality, dissimilar option significantly above the scale midpoint ($t(88) = 4.93$, $p < .001$), and the lower quality version of the preferred genre significantly below the scale midpoint ($t(88) = -2.49$, $p = .02$). This result suggests that the prediction error is considerable in magnitude: observers not only mispredicted their partners' choices, but also predicted an option that was liked significantly less.

Study 3 demonstrated that observers both misestimated others' ratings and mispredicted others' choices, believing both that dissimilar items were liked significantly less than they actually were, and predicting a rejection of the higher quality, dissimilar option in favor of a lower quality, similar option. Moreover, observers made this error despite being financially incented for accuracy.

STUDY 4

Study 4 tested the specificity of the error. As in Study 3, participants chose between two movies of varying quality; however, unique to Study 4, we used both a pair of dissimilar genres (i.e., thriller and documentary) *and* a pair of similar genres (i.e., thriller and action adventure). We expected participants to believe that others can like similar options, and thus predict others would be willing to switch to a higher quality, similar genre. Critically, however, we anticipated that participants would fail to predict the switch when the alternative option was dissimilar.

Procedure

Pretest. Study 4 used a similar pair of movies (i.e., thriller and action adventure) and a dissimilar pair of movies (i.e., thriller and documentary). A separate pretest confirmed that

thrillers and action adventures are perceived to be more similar (M = 4.45, SD = 1.46) than thrillers and documentaries (M = 2.03, SD = 1.28; $p < .001$).

Design. Study 4 was a 2 (role: *self* versus *observer*) x 2 (option set: *similar* vs. *dissimilar*) between-subjects design. Online participants (N = 601; 64% male; $M_{age} = 32.7$, SD = 10.6) first encountered a known choice task followed by a prediction task. To manipulate role, we asked participants to either report their own preferences (*self*) or to predict someone else's preferences (*observer*). To manipulate option set, we asked participants to either choose between two similar movie genres or two dissimilar movie genres.

Establishing a known choice. Participants in the self condition indicated which of two movie genres they preferred ("In general, which movie genre do you prefer?"). Half of these participants (i.e., one-fourth of all participants) selected between two similar options, thriller and action adventure; the other half selected between two dissimilar options, thriller and documentary. To make the genres tangible, we included three well-known and contemporary examples of each (thriller examples: Gravity, Captain Phillips, Lone Survivor; action adventure examples: The Dark Knight, Terminator, Star Trek; documentary examples: Blackfish, Man on Wire, Inside Job.)

Participants in the observer condition were told which of the two genres someone else preferred. Specifically, observers read: "Study participant (Participant #62013) was asked the following question and provided the following answer." (In reality, we randomly assigned participants to receive different information about Participant #62013's preferences.) A screenshot of the survey question was displayed, and a radio button was filled in to indicate the previous respondent's selection. Half of the observer participants saw Participant #62013's selection between two similar options, thriller and action adventure (counterbalanced); the other

half saw Participant #62013's selection between two dissimilar options, thriller and documentary (counterbalanced). To ensure that participants had processed this information, observer participants were asked: "Which movie genre does Participant #62013 generally prefer?" They were not able to advance until they had correctly restated the choice.

Trade-off choice. Next, participants faced a choice, either making it for themselves ("Which movie would you rather see?") or predicting it for someone else ("Which movie do you think Participant #62013 would rather see?"). As in Study 3, participants chose between a three-star version of the preferred genre (as established in the known choice task) and a five-star version of the alternative genre.

Results and Discussion

Trade-off choice. A logistic regression revealed that the propensity to choose the higher quality movie depended on both role (B = 2.09, SE = .27, $p < .001$) and option set (B = 2.29, SE = .28, $p < .001$). More importantly, the predicted interaction between role and option set emerged (B = -1.86, SE = .37, $p < .001$, Figure 15). Follow-up tests revealed that when the option set consisted of similar movies (i.e., thriller and action adventure, left side of Figure 15), there was no difference in actual and predicted choice (self = 73.5% vs. observer = 68.7%; $\chi^2(1) = .86$, $p = .35$), demonstrating that participants correctly predicted others' trade-off choices among similar items. However, when the option set contained dissimilar movies (i.e., thriller and documentary), a significant prediction error emerged. While most people (64.2%) preferred the higher quality, dissimilar movie for themselves, most observers failed to predict this preference: only 18.1% correctly predicted that others would select the higher quality dissimilar movie ($\chi^2(1) = 65.78$, $p < .001$).

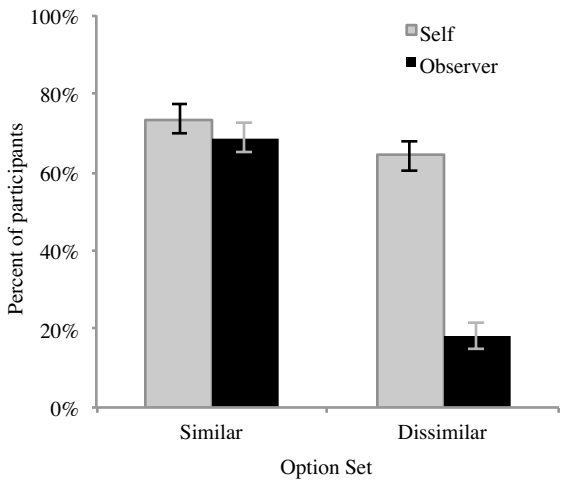

Figure 15: Percent of participants choosing the higher rated alternative genre (Error bars depict standard errors; Study 4)

While Study 3 documented the error in the prediction, Study 4 went further by demonstrating the specificity of the error: although people inaccurately predicted others' choices between dissimilar options (i.e., thriller and documentaries), their predictions were accurate when the options were similar (i.e., thriller and action adventure). Having demonstrated that observers mistakenly predict dissimilar items must be disliked, Study 5 shows why this effect occurs.

STUDY 5

We propose that the tendency to infer dislike from dissimilarity is driven by a belief that others have a narrow and homogeneous range of preferences, and therefore must not like dissimilar options. We draw on research from out-group homogeneity to support this account: where out-group homogeneity research documents the belief that out-group members lack *inter*personal variability (e.g., "all men are alike"), we suggest people also believe other individuals lack *intra*personal variability (e.g., "all of that person's tastes are alike"). In turn, this

perceived lack of variability drives people to think that others' preferences are mutually exclusive: if someone likes documentaries or lake vacations, he necessarily must not like dissimilar "opposites," thrillers or city vacations. Narrower and less variable preferences mean that more items fall outside the realm of "what is liked," and have a higher probability of being *dis*liked.

To test this account, Study 5 measured individuals' beliefs that others have homogeneous preferences using a distribution task akin to those used in out-group homogeneity research (Boldry, Gaertner and Quinn 2007; Judd et al. 1991; Linville et al. 1998; Park and Hastie 1987; Park and Judd 1990). As an example, Park and Hastie (1987) asked participants, "How many of these 100 [out-group] members would fall into each of eight boxes, which represent a continuum of the trait dimension *intelligence,"* where the boxes on the far left and right were labeled "stupid" and "intelligent," respectively; smaller ranges and tighter distributions reveal perceptions of homogeneity. We adapted this approach to directly measure participants' beliefs about the distribution of others' preferences in a particular domain: we asked participants to estimate the distribution of an average consumer's music preferences across multiple genres. Narrower distributions indicated a belief that others' preferences are homogeneous and one-dimensional, while broader distributions indicated a belief that others' tastes are more diverse and varied.

Anticipating variability in the strength of this belief, we then examined the role of the measured belief in how people make predictions. We replicated the basic design of Study 1, in which participants predicted how much a target likes a mountain and city destination after learning of his/her choice of a lake vacation. We expected that people who strongly believe others have homogeneous tastes would be more likely to predict that a dissimilar option is

disliked. Inversely, we anticipated that for people who believe others have diverse tastes, the effect would be mitigated and dissimilar options would be judged more favorably.

Procedure

Design. Study 5 was a two-condition, within-subjects design manipulating option (*similar* versus *dissimilar*). A second factor, *belief in preference homogeneity*, served as a measured independent variable. Online participants (N = 196; 59% male; M_{age} = 34.3, SD = 11.6) were all asked to take an observer role and make predictions about someone else's preferences.

Measuring the belief. To begin, all participants read the following: "Imagine that an average U.S. consumer was asked to indicate his/her music preferences by telling us how many songs from each music genre he/she would put in a 5-song playlist. We have disguised the specific music genres, but imagine the options included 5 very different kinds of music: Alternative Rock, Top 40 Pop, Classical, American Folk, and Hip Hop. How many songs do you think he/she would select from each genre?" Below the question, participants saw 5 textboxes with "Genre 1" through "Genre 5" labeled next to each; participants specified the number of songs from each genre, summing to five total songs. The distribution indicated participants' beliefs about others' preference homogeneity; for example, a distribution of [0, 5, 0, 0, 0] indicated a strong belief about another person's preference homogeneity (i.e., tightly clustered and unvaried tastes), while a distribution of [1, 1, 1, 1, 1], indicated a weak belief in preference homogeneity (i.e., dispersed and varied tastes).

Prediction task. On the following screen, participants were informed that they would next be asked to make predictions about a different person's vacation preferences. Rather than predicting for an average U.S. consumer, participants now made predictions for a "previous

survey respondent." (In reality, we created the answer this respondent had supposedly provided.) All participants read: "In a previous survey, we asked participants about their previous vacation choices. Participant #91811 provided the following response: Lake." As the dependent measures, all participants were asked to predict how Participant #91811 had answered the following two questions: "How much would you like or dislike taking a trip to a city destination?" and "How much would you like or dislike taking a trip to a mountain destination?" Participants predicted the target's response on a 7-point Likert scale (1 = *dislike very much* to 7 = *like very much*).

Results and Discussion

To measure the belief in preference homogeneity, we followed an approach used in out-group homogeneity literature and computed the standard deviation for each participant's allocation (e.g., Linville et al. 1989); higher "belief scores" represent stronger beliefs in preference homogeneity, while lower scores represent weaker beliefs in preference homogeneity. (For example, a relatively narrow distribution of [0, 1, 4, 0, 0] would be assigned a belief score of 1.73, while a relatively diverse distribution of [0, 1, 1, 1, 2] would be assigned a belief score of 0.71.) The maximum possible (and observed) value was 2.24, and the minimum possible (and observed) value was 0.00. The mean belief score was 1.06 (SD = 0.51).

To analyze the data, we used the method proposed by Spiller et al. (2013) to test for simple effects in a 2 (within-subjects) by continuous design. We created a contrast score for each participant by computing the difference in predicted ratings between the similar option, mountain, and the dissimilar option, city ($Z_{contrast} = Y_{mountain} - Y_{city}$). A regression was performed using the contrast score as the dependent variable, and belief in preference homogeneity as the independent variable. The results indicate a significant simple effect of option (B = 2.20, SE =

.29, $p < .001$), such that participants on average predicted higher ratings for mountain than city. The results also indicate a significantly positive slope for the measured variable, belief in preference homogeneity (B = .76, SE = .25, $p = .002$), indicating that the gap between ratings increased as beliefs in preference homogeneity strengthened. To further explore this interaction, we examined the slopes of predicted ratings for both mountain and city. While the slope of mountain ratings was positive but not significant (B = .26, SE = .16, $p = .10$), the slope of city ratings was both negative and significant (B = -.49, SE = .17, $p = .005$). In other words, people who tended to believe others' tastes are homogeneous were also significantly more likely to infer that dissimilar entails dislike.

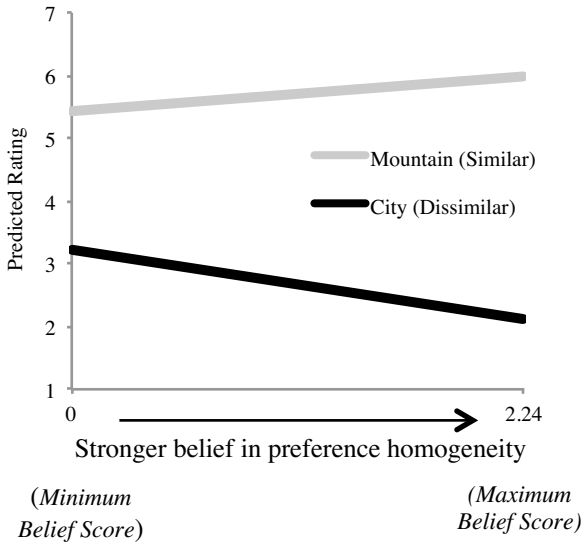

Figure 16: Regression lines by destination option (Study 5)

Study 5 provides evidence for why people infer dislike from dissimilarity. As our account suggests, people default to the belief that others have relatively narrow and homogeneous preferences, and thus predict that dissimilar items are disliked. Our results confirm that the

people who hold this belief most strongly are also the ones who predict the lowest ratings for dissimilar options. One possible limitation is that we elicited participants' preference homogeneity beliefs immediately prior to their preference predictions, which could have produced unintended spillover effects; however, such a limitation is likely mitigated by the fact that participants answered on behalf of two explicitly different targets (i.e., "an average U.S. consumer" versus "Participant #91811") about two different product categories using two distinct measures.

GENERAL DISCUSSION

Five studies demonstrated that despite liking dissimilar options themselves, people routinely predict that others do not. Relying on others' known choices, people predict that others will like similar options and dislike dissimilar ones (Studies 1 and 2). However, comparing predicted preferences to actual preferences, the error in this heuristic becomes clear: people readily select dissimilar items for themselves (particularly if the dissimilar item is of higher quality than a similar one), but fail to predict this choice for others (Studies 3 and 4)—even when monetary rewards are at stake (Study 3). Further, we show that the prediction error is driven by the misguided belief that others' preferences are homogeneous (Study 5).

Our basic paradigm—drawing inferences from someone's choice—is pervasive in everyday life; we routinely acquire information about other people's decisions and make judgments accordingly. Consequently, this prediction error may also be pervasive, leaving us to extrapolate more than is warranted by the bits of information we encounter. Further, these erroneous inferences happen even for options that are only implicitly (versus explicitly) forgone; for example, participants in Studies 1 and 5 knew only that the target had chosen a lake

vacation—not that the target had chosen the lake vacation *over* a city vacation—and yet, these participants still overwhelmingly inferred that cities were disliked because of their dissimilarity.

Particularly in contexts like recommendation-making and gift-giving, this bias may materially change what people select for someone else: if people are prone to making the wrong inferences about dissimilar options—even those options about which they have no explicit information—they may miss opportunities to diversify and optimize their recommendations or gifts. For instance, might a realtor forgo the opportunity to show a well-priced contemporary property after learning of her client's interest in a traditional home—mistaking a preference for the latter as an indication of dislike for the former? Or consider even more consequential contexts: knowing that a terminally ill patient had previously chosen an aggressive and experimental curative treatment, might a physician infer that her patient therefore rejects the more passive palliative option, and fail to initiate a more comprehensive discussion of all possible treatments?

Beyond actively mispredicting others' preferences, we suggest this mistaken inference has broader interpersonal implications, particularly in situations where it is important to gauge others' intentions. For example, consider what might happen if, in the course of a high-stakes negotiation, two dissimilar but equivalently viable options are put forth; would the other party's explicit endorsement of one option signal an implicit rejection of the other—even if that forgone option were only slightly less preferred? It is easy to see how this incorrect inference could lead to hostile or stalled negotiations. As a final example, one of us had a brief moment of panic when a senior colleague pointed out how *different* it was to collaborate with her versus another of his frequent coauthors; she fretted momentarily, assuming that if her colleague liked working with one, he could not possibly like working with the other, and worried that she might be the

dissimilar-and-therefore-disliked collaborator. But, recalling the takeaways of her own research, she realized (or at least hoped!) that she was underestimating her colleague's preferences for a diverse array of coauthors.

As a final thought, we propose that marketers, gift-givers, and recommenders should not constrain themselves to exclusively offering or suggesting similar products or options. While complex algorithms may steer marketers' product recommendations in Big Data-enabled settings (Linden, Smith and York 2003), there are myriad situations in which people give real-time, face-to-face, non-analyzed recommendations. Although it may be more difficult to predict which dissimilar items will have the most positive reception, there may be valuable upside to "going out on a limb" and suggesting something seemingly different. For example, someone who just watched a thriller may be bored by the obvious recommendation for yet another thriller—in which case the recommender risks appearing incapable of generating more nuanced and customized ideas; perhaps instead offering an acclaimed documentary could both pique consumer interest and make the recommender appear savvy and sophisticated. As our studies suggest, people are willing to sample diverse options, and in an environment where recommendations are routinely and recognizably based on item-to-item similarity, firms and individuals may risk appearing trite or unoriginal—or worse, like they have failed to recognize the variety of their clients' tastes. In this case, understanding that dissimilarity does not entail dislike may help people cater to the diversity of others' preferences.

REFERENCES

Amir, O., & Ariely, D. (2008). Resting on laurels: The effects of discrete progress markers as subgoals on task performance and preferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1158-1171.

Belk, R. W. (1994). Collectors and collecting. In S. M. Pearce (Ed.), *Interpreting objects and collections*. London: Routledge.

Boldry, J. G., Gaertner, L., & Quinn, J. (2007). Measuring the measures: A meta-analytic investigation of the measures of outgroup homogeneity. *Group Processes & Intergroup Relations, 10*(2), 157-178.

Botti, S., Orfali, K., & Iyengar, S. S. (2009). Tragic choices: Autonomy and emotional responses to medical decisions. *Journal of Consumer Research, 36*(3), 337-352.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty, 5*(4), 325-370.

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3*(1), 14-25.

Curley, S. P., Yates, J., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes, 38*(2), 230-256.

Danet, B., & Katriel, T. (1994). No two alike: Play and aesthetics in collecting. In S. M. Pearce (Ed.), *Interpreting objects and collections*. London: Routledge.

Dasgupta, N., Banaji, M. R., & Abelson, R. P. (1999). Group entitativity and group perception: Associations between physical features and psychological judgment. *Journal of Personality and Social Psychology, 77*(5), 991-1003.

Dawson, E., Savitsky, K., & Dunning, D. (2006). “Don't tell me, I don't want to know”: Understanding people's reluctance to obtain medical diagnostic information. *Journal of Applied Social Psychology, 36*(3), 751-768.

Day, S. B., & Bartels, D. M. (2008). Representation over time: The effects of temporal distance on similarity. *Cognition, 106*(3), 1504-1513.

Dhar, R., & Glazer, R. (1996). Similarity in context: Cognitive representation and violation of preference and perceptual invariance in consumer choice. *Organizational Behavior and Human Decision Processes, 67*(3), 280-293.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research, 24*(2), 215-231.

Dhar, R., Nowlis, S., & Sherman, S. (1999). Comparison Effects on Preference Construction. *Journal of Consumer Research, 26*(3), 293-306.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12*(1), 1-22.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics, 75*(4), 643.

Evers, E. R., Inbar, Y., & Zeelenberg, M. (2014). Set-fit effects in choice. *Journal of Experimental Psychology: General, 143*(2), 504-509.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics, 110*(3), 585-603.

Fox, C. R., Ratner, R. K., & Lieb, D. S. (2005). How subjective grouping of options influences choice and allocation: Diversification bias and the phenomenon of partition dependence. *Journal of Experimental Psychology: General, 134*(4), 538-551.

Frisch, D., & Baron, J. (1988). Ambiguity and rationality. *Journal of Behavioral Decision Making, 1*(3), 149-157.

Gneezy, U., List, J. A., & Wu, G. (2006). The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. *The Quarterly Journal of Economics, 121*(4), 1283-1309.

Grant, A. M., Campbell, E. M., Chen, G., Cottone, K., Lapedis, D., & Lee, K. (2007). Impact and the art of motivation maintenance: The effects of contact with beneficiaries on persistence behavior. *Organizational Behavior and Human Decision Processes, 103*(1), 53-67.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review, 103*(2), 336-355.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as Reference Points. *Cognitive Psychology, 38*(1), 79-109.

Houston, D. A., & Sherman, S. J. (1995). Cancellation and focus: The role of shared and unique features in the choice process. *Journal of Experimental Social Psychology, 31*(4), 357-378.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison of preferences. *Journal of Experimental Social Psychology, 25*(2), 121-141.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*(1), 1-24.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology, 61*(3), 366-379.

Kahn, B., & Wansink, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *Journal of Consumer Research, 30*(4), 519-533.

Kahn, B. E., & Luce, M. F. (2003). Understanding high-stakes consumer decisions: Mammography adherence following false-alarm test results. *Marketing Science, 22*(3), 393-410.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263.

Koffka, K. (1922). Perception: An introduction to the Gestalt-Theory. *Psychological Bulletin, 19*(10), 531-585.

Köhler, W. (1947). *Gestalt Psychology*. New York: Liveright Publishing Corporation.

Lefkoff-Hagius, R., & Mason, C. H. (1993). Characteristic, beneficial, and image attributes in consumer judgments of similarity and preference. *Journal of Consumer Research, 20*(1), 100-110.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 7*(1), 76-80.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57*(2), 165-188.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs, NJ: Prentice Hall.

Luce, M. (1998). Choosing to avoid: Coping with negatively emotion□laden consumer decisions. *Journal of Consumer Research, 24*(4), 409-433.

Luce, M. F. (2005). Decision making as coping. *Health Psychology, 24*(4, Suppl).

Luce, M. F., Bettman, J. R., & Payne, J. W. (1997). Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(2), 384-405.

Mcintosh, W. D., & Schmeichel, B. (2004). Collectors and collecting: A social psychological perspective. *Leisure Sciences, 26*(1), 85-97.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254-278.

Medvec, V. H., & Savitsky, K. (1997). When doing better means feeling worse: The effects of categorical cutoff points on counterfactual thinking and satisfaction. *Journal of Personality and Social Psychology, 72*(6), 1284-1296.

Miller, D. T., & Nelson, L. D. (2002). Seeing approach motivation in the avoidance behavior of others: Implications for an understanding of pluralistic ignorance. *Journal of Personality and Social Psychology, 83*(5), 1066-1075.

O'Connor, A. M. (1995). Validation of a decisional conflict scale. *Medical Decision Making, 15*(1), 25-30.

Park, B., & Hastie, R. (1987). Perception of variability in category development: Instance- versus abstraction-based stereotypes. *Journal of Personality and Social Psychology, 53*(4), 621-635.

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology, 59*(2), 173-191.

Payne, J. W., Laughhunn, D. J., & Crum, R. (1980). Translation of gambles and aspiration level effects in risky choice behavior. *Management Science, 26*(10), 1039-1060.

Pope, D., & Simonsohn, U. (2011). Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychological Science, 22*(1), 71-79.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717-731.

Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology, 38*(1), 141-152.

Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied, 1*(1), 34-49.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7*(4), 532-547.

Rubin, M., & Badea, C. (2012). They're All the Same!. . . but for Several Different Reasons: A Review of the Multicausal Nature of Perceived Group Variability. *Current Directions in Psychological Science, 21*(6), 367-372.

Sah, S., Elias, P., & Ariely, D. (2013). Investigation momentum: The relentless pursuit to resolve uncertainty. *JAMA Internal Medicine, 173*(10), 932.

Schwartz, B. (2004). *The paradox of choice: Why more is less*. New York: Ecco.

Simonsohn, U. (2009). Direct risk aversion: Evidence from risky prospects valued below their worst outcome. *Psychological Science, 20*(6), 686-692.

Smith, R. W., Faro, D., & Burson, K. A. (2013). More for the Many: The Influence of Entitativity on Charitable Giving. *Journal of Consumer Research, 39*(5), 961-976.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., & McClelland, G. H. (2013). Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression. *Journal of Marketing Research, 50*(2), 277-288.

Stewart, S. (1993). *On longing: Narratives of the miniature, the gigantic, the souvenir, the collection*. Durham, NC: Duke University Press.

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science, 3*(6), 358-361.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

Ubel, P. A. (2012). *Critical decisions: How you and your doctor can make the right medical choices together*. New York, NY: HarperOne.

West, P. M. (1996). Predicting Preferences: An Examination of Agent Learning. *Journal of Consumer Research, 23*(1), 68-80.