Attribution Bias in Consumer Choice

Kareem Haggag†
University of Chicago

Devin G. Pope*
University of Chicago and NBER

May 24, 2016

Abstract

When judging the value of a good, people may be overly influenced by the state in which they previously consumed it. For example, someone who tries out a new restaurant while very hungry may subsequently rate it as high quality, even if the food is mediocre. We produce a simple framework for this form of attribution bias that embeds a standard model of decision making as a special case. We test for attribution bias across two consumer decisions. First, we conduct an experiment in which we randomly manipulate the thirst of participants prior to consuming a new drink. Second, using data from thousands of amusement park visitors, we explore how pleasant weather during their most recent trip affects their stated likelihood of returning. In both of these domains, we find evidence that people misattribute the influence of a temporary state to a stable quality of the consumption good. We provide evidence against several alternative accounts for our findings and discuss the broader implications of attribution bias in economic decision making.

Keywords: misattribution, state-dependent preferences

*Contact information: kareem.haggag@chicagobooth.edu and devin.pope@chicagobooth.edu. We thank Marianne Bertrand, Seth Blumberg, Ben Bushong, Neil Davies, Tristan Gagnon-Bartsch, Christian Hansen, John List, George Loewenstein, Ed O'Brien, Ted O'Donoghue, Matthew Rabin, Joshua Schwartzstein, Anuj Shah, Richard Thaler, George Wu, and seminar audiences at Bonn University, Brandeis University, Carnegie Mellon University, Harvard University, University of California at San Diego, University of Chicago, University of Oregon, University of Southern California, University of Tennessee, and the 2016 Behavioral Economics Annual Meeting for helpful comments. We would also like to thank our contact at the Orlando amusement park.

1 Introduction

Across a variety of domains, people broadly recognize that their preferences are state-dependent. Food is tastier when hungry, outdoor vacation destinations are more pleasant in sunny, temperate weather, and going on a date is less enjoyable while sick. While standard models of state-dependent preferences assume that people properly appreciate the direction and degree to which their preferences vary with underlying states, both psychology and intuition suggest that this may not always be the case. One model of systematic errors related to state-dependent preferences is projection bias (Loewenstein, O'Donoghue, and Rabin 2003). Under projection bias, people inflate the degree to which they think their future tastes will match their current tastes, and are thus overly influenced by the state in which they make a decision. In this paper, we formalize and provide empirical evidence for a related, but distinct bias in state-dependent decision making: *attribution bias*.

When deciding whether to repeat a prior consumption activity, we draw on our past experiences with the item. However, in doing so, we may fail to account for the influence of the state under which we engaged in that prior consumption. We may overstate the quality of a restaurant that we last visited when we were extremely hungry, give an overly poor rating to a movie we watched while tired, or be less likely to use our next season ticket if it rained the last time we attended a baseball game. In short, we misattribute the influence of a temporary state to a stable property of the consumption good or activity. This type of attribution bias leads people to make systematic errors in economic decisions.

We start in Section 2 by providing a simple conceptual framework for attribution bias that draws on the model of projection bias provided by Loewenstein, O'Donoghue, and Rabin (2003). Suppose a person has state-dependent utility such that, in time t, she receives $u(c, s_t)$ from consuming c while in state s_t. When deciding whether to repeat a consumption activity, she forms a prediction $\tilde{u}(c, s_t)$ that lies between her true utility of consuming while in this state and the utility she received in her prior consumption experience in state s_{t-1}. If those states are different, then she will exaggerate the degree to which her current utility of consumption will match that of the prior experience. In our model, the predicted utility is a simple linear combination of the two utilities, with a parameter γ capturing the degree of attribution bias (i.e. the weight placed on the prior utility). We discuss important assumptions that our model makes about learning and consumption complementarity and also extend our basic framework to capture situations where an individual has experienced the same consumption good several times in the past.

In Section 3, we review the psychology literature on related types of attribution biases. Studies conducted over the last 40 years by psychologists have suggested that individuals misattribute temporary moods in life-satisfaction judgments, mislabel fear arousal as sexual attraction, and make misattributions in interpersonal explanations of behavior. By being somewhat broader in some dimensions of our model and more specific in others, we hope to build on this literature by providing a framework amenable to studying economic decision making. We also review the literature on projection bias and discuss a few papers within economics that have hinted at a role for misattributions in important contexts.

In Section 4, we provide evidence of attribution bias with respect to the commonly-experienced state of thirst. To do so, we ran an experiment with 427 subjects in which we randomly manipulated their thirst states (by assigning them to drink either $\frac{1}{2}$ or 3 cups of water) prior to having them consume a new mixed drink. We verified that our manipulation did in fact move thirst levels, and that the drinking experience exhibited state-dependence (i.e. subjects assigned to the $\frac{1}{2}$ cup treatment reported higher enjoyment levels for the subsequent new drink). To elicit our measure of attribution bias, we followed up with subjects a few days later while they were in a state of thirst orthogonal to their treatment condition in the baseline survey. Using both simple mean comparisons and by instrumenting baseline enjoyment with the randomized treatment, we find evidence of attribution bias. Subjects that were assigned to drink 3 cups of water report that they would be less likely to drink the mixed drink again if it were in front of them, less likely to make it again in the future, and would require a higher payment in order to drink it. To get a sense for the size of attribution bias, we map our regressions to the theory to estimate the degree of misattribution that occurred. Under our preferred set of comparisons, we find a rough estimate of $\hat{\gamma} = 0.7$. As an additional way of assessing the economic significance of attribution bias in this context, we perform a comparison of attribution and projection bias, finding that the former is at least as important as the latter in this context.

In Section 5, we provide further evidence of attribution bias in a field setting. We partnered with a major theme park operator in Orlando, Florida to study the influence of weather misattributions in how visitors evaluate Orlando theme park vacations. To do so, we designed a survey experiment that the operator administered to a sample of 9,340 prior visitors to the park. We asked visitors to recall the weather during their most recent visit, to evaluate how enjoyable they found the trip, and to state how likely they are to return and how likely they are to recommend Orlando theme park vacations to friends and family. The amusement park survey experiment supplements our lab experiment by testing for attribution bias using a different temporary state (weather instead of thirst), by using a larger sample of respondents, by using a more well-liked and familiar consumption good, and by allowing us to include additional measures and treatments to more carefully explore the mechanism (learning and de-biasing). The amusement park is also an ideal context to test for attribution bias since a large segment of customers buy their tickets in advance, mitigating the selection concern that could plague many other field settings. Using reduced-form and IV specifications, we find evidence of attribution bias: the pleasantness of the weather during a customer's most recent trip influences both her own future demand and her likelihood of recommending it to others. This effect is slightly attenuated, though similar, when subjects are randomly assigned to a treatment that provides detailed information on the weather patterns in Orlando. We also find that the effect of the weather during one's most recent visit diminishes with the stock of past experience with the good (consistent with our extended model). Finally, we find some evidence that participants can be de-biased by prompting them to think about the weather during their trip, prior to eliciting their evaluation of the trip.

Attribution bias is a large and important literature within psychology; however, it has received very little attention in economics. The goal of this paper is to help bridge this divide by studying attribution bias in economic decisions. We conclude the paper in Section 6 by providing a discussion of how attribution bias may matter in a broader set of economic domains.

2 Conceptual Framework

In this section, we sketch a simple model of *attribution bias* that closely follows the model of *simple projection bias* presented in Loewenstein, O'Donoghue, and Rabin (2003). To start, the agent has state-dependent utility so that at time t, her true instantaneous utility is $u(c, s_t)$, where c is consumption and s_t captures her state. As in the model of projection bias, the state parameterizes her tastes and is left intentionally broad. However, unlike the model of projection bias, we limit our focus to transient states at the time of consumption, including those external (e.g. weather, noise, smells) and internal (e.g. hunger, thirst, fatigue, mood) to the agent.[1]

An agent trying to predict her instantaneous utility from consuming c while in (current or future) state s_t will form a prediction that depends on her consumption in a prior, different state s_{t-1}. We allow for this prediction to fall between her true utility $u(c, s_t)$ and her realized utility in the prior state, $u(c, s_{t-1})$.[2] Our primary specification of attribution bias in predicted utility is a linear combination of the terms.

[1] In contrast, the model of projection bias allows the state to reflect the stock of past consumption or permanent health changes (e.g. losing a limb). This allows them to cover underappreciation of adaptation to major life changes, as well as the implications of habit formation under projection bias.

[2] Projection bias is: $\tilde{u}(c, s|s') = (1 - \alpha)u(c, s) + \alpha u(c, s')$ where s' is the current state and s is the state under which the future consumption occurs.

Definition 1. Predicted utility exhibits attribution bias if there exists a $\gamma \in [0,1]$ such that for all c, s_t, and s_{t-1}, $\tilde{u}(c, s_t) = (1-\gamma)u(c, s_t) + \gamma u(c, s_{t-1})$

Thus if $\gamma = 0$, the agent is unaffected by attribution bias, so that her predicted utility is correct: $\tilde{u}(c, s_t) = u(c, s_t)$. If she exhibits "full" attribution bias, then she her predicted utility is identical to what she received when consuming it in her prior state s_{t-1}.

Our two-period model of attribution bias is largely sufficient for the empirical examples in this paper, as we consider a new consumption experience (Section 4) and one that is relatively infrequent (Section 5). However, we can extend the model to include multiple prior consumption experiences, for example:

$$\tilde{u}(c, s_t | s_1, ..., s_{t-1}) = (1-\gamma)u(c, s_t) + \gamma \frac{1}{t-1} \sum_{\tau=1}^{t-1} \delta_\tau u(c, s_{t-\tau})$$

Where t indexes past consumption experiences, and $\delta_\tau \in [0,1]$ is a period-specific discounting term. Insofar as agents broadly sample the space of underlying states across their consumption experiences with a good, the influence of attribution bias on the accuracy of utility predictions should diminish with this experience. However, the degree to which the influence of attribution bias diminishes obviously depends on how prior consumption experiences are discounted. For the purposes of this paper, we remain agnostic about the specifics of the discounting function and whether it is context-specific. As a few examples, the function could weight all experiences equally (i.e. $\delta_\tau = 1 \; \forall \; \tau$), exponentially discount (i.e. $\delta_\tau = \delta^\tau$) or reflect order effects such as recency and/or primacy effects (e.g. $\delta_1 = \frac{t-1}{2}$,$\delta_{t-1} = \frac{t-1}{2}$, and $\delta_2 = \delta_3 = ... = \delta_{t-2} = 0$).[3]

Our model abstracts from some of the underlying psychology of misattribution. For example, we do not specify *when* the misattribution happens. In some cases, people may be aware of the role of hunger in inflating their enjoyment of a restaurant while they eat there, but make the misattribution in later considering whether to return (e.g. they vaguely recall the positive experience, but forget their underlying hunger state at the time). In other cases, the misattribution may happen in the moment; though the person recognizes that they are hungry, they do not appreciate how much this inflates their assessment of the restaurant. While the distinction between an immediate misattribution and misattribution-in-recall may

[3] Serial-position effects such as primacy and recency typically refer to how items presented toward the beginning or end of a set are remembered relative to other items presented in the same session (e.g. Murdock 1962). As an example, Garbinsky, Morewidge, and Shiv (2014b) argue that recency effects dominate primacy effects for gustatory experiences (i.e. enjoyment at the end of an eating experience, relative to enjoyment at the beginning, is more predictive of how soon people desire to repeat it). Researchers have also studied recency effects across longer periods of time. For example, Agarwal et al. (2013) present evidence of recency effects in consumer learning related to credit card fees and repayment that occur month to month.

have implications for de-biasing strategies and other interventions, grouping the two types of errors together allows for a simple description of the relationship between transient states at the time of consumption and later behavior.

It is worth noting two things that are not part of the simple model described above. First, we do not specify consumption complementarities from one time period to another. In other words, the actual instantaneous utility of consuming c in state s_t does not depend on s_{t-1}.[4] Attribution bias suggests that consumers may predict an association, but the model does not allow for actual utility to be affected by a previous state of the world. One could imagine situations where this assumption does not hold. For example, it is possible that having magnificent weather on a particular vacation creates memories in one's mind in a way that going back to that same vacation spot is actually more enjoyable (not just predicted to be more enjoyable) due to a form of consumption complementarity.[5] While we think this story is an unlikely explanation of our results, it is an alternative mechanism to attribution bias. However, we will provide some direct evidence that goes against this account. For example, in our empirical analysis about amusement parks, we not only ask if people who had good weather previously are more likely to return themselves, but we also ask if they would recommend Orlando, Florida as a nice vacation spot for others. We find strong effects of weather on willingness to recommend Orlando to others, and because the recommendation is for somebody else who hasn't had the same previous consumption experience, this cannot be explained by a consumption complementarity account.

Another potentially important aspect that is left out of the model are certain types of learning. First, it is possible that consuming c in state s_t provides information about the distribution of states of the world. For example, getting rained on in Orlando may change one's priors about how often it rains in Orlando and therefore change one's expressed interest in returning.[6] To test for evidence of learning about the distribution of states, we introduce an information treatment to the survey takers in our Orlando experiment that provides them with detailed information about typical weather in Orlando (more information than could have been obtained about Orlando weather than by simply visiting once).

While the learning treatment may rule out an explanation related to learning about the distribution of states, there remains a conceptually distinct alternative that concerns asymmetric updating about the stable quality of Orlando vacations. Specifically, people may

[4] The conditioned taste aversion that can form after consuming an item that causes nausea or other sickness, known as the Garcia effect (Garcia, Kimeldorf, and Koellino 1955), is one example of a consumption complementarity.

[5] Of course, it is also the case that having bad weather could be consumption complement to returning to the same vacation spot (because you didn't get to "fully experience" the location on your previous trip).

[6] Learning about the distribution of states is not a good alternative explanation for many other types of states such as thirst which we use in our laboratory experiment.

start with biased beliefs about the quality of the vacation and only update their priors in some states of the world. For example, if visitors have prior beliefs that overestimate the quality of Orlando vacations across all states, but *only* update their beliefs if they experience poor weather, then they will be less likely to return to Orlando if they experience poor weather. For any given consumption experience, priors could have been systematically wrong and thus could lead to evidence that looks like attribution bias. However, it is just as likely that this sort of bias would work against us finding evidence of attribution bias. It is difficult to ever fully rule out a learning story where people happened to have priors that would lead them to update in the direction that attribution bias also predicts. However, the fact that we find effects in the direction predicted by attribution bias in both our experiment and the Orlando field setting makes this learning story less likely. Further, in our Orlando survey, we find attribution bias even among the subset of survey respondents who indicated they had visited Orlando on many occasions. The fact that this subset has likely experienced many different weather states in Orlando also casts doubt on the ability of this learning story to drive the effects that we find.

3 Literature Review

In this section we review evidence related to attribution bias, with the goal of clarifying how our formulation of attribution bias fits in with the extant literature. We briefly cover the psychology literatures on feelings-as-information theory, misattribution of arousal, and interpersonal attribution biases (e.g. fundamental attribution error).[7] We then cover the cross-disciplinary literature on projection bias, as well as a handful of studies within economics that relate to misattribution.

The influence of temporary state variables on judgment and decision making has perhaps been most widely studied with respect to mood/affect.[8] Manipulations of affect have been shown to influence time preferences (Ifcher and Zarghamee 2011), risk aversion (Isen and Geva 1987), reciprocity (Kirchsteiger, Rigotti, and Rustichini 2006), and the endowment

[7]We limit our discussion to the three psychology literatures that we deem most relevant; however, attribution bias is a broad topic within psychology and there are other literatures that certainly relate to attribution bias that we do not cover. For example, we do not discuss anchoring and adjustment. We also do not cover the halo effect that has been described as, "a fundamental inability to resist the affective influence of global evaluation on evaluation of specific attributes," such as labeling an instructor's accent and appearance as irritating if that instructor was viewed in a video in which he was warm and friendly vs. cold and distant (Nisbett and Wilson 1977).

[8]For an extensive review of the literature on affect in consumer decision making, see Cohen, Pham and Andrade (2008). Noting the scope of the literature, they write, "Within psychology more generally, Schimmack and Crites (2005) located 923 references to affect between 1960 and 1980 and 4,170 between 1980 and 2000."

effect (Lerner, Small, and Loewenstein 2004). Within the mood-effects literature, the closest papers to our work are those that study affect-based misattribution. The seminal paper on this type of misattribution is Schwartz and Clore (1983), who studied the effect of mood manipulations on subjective measures of wellbeing.[9] They found that subjects who were asked to write about a happy (vs. sad) experience (N=61) or interviewed on a sunny (vs. rainy) day (N=84) reported higher levels of happiness and life satisfaction. While more recent research (e.g. Lucas and Lawless 2013) has failed to find reliable evidence that weather influences life satisfaction measures in much larger samples (e.g. N $>$ 1 million), the initial results spawned a large literature on "feelings-as-information" theory. Schwarz and Clore (1983) presented these results as evidence that people use momentary affective states as information about how their lives are going in general. In addition, in their first experiment, the authors were able to eliminate the mood effect by inducing subjects to attribute it to a separate factor (they were told that the soundproof experiment room could make them feel good or bad, depending on the experimental condition). Later research found that simply having subjects label their specific emotions was as effective as the more elaborate misattribution treatments in eliminating the effect of mood manipulations on life satisfaction measures (Keltner, Locke, and Audrain 1993).

While the "feelings-as-information" literature provides evidence that it is difficult for individuals to parse incidental from stable factors in making judgments, our formulation of attribution bias is distinct in a few specific ways. First, our formulation of attribution bias is less focused on overall affect and more focused on the complementarity of temporary states to the quality of the consumption experience (in a situation with state-dependent preferences). This distinction is important because in some domains the most relevant state to the consumption decision may work in an opposing relationship to its affective correlates. For example, hunger may make food more appealing; however, it may also be correlated with negative valence affect (e.g. that which produces aggressive impulses as in Bushman et al. 2014). If one were narrowly focused on affect-based misattribution, one might predict that tasting a food for the first time while hungry might reduce subsequent demand for it.[10] Our formulation of attribution bias on the other hand makes a very strong prediction that hunger during consumption will increase subsequent demand for the product because

[9] Gorn, Goldberg, and Basu (1993) extended Schwarz and Clore (1983) to a consumer decision-making context. They found that speaker systems evaluated after listening to liked (vs. disliked) music were rated higher overall and on hypothetical purchase intentions, but only if they were not first asked to rate how much they liked the music (which was interpreted as a way of making subjects aware of the source of their temporary mood).

[10] Another example of a case in which the consumption-state complementarity is more direct and counter to the implied mood effect would be a vacation destination in which the utility of consumption is higher when it is cloudy.

it is a direct complement to the consumption experience.[11] Another distinction of our work from the "feelings-as-information" literature is that most of the studies in this literature are focused on decisions that are coincident with the emotion, whereas our work, which applies attribution bias to repeated consumer decisions, is more focused on our remembered experience. So, while the "feelings-as-information" literature might ask how happy someone is with their life and see if the weather at the time of the decision impacts their response, our work – because of its focus on consumption experiences – is more interested in knowing if the weather that occurred during a previous consumption experience changes one's willingness to consume something again (even though the decision is now being made in a neutral weather state). Overall, our work is closely related to the "feelings-as-information" literature, but focused more on state-dependent preferences and consumption experiences than the impact of general affect. Our hope is that our simple model and empirical work using more of an economics framework can complement the existing work in this space.

A second, related literature concerns misattribution of arousal. The seminal study is Dutton and Aron (1974), who found that male subjects approached by a female surveyor immediately after crossing a suspension bridge were more likely to call her than subjects that were approached 10 minutes after crossing ($N = 45$). The authors provided a number of possible explanations, but were motivated by the Schachter and Singer (1962) two-factor model of emotion which argues that environmental cues are sometimes used to provide emotional labels for ambiguous states of arousal. The authors go further to suggest that cognitive relabeling of strong emotions (such as fear or sexual arousal) may occur even in situations in which the source is unambiguous. While the Dutton and Aron study and subsequent research in this vein have again highlighted an attribution bias with respect to incidental states, their focus is somewhat different than ours. Whereas individuals may at times make misattributions that operate through mislabeling one state (fear) as another (attraction), the misattribution bias that we formulate occurs without first mislabeling one state as another state. Once again, however, we are clearly drawing from the results and implications from the misattribution of arousal literature when creating a framework for misattribution in consumer decision making.

Within social psychology, attribution bias often refers to a bias in interpersonal explanations of behavior. For example, fundamental attribution error is the "general tendency to overestimate the importance of personal or dispositional factors relative to environmen-

[11] An additional value of focusing on how states directly complement the consumption experience in a state-dependent preferences sense as opposed to overall affect is that recent work has shown that similar valence emotions may operate differently. For example, Lerner, Loewenstein, and Small (2004) found that two emotions with the same valence (negative) produced opposing outcomes, with sadness increasing willingness to pay and disgust reducing it.

tal influences" when considering the causes of others' behavior (Ross 1977). More broadly, actor-observer bias is the idea that, "actors tend to attribute the causes of their behavior to stimuli inherent in the situation, while observers tend to attribute behavior to stable dispositions of the actor" (Jones and Nisbett 1971). While an important part of the literature on attribution bias in psychology, this interpersonal misattribution bias is quite different than the intrapersonal misattribution bias that we formulate and discuss in this paper which is more useful for consumer choice contexts.[12]

A few other papers within psychology have touched on related concepts. In a discussion paper, Ariely and Norton (2008) similarly synthesize a number of strands of the psychology literature related to misattribution. They take a broad view of the situational factors that may be misattributed (e.g. price anchors) and a specific focus on how the act of choosing a good under those situational factors may affect one's preferences (whereas we consider situations in which choice is not coincident with the situational factor). A pair of papers by Garbinsky, Morewedge, and Shiv (2014a, 2014b) consider the dynamics of consumption experiences and liking vs. wanting in predicting future consumption timing. In the 2014b paper, the authors were concerned with why recency bias affects when a food is consumed again. Specifically, they hypothesize that if the last part of a consumption experience is very positive, people will want to consume the product sooner. In Experiment 2 of that paper, the authors manipulate how positive the last part of a consumption experience is by assigning some subjects to a control group that drank a cup of grape juice, while subjects in the treatment ("reset") group drank the same amount of juice but then had a break, ate crackers, and then took one last sip of grape juice. Subjects in the treatment group reported higher enjoyment at the end of the session, and upon follow-up the next day also reported that they would prefer an earlier delivery date for a free half-gallon of the grape juice. Insofar as delivery timing is a reflection of the strength of one's preference, this study suggests that attribution bias, coupled with recency effects, can have an impact on future consumption decisions.

Within economics, psychological insights into state-dependent preferences have been primarily focused on projection bias. Projection bias, as formalized by Loewenstein, O'Donoghue, and Rabin (2003), is the tendency to overpredict the degree to which one's future preferences will resemble one's current preferences. Loewenstein, O'Donoghue, and Rabin (2003) pro-

[12]Within economics, attribution theory has been used to explain patterns of discrimination in different markets (Gneezy, List, Price 2012). Specifically, they draw on the strand of attribution theory that argues that attributions about the cause or controllability of certain behaviors "lead to emotional reactions that affect the likelihood of helping or punishing behaviors." The authors find that when the "discriminator believes the object of discrimination is controllable, any observed discrimination is motivated by animus. When the object of discrimination is not due to choice, the evidence suggests that statistical discrimination is the underlying reason for the disparate behavior."

vided a detailed overview of the psychology literature, provided a formal model, and derived the implications of projection bias for a life-cycle consumption problem (i.e. under-saving) and for durable good purchases (i.e. if returning goods is costly, then over-purchasing). Since its publication, a number of studies have tested for projection bias in the field, including purchases of cold-weather clothing (Conlin et al. 2007), vehicles (Busse et al. 2015), and outdoor movie tickets (Buchheim and Kolasak 2014), as well as cigarette addiction (Levy 2010) and gym attendance (Acland and Levy 2015). The most closely related study to our own within this literature is Uri Simonsohn's work on the effect of weather during campus visit days on the subsequent attendance decisions of 562 students admitted to a rigorous US university (Simonsohn 2009). Unlike other studies of projection bias, in which the transient state is coincident with the decision, Simonsohn studied the effect of a transient state during a sample experience (college visit day) on a decision made at a later date (whether to enroll). Though he described the study as one testing projection bias, he noted this distinction, drawing a parallel to the misattribution of arousal study by Dutton and Aron (1974) and writing that, "rather than projecting current utility, people appear to be projecting their remembered utility." He finds the somewhat surprising result that students who visited the college on a cloudy day were *more* likely to enroll. He explains this result by arguing that sadder moods induced by cloud cover make academically-demanding activities more appealing, and conditional on the college's academic attributes being more favorable than its non-academic attributes, the overall assessment of the university is thus higher under cloud cover. If the decision to enroll is effectively made at the time of the visit/interview, then this study is best categorized as a test of projection bias, albeit one with a bit of nuance with respect to multidimensional trade-offs. Our study thus adds to Simonsohn's work by making a precise distinction between attribution bias and projection bias, and testing directly for the former.

Finally, a few additional studies within economics bear mention. First, there is a growing literature on whether individuals experience-weight their past outcomes in guiding future behavior. For example past, personally-experienced outcomes have been shown to affect risk taking and stock expectations (Malmendier and Nagel 2011), inflation expectations (Malmendier and Nagel 2015), video-rental compliance (Haselhuhn et al. 2012), and investment in IPOs (Kaustia and Knupfer 2008).[13] These studies relate to our own in that they document the difficulty of parsing irrelevant information in judgment; however, the channel through which these mistakes likely operate is biased beliefs about underlying probabilities, rather than a misattribution of consumption utility. There is also theoretical work that relates to

[13] More broadly these also relate to studies on reinforcement learning, e.g. being more likely to repurchase a stock if one previously sold it for a gain (Strahilevitz, Odean, and Barber 2011) or to increase one's 401k savings rate in response to personally experiencing a high average or low variance return in a given period (Choi, Laibson, Madrian, and Metrick 2009).

attribution bias. For example, Schwarzstein (2014) models the role of selective attention in learning. In his model, an agent decides what information to attend to and subsequently fails to attend to other relevant information when updating their beliefs. Because the agent thus overweights the variables she attends to, she may exhibit behavior consistent with attribution bias. In contrast to his learning model, we narrowly focus on state-dependent utility where we have strong predictions about what information will be overweighted. Cunningham (2016) provides a model of hierarchical aggregation of information in the brain, and explores how such a 2-system model could explain a variety of behavioral biases, including an application to weather-related misattributions. Bushong & Gagnon-Bartsch (in progress) examine the implications of misattribution of reference-dependent utilities, i.e. neglecting the component of utility that is generated by deviations from expectations when assessing the quality of a good. Outside of the projection bias literature, perhaps the closest field studies within economics to our own are a series of papers which suggest that various key leaders are rewarded for luck, including CEOs (Bertrand and Mullainathan 2003), US governors (Wolfers 2007), and Indian politicians (Cole, Healy, and Werker 2012). Though each of these papers has plausible alternative explanations, they each discuss attribution bias on behalf of observers (shareholders or voters) as a possible mechanism. Beyond the difficulty of identifying attribution bias as the mechanism in these studies, they also concern judgments in which there is substantial complexity. For example, shareholders may not have adequate data or processing power to separate the relative influences of particular commodity price movements from CEO decisions in determining stock prices. Thus, rather than a psychological bias, these failures to parse luck from skill may reflect simple bounded rationality. Just as we would not say that an economist who runs a regression with omitted variable bias is exhibiting a psychological bias, the same may hold in these other contexts. Nonetheless, the studies are suggestive that attribution bias may have some role in these important economic domains; we contribute to this literature by cleanly testing for attribution bias and by showing that misattribution can occur even in low-dimensional judgments.

4 New Consumer Experience Experiment

To cleanly test for attribution bias, we ran an online experiment in which we manipulated subjects' thirst prior to consuming a new mixed drink. We then followed up with those same subjects a few days later to test whether their thirst during the first consumption experience influenced their preference for the drink while in an orthogonal (thirst) state.

4.1 Experimental Design and Data Description

We recruited 448 subjects from Amazon Mechanical Turk (MTurk) between September 7 and 9, 2015.[14,15] We posted the survey with the title "New Consumer Experiences Survey ($\sim$ 20 minutes)" and the description, "Answer a 15-20 minute survey on your preference and habits related to consuming liquids." The survey paid \$3 (i.e. a projected \$9/hour wage) and was only shown to MTurk workers that had a US residence, completed at least 500 prior tasks, and had a 95% approval rate on those tasks. Before accepting the task, subjects were informed that they would be required to consume a drink consisting of milk, orange juice, and sugar, and to answer survey questions.

The survey started with an informed consent page that asks subjects to confirm that they were (1) 18 or older, (2) had a set of ingredients that they were prepared and willing to consume during the experiment (and to commit to not consume anything else during the survey), and (3) had a camera and were willing and able to upload a photo of the ingredients with their Worker ID number.[16] Subjects were informed that participation was voluntary and that they could omit responses to any question, though each page included a prompt asking subjects if they intended to leave a question blank if it was left unanswered. Following the consent page, subjects were presented with a request to upload a photo of the full set of ingredients. They were asked to (1) measure 3 cups of water and to place it in one (or more) glass, (2) measure $\frac{1}{2}$ cup of water and place in a separate glass, (3) measure 1 cup of milk and place it in a third glass, (4) measure 1 tablespoon of sugar and 1/3 cup of orange juice and place each in separate cups, bowls, or other containers, and finally to (5) place all of the bowls and glasses on top of a piece of paper on which they had written their Mechanical Turk Worker ID (handwritten and visible).[17] The page included an example photo that we asked subjects to match as closely as possible. They were further reminded to not eat or drink any of the ingredients at that time, and that we would later ask them to consume the ingredients in a specific way.

Following the image upload, subjects were randomly assigned into one of two treatment conditions. Half of the subjects (N=224) were told to drink 3 cups of water, while the other

[14]Using MTurk allowed us to obtain a much larger sample than was possible in a laboratory setting. Of course there is a trade off is between sample size and experimental control. As we discuss below, we take several steps to try to ensure we have reasonable compliance with our protocol.

[15]448 reflects the number of subjects we kept in our sample. We received 456 responses in total, but dropped 7 duplicate survey completions and 1 respondent that uploaded an inaccurate photo.

[16]To comply with IRB, the following disclaimer accompanied the second condition, "Please do not participate in this study if you have allergies to any of these ingredients. If you have any health conditions that would be negatively affected by drinking 3 cups of water all at one time, please do not participate in this study." This was followed by, "You commit to not eat or drink anything else at any point during the survey, which will last approximately 20 minutes."

[17]See Appendix B, Figure B.1 for a screenshot of this upload page.

half were told to drink $\frac{1}{2}$ cup of water. Next, subjects were asked for their age, gender, weight, height, the last time they drank anything prior to starting the survey (14 categories), how thirsty they were (on a 7-point scale from "not at all thirsty" to "very thirsty"), how many glasses of water they typically drink per day, and how many glasses of any liquid (water, coffee, juice, soda, etc.) they typically drink per day. Subjects were then told to mix the milk with the sugar and orange juice and to stir together. On the next page they were told to drink half of the mixed drink before proceeding forward. They were then provided with a text entry box to answer the following prompt: "In two or three sentences, please describe your overall feelings right now. This could be about your mood, how drinking the first half of the drink made your stomach feel, or any other thoughts you would like to share." They were then asked to report how excited they were to drink the second half of the drink, on a 7-point scale ranging from "not at all excited" to "very excited". The goal of these two questions was to make sure that subjects fully appreciated and would remember the drinking experience. The second question also provided an intermediate check that our manipulation affected the utility of consumption (through excitement to complete the drink). On the next page, they were instructed to drink the second half of the mixed drink. They were then asked to answer on a scale from 1 (not at all enjoyable) to 7 (very enjoyable) how enjoyable drinking the mixed drink was. Finally, subjects were asked to dispose of their remaining water and to share any comments or questions they had regarding the survey.

On the morning of September 10, we sent an email to the 448 baseline participants to request their participation in a 2-minute follow-up survey that paid \$2. We kept the survey open for two days and received completed surveys from 427 (95%) of the baseline participants over this period.[18] Following the informed consent page, we had a single page that consisted of five questions with their order randomized. Four of the questions were: (1) "On a scale from 1 (not at all thirsty) to 7 (very thirsty), how thirsty are you right now?", (2) "On a scale from 1 (not at all enjoyable) to 7 (very enjoyable), how enjoyable was drinking the mixed drink we asked you to consume during our last survey?", (3) "If you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now, how likely would you be to drink it?", (4) "How likely are you to mix and consume the same mixed drink we asked you to prepare in the last survey?". We also asked a multiple price list question similar to (3) with the text, "Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so," which was followed by the amounts \$0.00,

[18]Subjects were quick to start the survey following our email. Roughly 70% of those that took the follow-up survey started it within 5 hours of the email, and $\sim 97\%$ started within the first 24 hours.

\$0.05, \$0.10, \$0.25, \$0.50, \$1.00, \$2.00, and \$5.00.[19] As with the baseline survey, subjects were allowed to enter any comments or questions on the last page.

Table 1 shows summary statistics split by our experimental treatment conditions. We have experimental balance on all of the non-outcome questions, as well as on attrition. Roughly 59% of participants were female, and on average they were 34, weighed 177 pounds, and reported consuming ~ 7 glasses of liquid a day. The survey took a bit longer than our projection, with an average completion time of 27 minutes.[20] The baseline and follow-up survey completions were spaced out by about 54 hours on average, with a range from 11 to 103 hours. Finally, we have experimental balance on the thirst levels in the follow-up survey; as would be expected, whether we assigned a subject to drink $\frac{1}{2}$ or 3 glasses of water does not predict their thirst in this latter survey.

4.2 Experimental Results (Reduced-Form and IV)

Our theory predicts that consumers will misattribute the utility of consumption in a prior state s_{t-1} to the utility of consumption in a later consumption experience while in state s_t. The most direct test of this theory is thus one in which we induce exogenous variation in that prior utility, $u(c, s_{t-1})$, and test whether that induced variation in utility carries over into predicted utility $\tilde{u}(c, s_t)$. The empirical specification that most closely captures this thought exercise is an instrumental variable approach in which we use the exogenous variation in s_{t-1} induced by our experiment as an instrument for the endogenous baseline enjoyment (our proxy for $u(c, s_{t-1})$). However, before turning to this primary specification, we first show that our treatment did affect our proxy for $u(c, s_{t-1})$, followed by a simple comparison of means for our proxies of $\tilde{u}(c, s_t)$ (i.e. the reduced form of our subsequent IV analysis).

Our baseline experimental condition affected subjects' consumer experience, as shown in Figure 1 and the first panel of Table 2. Panel A first shows a manipulation check of our experimental treatment. Subjects assigned to drink 3 cups of water were less thirsty, reporting thirst levels that were on average 1.29 points lower on the likert scale. That difference corresponds to roughly 83% of a standard deviation of the control ($\frac{1}{2}$ cup) group's thirst level. Whereas 61% of the treatment group reported a 1 ("Not at all") for the thirst question, only 20% of the control group reported this level. Panel B shows that the treatment was also reflected in the intermediate question of how excited the participant was to finish the drink, moving this measure by 0.71 points on average (41% of a control group SD). Finally, Panel

[19]See Appendix B, Figure B.3 for a screenshot of this question.

[20]Though subjects may likely have spent much less time on the survey itself, as this number is inflated by subjects that left the survey open while completing other tasks. This average also truncates the 52 surveys that were open for longer than a hour, replacing their survey length values with 60 minutes.

C shows the treatment effect on how subjects enjoyed the experience of drinking the mixed drink (our proxy for the utility of that first consumption experience), with a mean difference of 0.64 points (32% of a control group SD). In spite of subjects administering the protocol in their own homes and having the freedom to surreptitiously fail to comply, we were able to induce a reported change in their experience. The baseline results provide evidence of simple state-dependence with respect to the utility of consuming the mixed drink. We next turn to the follow-up survey to see whether there is evidence of attribution bias.

Figure 2 and the second panel of Table 2 show simple t-tests of whether our experimental treatment carried over into preference measures at a later, orthogonal state of thirst. Our first measure is our question of how likely they would be to drink the mixed drink if it were prepared and ready to drink in front of them now. On this measure, we find evidence of attribution bias significant at the 10% level (p-value = 0.07), with the control group reporting a higher likelihood of drinking of 0.35 points on average than the treatment group (17% of a control group SD). On the outcome of how likely they are to prepare and consume the drink again, we find a similar treatment effect of 0.25 points (14% of a control group SD), with a p-value of 0.14. As would be expected, this measure of "Would Make" suffers from more truncation than the "Would Drink" question, with 52% of the control group reporting a 1 on the 7-point scale, in contrast to 35% for the first question. Finally, we report results from the multiple price-list. Similar to the "Would Drink" measure, the question presents the prospect of the drink having already been prepared and being ready to consume; however, to deal with truncation and to elicit a monetary measure, we ask if they would drink it if they were paid various amounts of money. We find that subjects in the control group are more likely to respond, "Yes, I Would Drink it" for each of the monetary amounts posed, with the difference being significant at the 5% level for $0.00, $0.05, $0.10, $0.25, and $0.50. Further, we construct a minimum willingness to accept measure by taking the lowest value for which each subject reports that they would drink the mixed drink. Using this measure, the control group would drink for $0.26 less on average than the treatment group (15% of a control group SD), with the p-value of 0.12.[21,22]

Table 3 presents our primary analysis of the experiment results. The odd columns of

[21] Multiple price-list (MPL) questions are subject to a few criticisms. First, subjects may display multiple switching points (e.g. reporting "Yes" to $0.00, "No" to $0.05", and "Yes" for $1.00). In our sample, however, this not a strong concern, as only 2 of the 427 respondents displayed this pattern of response. A second concern is that the MPLs only elicit interval valuations (i.e. we know the subject is willing to accept $0.25 and not accept $0.10, but not the point estimate of the switch point), with this censoring potentially being most extreme at the ends of our ranges (i.e. $0 and $5). To address this concern, in Appendix Table A.1, we report results from an interval regression to account for this interval censoring (using *intreg* in *Stata*). Our point estimate in that specification is $0.36 and it is significant at the 10% level.

[22] Appendix Table A.1 also reports an ordered probit specification of the WTA measure, which also shows a lower WTA for the control group, with the coefficient significant at the 5% level.

Panel A are equivalent to the t-tests reported in Table 2, while the even columns show that we obtain similar results when we control for all of the variables summarized in Table 1. Panel B reports results from using the treatment as an instrument for the baseline enjoyment levels.[23] The identifying assumption is that treatment only affects the follow-up preference measures through the channel of baseline enjoyment. Using this IV approach, we find that a 1-point increase in baseline enjoyment induced by our experimental treatment results in a 0.55 point increase in the likelihood that a subject would drink the mixed drink (p-value = 0.01). Similarly we find that a 1-point increase in baseline enjoyment results in a 0.40 point increase in the likelihood that they would make and consume the mixed drink (p-value = 0.08) and a \$0.42 decline in their minimum WTA (p-value = 0.08).

Finally, Panel C of Table 3 shows results from using all of the baseline variation in the underlying state (thirst) as an instrument for baseline enjoyment. While our treatment did have a strong effect on thirst, it only moved it by 0.83 control group SDs on average. Using all of the baseline variation instead provides a more powerful instrument, as reflected in the higher first-stage F-stats (26.7 and 25.44). However, it also requires the stronger assumption that the natural variation in baseline thirst does not affect follow-up enjoyment through a channel other than baseline enjoyment. This assumption could be violated, for example, if people that come into the survey with higher thirst levels are "types" that are more likely to enjoy a sugary drink. While we know of no *a priori* reason to expect this particular relationship to hold, it is a possibility. In the even columns of the table; however, we also run the specification with controls, including one for the participant's thirst level in the follow-up survey. Thus, to violate our exclusion restriction with that particular example, we would need there to be "types" that are only different by their baseline thirst levels but not in their follow-up survey thirst levels. We however find similar estimates across the even and odd columns. Using the baseline thirst as our instrument, we find that a 1-point increase in baseline enjoyment results in a 0.73-point increase in likelihood of drinking (p-value<0.001), 0.77-point increase in likelihood of making and drinking (p-value<0.001), and a \$0.31 reduction in minimum WTA (p-value = 0.04).

4.3 Estimating the Magnitude of Attribution Bias

Section 4.2 showed that we can reject the null hypothesis that $\gamma = 0$; exogenous variation in the state during an initial consumption experience is misattributed when predicting utility in an orthogonal state. While we provide estimates of the variation in our likert scales and in terms of a minimum WTA measure, it may remain unclear whether attribution bias is

[23] The first-stage F-statistic for a model with and without the additional controls are 11.93 and 13.40 respectively.

economically significant. We have two methods for shedding some light on this question. In this section, we will relate our empirical measures to our theoretical framework to get a rough estimate for the attribution bias parameter within the context of this experiment. To begin, recall our theoretical and empirical models:

$$\textbf{Theoretical Model}: \tilde{u}(c, s_t) = (1-\alpha)u(c, s_t) + \alpha u(c, s_{t-1})$$

$$\textbf{Regression Model}: Demand_t = \eta + \beta^{IV} Enjoyment_{t-1} + \epsilon$$

Whereas our theoretical model maps utils into (predicted) utils, our regression instead maps enjoyment into measures of demand (e.g. whether the subject would consume the drink again). Given the incongruity in mappings, we need to make a series of claims to derive a rough estimate of γ from our estimates of β^{IV}. First, if there were no attribution bias ($\gamma = 0$), then predicted utility would be equivalent to true utility ($\tilde{u}(c, s_t) = u(c, s_t)$). Such an agent would be able to completely parse transient variation in an underlying state from her permanent preferences. Therefore any transient variation in her enjoyment (induced by our treatment) would have no predictive power for her demand, i.e. $\hat{\beta}^{IV} = 0$. Second, if there is full attribution bias ($\gamma = 1$), then predicted utility would be equivalent to experienced utility in the initial consumption experience ($\tilde{u}(c, s_t) = u(c, s_{t-1})$). Any variation in her enjoyment due to exogenous variation in her state would then have equivalent predictive power for demand as any variation due to her permanent preferences. We therefore need an estimate of how permanent preferences (translated into baseline enjoyment) map into demand. To get a benchmark for this mapping, we attempt to isolate baseline enjoyment induced by the permanent component of preferences from any transient influences. We do this by running the regression, $Demand_t = \eta + \beta^{\text{Permanent Preference}} Enjoyment_{t-1} + Thirst_{t-1} + \epsilon$, finding that $\hat{\beta}^{\text{Permanent Preference}} = 0.80$ (i.e. Table 4, Panel D, Column 1). This approach assumes that the only transient state factor that influences the utility of consuming the drink is one's thirst, so that the residual variation in $Enjoyment_{t-1}$ reflects only non-state-dependent preference.[24]

To summarize the above logic, when $\gamma = 0$, $\beta^{IV} = 0$ and when $\gamma = 1$, $\beta^{IV} = 0.8$. To get an estimate of γ from our estimate of $\hat{\beta}^{IV}$ we therefore need to make one more assumption. We assume linearity in equating the proportionality of the two β terms to estimate γ, i.e. $\frac{\hat{\beta}^{\text{IV}}}{\hat{\beta}^{\text{Permanent Preference}}} = \gamma$. Our preferred estimate of attribution bias uses the randomized treatment as our instrument for baseline enjoyment and "Would Drink" as the measure of demand, for $\hat{\beta}^{\text{IV}} = .55$. With these measures we estimate γ as ~ 0.7 ($\frac{\hat{\beta}^{\text{IV}}}{\hat{\beta}^{\text{Permanent Preference}}} = \frac{0.55}{0.80} = 0.69$).

[24]Other state variables could include mood, hunger, and stress; omitting these state variables could thus lead to either over or underestimates of $\hat{\beta}^{\text{Permanent Preference}}$, depending on the covariance structure of these variables with the included terms. Moreover, we are likely to have measurement error in $Enjoyment_{t-1}$ would would lead to attenuation bias in $\hat{\beta}^{\text{Permanent Preference}}$.

While our estimate suggests that attribution bias is large in this context, relaxing our assumptions may lead to unreasonably large bounds (e.g. if one is willing to posit a multitude of omitted variables).

4.4 Comparing Attribution Bias to Projection Bias

Another approach to gauging the importance of misattribution is to compare it to another psychological bias of import, e.g. projection bias. An ideal test would compare how predictions of future utility (i.e. $u(c, s_{t+1})$) are influenced by exogenous variation in one's state during the time of the decision (i.e. s_t) relative to exogenous variation in one's state during the time of the previous consumption experience (i.e. s_{t-1}). The former would provide a measure of projection bias (i.e. overly weighting one's current state) while the latter would provide a measure of attribution bias (i.e. overly weighting one's state during the prior consumption experience). Since we elicited thirst during both the baseline and follow-up surveys, we have natural variation in thirst that we can use as our measures of variation in s_t and s_{t-1}. In so far as these sources of variation are comparable and exogenous to other determinants of predicted future utility, they provide our regressors of interest for this comparison. The closest measure we have for predicted future utility is the "Would Make" variable. This question asks how likely subjects are to prepare and consume the drink again. We did not specify that this consumption experience would happen in the future. Thus, some subjects might have interpreted it as a prediction of the utility while still in their current state, while others may have interpreted it as a measure of future utility (across other states). If subjects used the former interpretation, then an influence of the follow-up thirst on the dependent variable would not reflect a bias, but rather simple evidence of state-dependent preferences. As a result, this test favors the projection measure relative to attribution, as it will reflect a combination of a bias and a standard determinant of their true utility of consumption.

Table 5 presents the comparison of projection bias and attribution bias. Column 3 shows the regression of "Would Make" on baseline and follow-up thirst in the full sample. We find that a 1-point increase in baseline thirst is associated with a 0.25-point increase in the likelihood of mixing and consuming the mixed drink again (alternatively, a 1 standard deviation increase in baseline thirst is associated with a 0.19 standard deviation increase in "Would Make"). The point estimate for thirst in the follow-up survey is 0.08 (i.e. 1 SD increase in follow-up thirst associated with a 0.08 SD increase in "Would Make"). A Chow test of the equality of the two coefficients produces a p-value of 0.14. One may be concerned that because half of the subjects in the baseline survey were asked to drink 3 cups of water, a 1-point movement in baseline thirst means something different than a 1-point movement in the follow-up thirst that had no experimental variation. This would potentially be the case if subjects were less likely to be biased by large and salient movements in their thirst levels. To address this concern, we repeat the analysis in the control group and find similar estimates. Overall, the results suggest that in spite of the exercise favoring projection bias, attribution bias appears to be of at least similar importance to projection bias in this particular context.

5 Weather Misattribution in Theme Park Vacations

In this section, we test for attribution bias in a natural field context, using variation in a different consumption state (weather instead of thirst). We designed a survey experiment that was administered to prior visitors to a large theme park in Orlando, Florida. We asked respondents to recall the weather during their most recent trip, to recall how enjoyable they found the trip, and to assess how likely they are to recommend Orlando theme park vacations to friends/family and how likely they are to return. Broadly, our hypothesis is that poor weather will not only affect how enjoyable the trip was, but that respondents will then misattribute the influence of this temporary weather to the fixed quality of the vacation destination, as reflected in return and recommendation likelihoods. We implement a few randomized treatments and examine heterogeneity to rule out alternative explanations.

5.1 Institutional Context and Experimental Design

We partnered with a large amusement park operator located in the Orlando, Florida area. The operator maintains a database of several million prior visitors that they periodically survey. Customers enter the panel through a variety of methods, including: (a) directly selecting into the panel through a website maintained by the operator, (b) being approached by the customer insights team at the park, or (c) taking an exit survey at the park. We designed a Qualtrics survey experiment that the operator emailed to this panel on October 26, 2015. The survey bore no marks of our involvement, and was administered entirely by the operator. To encourage participation, in accordance with their standard practice, the operator noted in the email that participants would be entered into a sweepstakes to win $250. The email also noted that the survey would take less than 5 minutes and that it concerned their most recent Orlando theme park vacation.

Because Orlando tourists often visit multiple parks during a vacation, we framed our survey as one concerning Orlando theme park vacations rather than one about the specific park with which we partnered. Moreover, asking subjects to recall the weather on a particular theme park visit may have been too demanding on their memory, and poor weather experienced during one park visit may spill over into another. Our survey included two randomized treatments and consisted of 12 questions spread over 3 sections. In the first section, we asked subjects: (1) the number of separate trips they had made to Orlando to visit theme parks, (2) during their most recent trip, how many days they spent visiting theme parks, (3) to recall the exact date of the last day of their most recent trip if they could,[25] (4) to select all parks that they visited during the trip, (5) to select the parks that they visited during the last day of the trip, (6) to select the parks that they visited during the first day of the trip. Between questions 3 and 4, half of the subjects were randomly assigned to receive a weather "information treatment". The goal of this treatment was to understand whether any effects of prior weather during the trip could be explained by simple Bayesian updating with the weather during one's most recent trip. While a single instance of weather is not particularly informative, subjects may have strongly updated their priors on the likelihood of poor weather based on their experienced weather. Such an explanation would be particularly compelling if collecting better weather information is high cost or low benefit. While individuals may have sufficient incentive to collect accurate weather data if they were booking a trip, this may not be the case during a hypothetical survey. We therefore provided subjects with very detailed and accurate information displayed in a simple table. The table of information was preceded by the following description: "As a short detour from the survey, we would like to provide you with some information on the kind of weather you can typically expect in Orlando. Here is the monthly average high temperature, low temperature, precipitation (inches), and the number of days with at least 0.01 inches of rain". In addition to the monthly averages, we also included a row with the yearly averages.[26]

Our other randomized treatment was the order in which we displayed the second and third sections of the survey. One of these sections consisted of a page in which we asked subjects two questions about the weather during their most recent trip: (1) "How pleasant was the weather while visiting theme parks during your last trip to Orlando?" (on a 7-point likert scale from 1 (not at all) to 7 (very)), (2) "How much did it rain while visiting theme parks during your last trip to Orlando?" (on a 3-point scale). The other section asked subjects to evaluate their trip and Orlando vacations more broadly, all on 7-point likert scales: (1) "How enjoyable was your theme park experience during your last trip to Orlando?", (2) "How likely are you to recommend an Orlando theme park vacation to your friends and family?", (3) "How likely are you to do another Orlando theme park vacation in the next 12 months?", (4) "How likely are you to do another Orlando theme park vacation ever

[25] The text was, "Please indicate the last day you visited a theme park during your most recent trip to Orlando. The exact date is important to us; please consult your records/calendar if possible. If you still do not know the exact date after checking your records, then please skip this question."

[26] See Appendix Figure B.4 for an screenshot of the page.

again?".[27] Half of the subjects were randomly assigned to see/answer the weather questions first, while the other half was assigned to first see/answer the trip evaluation questions. We randomized the order for two reasons. First, the strength of the correlation between weather and enjoyment of the trip may be influenced by an "assimilation" effect. For example, Strack, Martin, and Schwarz (1988) found that responses to the following two questions were much more strongly correlated if the first was asked before the second than vice versa, "How often do you normally go out on a date?" and "How happy are you with life in general?" In our cases, subjects prompted to recall the weather during their trip may use that specific aspect of the trip to inform their response to how enjoyable it was. A second reason to randomize the order of questions is the possible de-biasing effect of asking about weather before asking about return likelihoods. That is, while asking about weather first might yield a stronger correlation between weather and enjoyment, the portion of enjoyment explained by weather should have weaker predictive power for return likelihoods. This hypothesis is informed by Schwartz and Clore (1983) who found that asking subjects to first report the current weather conditions in their location before asking life satisfaction questions attenuated the relationship between the two.

The operator received 9,532 responses to the survey within one week of survey mailing date (i.e. by November 2, 2015). We dropped the 192 respondents that reported having never made a trip to Orlando to visit theme parks. Table 6 reports summary statistics on these remaining 9,340 respondents. Most respondents answered all questions in the survey, with 9,330 responding to the second and third sections in their entirety. Roughly 88% were able to recall the date of their most recent trip; among this group, participants took their most recent trip a bit under 1 year before the survey. On average, respondents made a bit over 8 separate prior trips to Orlando to visit theme parks, with the most recent trip taking roughly 5 days and covering 3 parks. As may be expected among a sample that agree to participate in repeated surveys, they are very fond of Orlando, maxing out most of our measures evaluating their trips and Orlando. Finally, the second panel of Table 6 includes the (self-reported) demographic variables provided by the operator from their prior records. Of note, roughly 30% of respondents are Florida residents and 22% are annual pass holders.

5.2 Results

As with the new consumer experiences experiment, our hypothesis is that people will misattribute the temporary, exogenous state they experienced during consumption to a fixed quality of the good under consideration. In this context, we predict that poor weather will (ra-

[27] We randomized the order of the questions in this section as well.

tionally) reflect itself in worse evaluations of the trip, but that individuals will then carry this over into their judgment of Orlando vacations broadly, resulting in lower stated likelihoods of return and of recommending the destination to friends/family. To test these predictions, we follow a similar approach to that of Section 4. We start by presenting the reduced-form results in which we regress our outcome variables (stated return and recommendation likelihoods) on measures of the underlying consumption state (how pleasant the weather was during the trip). We then map the data closer to our theoretical framework by using an instrumental variable strategy in which we instrument for the utility of consumption in the most recent trip experience (i.e. how much they enjoyed the trip) with the state variable. This IV approach requires that the stated pleasantness of the weather only affects the stated return likelihood through the channel of stated enjoyment on the most recent trip.

The four scatterplots presented in Figure 3 summarize our results in the raw data. The graphs plot responses to how pleasant the respondent found the weather during their most recent trip against enjoyment during that trip (our first-stage of the IV) and stated return and recommendation likelihoods (the reduced-form). For visual clarity, we take means within each value of the X variable for 5, 6, and 7, while we group together responses of 1, 2, and 3 (roughly 6.8% of observations) together with 4 for the "Less than 5" value. Figure 3(a) shows our significant first-stage in the raw data. Without any controls, we see that enjoyment during the trip increases by about 0.63 points ($\sim$ 72% of a standard deviation) as the weather assessment moves from the lowest to the highest category. We also see significant movements on the reduced-form figures, e.g. the likelihood of recommendation increases by 0.23 points ($\sim$ 36% of a standard deviation).

Table 7 presents our primary analysis. Columns 1, 4, and 7 of Panel A display similar estimates to Figures 3b, 3c, and 3d (without grouping values 1-4 in the independent variable). Across all estimates, we cluster our standard errors by the date that respondents report as the last day of their most recent trip (we thus drop the 1,008 subjects that did not report a date). We find that a 1-point increase in the pleasantness of the weather translates into a 0.061 point increase in the likelihood of recommending the destination to friends and family (i.e. a 1 standard deviation increase in weather pleasantness translates into a 0.13 standard deviation increase in likelihood of recommending). Similarly, we find coefficients of 0.120 for the likelihood of returning within the next 12 months and 0.039 for returning ever again (a 1 SD improvement in weather pleasantness translates into 0.089 and 0.071 SD increases in return likelihoods respectively). Adding demographic controls and week-of-the-year and year fixed effects produces very similar estimates in columns 2, 5, and 8. All of these estimates are statistically significant with $p < .001$. Panel B presents the effect of the portion of stated trip enjoyment induced by stated weather quality on these outcome variables. Again, we find significant evidence of attribution bias. A 1-point increase in enjoyment translates into a 0.334-point increase in the likelihood of recommending Orlando theme park vacations to friends and family (a 1 SD increase in enjoyment translates into 0.46 SD increase in recommendation likelihood). Return likelihood coefficients are 0.661 and 0.215 (for 0.32 and 0.26 SD translations). Finally, in Panel C we present regressions of (endogenous) trip enjoyment on the outcomes, while controlling for how pleasant they found the weather. These estimates allow us to perform a similar calculation of γ as the one presented in Section 4.4. If we use the recommendation likelihood or return (ever) coefficients, we estimate γ parameters close to 1 (e.g. $\gamma = 0.97$ using the Column 2 estimates). However, when we turn to the likelihood of returning within the next 12 months, we estimate γ to be even larger than 1.

Table 7 also demonstrates the results of our information treatment. As discussed in Sections 2 and 5.1, evidence of attribution bias could also be consistent with simple updating over weather patterns in Orlando using one's most recent experience. To address this confound directly, we provided half of the subjects with detailed information about weather patterns in Orlando prior to eliciting any measures of demand. One concern with this approach is that providing unprompted information about the weather might have a de-biasing effect similar to asking them about the weather prior to asking them about their trips. That is, in addition to giving them proper beliefs about the weather process, it could remind them that their experienced weather was a temporary state and thus they may be less likely to misattribute it to a stable quality of the Orlando vacation experience. Nevertheless, it provides a useful test, albeit one potentially less likely to find evidence of misattribution. Columns 3, 6, and 9 of Table 7 show the effects of weather on return and recommendation likelihoods within the random sample of subjects exposed to the information treatment. Across specifications 3 & 6, we find that our estimates are slightly attenuated relative to the full sample coefficients; however, we still find strongly significant effects of relatively similar magnitude. These results suggest that, while there may have been some personal-experience-weighted learning, attribution bias in this context is largely robust to that alternative explanation. Panel C of Table 8 sheds some light on the alternative *biased* learning explanation discussed in Section 2. While asymmetric updating of biased beliefs may be a plausible alternative account, it seems less likely in cases in which a prior visitor has experienced the consumption good across many different states.[28] However, columns 3, 6, and 9 of Panel C show that we still find significant evidence of attribution bias among customers that have visited Orlando theme parks on at least 6 separate, prior trips. Beyond providing evidence against the biased learning account, Panel C also suggests that the influence of attribution bias attenuates with

[28] We additionally control for whether the respondent is a Florida resident in Panel C, since frequent visitors are more likely to be locals, which poses its own selection concern.

the stock of past experience. Such a pattern would be consistent with a multi-period model of attribution bias that places sufficiently high weight on the most recent experience, even with a large number of periods.

Panel A of Table 8 examines the influence of order effects. In Section 5.1, we hypothesized that we would find strong assimilation effects, but possible de-biasing, if subjects were asked to evaluate the weather during their trip before evaluating how enjoyable the trip was as a whole. Panel A supports this hypothesis. We find a stronger first-stage f-statistic (239 vs. 134) in the sub-sample of subjects that were first asked about the weather, but a larger estimate of attribution bias. We find some evidence of a de-biasing effect on the recommendation question (0.283 vs. 0.376; $p = .15$), in the likelihood of ever returning (0.110 vs. 0.288; $p = .02$), and in the likelihood of returning within the next 12 months (0.316 vs. 0.585; $p = .13$). Finally, Panel B of Table 8 examines heterogeneity by state of residence. Visitors from Florida present two possible challenges for estimating the effect of temporary weather variation on subsequent return likelihoods. First, they pose a particular selection problem. Because it is likely easier for Florida residents to plan their trip to avoid poor weather (e.g. in the extreme case, an Orlando resident can check the forecast the same day and adjust plans accordingly), those that do attend in spite of poor weather may not value weather as heavily in their enjoyment of the vacation. We see some evidence consistent with this selection story in the relatively weaker first-stage f-statistic for Florida vs. Non-Florida Residents (97.8 vs. 226.7), though the strong first-stage among even Floridians suggest this isn't too much of a concern. Second, Florida visitors are more likely to be regular patrons, and thus we are truncated on our outcome measures (though we do control for the respondent's number of past visits in Panel B). This truncation can be seen in Table 8 by comparing the mean of the dependent variable in columns 4 vs. 5 (5.59 vs. 6.58). We are thus left with little variation within this sample on which to detect an effect. We end up finding the largest difference in our IV estimates for the outcome variable that has the biggest difference in truncation: returning in the next 12 months (0.530 vs. 0.178). In contrast, we find relatively similar effects for the likelihood of recommending the destination (0.327 vs. 0.293) and in the likelihood of ever returning (0.165 vs. 0.174), which both had similar dependent variable means across the two samples. Thus while focusing on non-Florida residents is a good way to get around the selection concern, it appears that selection effects are not a large concern even among Florida residents.

5.3 Robustness

Throughout the previous section, we used a self-reported survey measure of weather; however, this approach may suffer from at least two potential issues. First, our independent and dependent variables all use a common 7-point likert scale. While using a common scale has advantages (e.g. reducing cognitive complexity), some of the covariation between measures may reflect fixed response styles of participants (Podsakoff, MacKenzie, and Podsakoff 2012). For example, if we have a subset of respondents with a tendency to acquiesce across all measures (i.e. only using the upper portion of the scale, regardless of the content) as well as a subset that tend to disacquiesce (i.e. only using the lower portion of the scale), the correlation between variables may reflect this fixed effect rather than a true relationship between the underlying constructs. A second, related issue with correlating the likert scales is sequential anchoring (Hitczenko 2013). Responses to the first-encountered question may influence responses to subsequent ones. While similar to the "assimilation" effect discussed in the previous section, the underlying psychology could be broader. Instead of using the first question as a source of information about the second (e.g. using dating frequency to inform a life satisfaction question), it could simply reflect a decision to minimize variability to respond as quickly as possible. Although we find fairly similar treatment effects regardless of question order, that result doesn't entirely rule out sequential anchoring driving some of the observed correlation.

To address the first issue, we start by using the other subjective weather measure we collected, which partially breaks the common scaling. In addition to an overall assessment of weather pleasantness, we asked subjects to report "How much did it rain while visiting theme parks during your last trip to Orlando?" with four options: "Not at all", "A small amount", "A medium amount", and "A large amount". While this variable maintains a clear ordinal property, it less closely mimics the 7-point likert scales of the endogenous ("Enjoyment on Trip") and outcome variables. Prior research has shown that even minor changes to scales can reduce the degree of (possibly spurious) correlation between two variables (e.g. moving from labeling all points of a scale for both the X and Y variables, to only labeling the endpoints for one, as in Weijters et al. 2010). In Appendix Table A.3, we repeat Table 7, instead using "How Much Rain" to construct our instrument rather than "How Pleasant Was Weather"; we use three dummy variables for the categories of rain (with the omitted category being "Not at all"). Instead of OLS, we show the first-stage of the IV estimation in Panel C. We first see that the first-stage relationship is primarily driven by the "A large amount" category ($\sim$ 5% of respondents), with those participants reporting a trip enjoyment level roughly 0.3 likert points lower than participants that experienced no rain at all ($\sim$ 33% of the sample). Interestingly, participants that experienced "A small amount of rain" ($\sim$ 47%

of the sample) reported enjoying the trip more than those that experienced no rain. As this measure captures only one dimension of the weather (and also is less likely to suffer from common scale issues), it ends up being a much weaker instrument. In Column 2 of Panel B, we see a first-stage joint F-statistic for the instruments of 12.9 (in contrast to the F-statistic of 372.4 in Table 7). However, we find fairly similar IV estimates for the likelihood of recommending the destination (0.321 vs. 0.305) and the likelihood of ever returning (0.186 vs. 0.229); both of which remain significant, though with larger standard errors. In contrast, the estimates on the likelihood of returning within 12 months become insignificant (as does the estimate on the likelihood of returning ever again within the information treatment subsample).

An alternative approach that addresses both criticisms is to instead use objective weather data to construct our instruments. We asked subjects to recall the last day of their trips, and we can merge this to historical weather data. However, there are a few points of caution in taking this approach. First, we asked subjects to recall the date of their most recent visit, which took place roughly a year prior to the survey on average. The date they report is surely different than the actual visit date in many cases, producing a large amount of measurement error when matched to objective weather data. Second, subjects spent roughly 5 days visiting amusement parks during their trip. This presents a challenge for knowing which days are appropriate for shaping perceptions of the weather during the trip. Third, it's not clear which variables are important for shaping the likely heterogeneous mappings between aspects of the weather and how pleasant it was as a whole. For example, some people may find a light rain to be enjoyable, whereas others may find it to be a nuisance. Similarly, there may be important nonlinearities in the mapping between the objective measures of weather and the subjective assessment. With these limitations in mind, we used Weather Underground to collect historical daily weather data (January 1, 1995 to November 1, 2015) from the Orlando International Airport station (KMCO), originally sourced from the National Weather Service.

We start by merging the maximum temperature (°F), minimum temperature(°F), and total precipitation (inches) on the date reported for the question, "Please indicate the last day you visited a theme park during your most recent trip to Orlando". We did not ask the date on which the trip started, but instead asked, "During your most recent vacation in Orlando, how many days did you spend visiting theme parks (including Universal, Disney, Seaworld, etc.)?"; we subtract this number from the date reported as the last of the trip, thereby assuming that respondents visited the parks consecutively on all days prior to the last. We then merge the weather variables for each day in this interval. We use these measures to construct variables that characterize the weather during the entire trip and on the last day of the trip. To summarize the trip, we construct six measures. First, we characterize average conditions by taking the mean of the maximum temperature and the total precipitation (rainfall). We additionally denote days of the trip as "rainy" by an indicator for whether the rainfall exceeded 0.1 inches; we use the fraction of rainy days during the trip as our third characterization of average conditions. Second, we constructed three variables to capture the most extreme dates during the trip with respect to temperature and rain: the minimum temperature on the coldest day of the trip, the maximum temperature on the hottest day of the trip, and the total rainfall on the wettest day of the trip. Since temperatures and rainfall are likely to have differential effects by week of the year, we can additionally interact these measures with dummies for the month, giving us 72 total trip summary instruments. To characterize the last day of the trip, we limit our focus to the maximum temperature and the total rainfall. We split the temperature and rainfall into 20 dummy variables corresponding to 5-degree bins (Under 55, 55 to 59,...,90 to 95, Over 95) and quarter-inches (Under 0.25 inches, 0.25 to 0.49,...,2.00 to 2.25, Over 2.25 inches). In total, we have 90 potential instruments (after dropping one of the temperature bins and one of the rain bins), though we could make this set arbitrarily larger by allowing for more interactions and more flexible functional forms.

Weather thus presents a many instruments challenge. Naively including all of the variables results in a weak instruments problem, and estimates from two-stage least squares (2SLS) may be biased. However, our context provides no clear theory to guide instrument selection *a priori*. One way forward is to select the optimal instrument set through a principled data mining approach (i.e. one that finds instruments with predictive power, while guarding against overfitting and false discovery). The Least Absolute Shrinkage and Selection Operator (LASSO) method presented in Belloni, Chernozhukov, and Hansen (2010) provides such an approach, and it has been used by Gilchrist & Sands (2015) in an empirical setting similar to our own (using weather to instrument for movie viewership).[29] However, LASSO requires that a sparsity condition be satisfied; the conditional expectation of the endogenous variable given the full set of instruments needs to be well-approximated by a small subset of those instruments. If this condition is not satisfied, LASSO may end up choosing no instruments or too many instruments. As noted in Belloni et al. (2012), this assumption is likely to be violated in the case of many weak instruments. Unfortunately, we find that supplying LASSO with the full set of instruments in our data results in it choosing none. While there are a

[29]LASSO provides a formal way to make a bias-variance trade-off. Specifically, it minimizes the sum of squared errors subject to a constraint that the sum of the absolute value of the coefficients falls below a constant (the "tuning" parameter). If one relaxes this constraint completely (i.e. sets the tuning parameter to infinity), this reduces to OLS. The procedure ends up setting some of the coefficients exactly equal to zero, leaving the reduced set of instruments. LASSO is particularly appropriate in cases in which the (# of instruments) > N.

few instruments that exceed heuristic thresholds for relevance (e.g. "Max(MaxTemp)" gives a first-stage F-statistic of 15.8), they do not produce enough signal to be chosen by LASSO. Naively hand-picking the instrument with the largest F-statistic (i.e. using unprincipled data mining) risks choosing a variable that is most correlated with the noise in the first stage, thereby biasing the resulting IV estimate.

Given the lack of sparsity, our primary approach is to use the full set with estimators that are somewhat more reliable than 2SLS in the case of many weak instruments. In particular, we present results using limited information maximum likelihood (LIML) and the Angrist, Imbens, and Krueger (1999) jack knife instrumental variable estimator (JIVE).[30] While these estimators are often suggested as alternatives to 2SLS in the weak instruments case, Hahn & Hausman (2003) caution that "these estimators sometimes perform well and sometimes poorly in the WI situation" because they lack finite sample moments. Hansen, Hausman, and Newey (2008) note that LIML standard errors will be too small, but can be corrected using an extension of Bekker (1994). In simulations, Hansen, Hausman, and Newey (2008) find that LIML with these corrected standard errors performs well if the concentration parameter exceeds 32 (for a single endogenous variable, the concentration parameter is K*[F-1], where K is the number of instruments and F is the first-stage F-statistic for the instruments).

It's important to note that the preceding discussion assumes that our weather instruments are valid (i.e. that the residual variation in the weather variables, after controlling for demographics and week and year fixed effects, only affect stated return and recommendation likelihoods through stated trip enjoyment). As shown by Bound, Jaeger, and Baker (1995), even a weak correlation between the instruments and the error in the structural equation can lead to IV estimates being even more biased than OLS.

Appendix Table A.2 presents 2SLS, LIML, and JIVE estimates using the full set of instruments (Column 7) and six different subsets, with the three outcome variables across different panels. The table presents results from 63 IV estimations in total. All specifications include controls for demographics and week and year fixed effects. Column 2 shows that using two reasonable hand-picked instruments (the average maximum temperature and the average rainfall during the trip) results in extremely weak instruments, with a F-statistic of 1.060. The IV estimates are correspondingly imprecise. By contrast, using the arbitrary maximum temperature on the hottest day of the trip gives a much strong first-stage rela-

[30]With weak instruments, 2SLS is biased towards OLS and the degree of bias grows with the set of instruments. This bias is due to correlation between the fitted value from the first-stage regression for observation i with the error term for that same observation in the structural equation (e_i). The leave-one-out jack knife estimator attempts to break this dependence by using all observations except for i to estimate the coefficient in the first-stage and uses this along with z_i to construct the fitted value of the instrument for i. This process is repeated for all observations in the sample. We use the UJIVE1 estimator in Stata (Poi 2006).

tionship and much smaller standard errors. However, as discussed above, this estimate is unreliable. Using all six of the trip summary variables gives a first-stage F-statistic of 5.666. There we find similar estimates to column 1, and across the different estimation strategies. However, interacting these trip summary variables with month dummies cuts the resulting IV point estimates in half (Column 4). Finally, in Column 7 we examine including the full set of 90 instruments. The first-stage F-statistic is 2.324, implying a concentration parameter of 119, well beyond the simulation threshold shown in Hansen, Hausman, and Newey (2008) to produce reasonable estimates for LIML with corrected standard errors.[31] We see that LIML and JIVE give fairly similar results for the likelihood of recommending the destination. These estimates are larger than the IV estimates presented in Table 7 (0.654 and 0.762 vs. 0.319) with larger standard errors. The estimates remain significant at the 1% level. The LIML and JIVE point estimates for the likelihood of ever returning have the same sign (0.046 and 0.275 vs. 0.185), though are no longer significant at conventional levels. Finally, the coefficients are negative for the likelihood of returning in 12 months; with the estimate marginally significant for LIML.

To summarize, there are many objective measures of weather that could impact people's enjoyment of their trip (the first stage). Some of these measures produce a strong first stage (e.g. the maximum temperature during a trip has a strong effect on trip enjoyment), which then produces strong evidence of attribution bias in the second stage. Other intuitive choices such as average rain and temperature, do not produce a strong first stage, and therefore do not allow us to test for attribution bias (second stage). However, one cannot simply choose a variable like maximum temperature ex-post to use for the first stage. Approaches that use multiple instruments, but also guard against aspects of weak instrument bias (e.g. LIML and JIVE), provide evidence of attribution bias that is generally consistent with our analysis using subjective weather measures.

6 Conclusion

In this paper we aim to advance the study of attribution bias for economic decision making. To that end, we sketch a conceptual framework that closely follows the model of simple projection bias by Loewenstein, O'Donoghue, and Rabin (2003). We then review the psychology evidence on related biases, while highlighting the advantages of our approach for studying economic behavior. We find evidence of significant misattribution in a new con-

[31] We implement the Bekker standard errors for LIML using Stata code provided to us by Neil Davies, originally written Helmut Farbacher for Davies et al. (2015). As noted by Hansen, Hausman, and Newey (2008), the consistency of the LIML estimate relies on the assumption of homoskedasticity.

sumer experience that we created in a controlled experiment. Finally, we illustrate a test of attribution bias that leveraged naturally-occurring, exogenous variation in a state (weather) among prior customers to an amusement park. Together, we hope that the sections provide a foundation to encourage further theoretical refinement of the bias, as well as the collection of more empirical evidence in high-stakes decisions.

Though the tests of attribution bias presented in this paper all concern consumer decision making, we believe that attribution bias has relevance in a variety of economic domains. Within education, college students may perceive courses as more difficult or less interesting if they are taken during an early-morning section instead of later in the morning.[32] If a student takes a key course for their major field of study in one of these early slots, they may then misattribute the utility of taking further classes in the major due to the temporary variation in their tiredness at the time of the first course.[33] Likewise, students' misattributions may also influence professors' performance ratings.[34] In the labor market, an employer may find interviewing an applicant less enjoyable if they are tired during that interaction. That employer may later penalize the applicant if the employer misattributes her own temporary tiredness to a stable quality of the candidate.[35] Health investments may also be skewed by attribution bias. For example, an individual that tries a new exercise on a day they are not feeling well may consequently be less likely to re-engage in that activity. There are challenges associated with isolating attribution bias in each of these examples, but they highlight the broad relevance of the bias once we relax the definition of "consumption" in our model. In particular, each of these examples exploit an asymmetry in the costs of attribution bias. Cases where initial discouragement can lead to failure to ever resample the good or experience may have particularly large welfare effects. Moreover, if we find evidence of the bias in these decisions, there may be important policy implications. As one example, if college major choice is in fact skewed by the timing of required introductory courses, a university administrator may schedule these courses in a way that favors careers with the highest social returns. By enriching our understanding of state-dependent preferences and

[32] Pope (2016) and Shapiro and Williams (2015) find evidence of time-of-day effects in student productivity.

[33] Beyond perceived difficulty, students' lower grades in these courses could independently discourage them from pursuing a career in the morning-class subject for standard reasons (e.g. schools may screen on course-grouping GPA levels).

[34] Identifying this correlation as attribution bias assumes that professors' teaching abilities are independent of the course start time. If professors do a worse job teaching the early sections, then a lower rating would be a rational reflection of their lower productivity in these courses insofar as the rating is interpreted as a measure of the professors' realized performances.

[35] Both this example and the professor rating involve interpersonal misattributions; however, in both cases the process can be embedded within our model as long as we consider the interaction to be consumption. That is, the agent must be trying to form an assessment of the utility of interacting with the employee/professor in the future when forming that judgment.

intertemporal choice, we hope that the study of attribution bias will lead to further insights into behavior and associated policy improvements.

References

Acland, Dan, and Matthew R Levy. 2015. "Naivete, Projection Bias, and Habit Formation in Gym Attendance." *Management Science*, 61(1): 146–160.

Agarwal, Sumit, John C Driscoll, Xavier Gabaix, and David I Laibson. 2013. "Learning in the Credit Card Market." *Working Paper*.

Andrade, Eduardo B., and Dan Ariely. 2009. "The enduring impact of transient emotions on decision making." *Organizational Behavior and Human Decision Processes*, 109(1): 1–8.

Angrist, J D, G W Imbens, and A B Krueger. 1999. "Jackknife instrumental variables estimation." *Journal of Applied Econometrics*, 14: 57–67.

Ariely, Dan, and Michael I. Norton. 2008. "How actions create - not just reveal - preferences." *Trends in Cognitive Sciences*, 12(1): 13–16.

Bekker, Paul. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica*, 62(3): 657–681.

Belloni, A., D Chen, V C Chernozhukov, and C H Hansen. 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica*, 80(6): 2369–2429.

Belloni, A, V Chernozhukov, and C Hansen. 2010. "Lasso Methods for Gaussian Instrumental Variables Models." *arXiv Working Paper*, , (2010): 1–34.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Are CEOS Rewarded for Luck? The Ones Without Principals Are." *Quarterly Journal of Economics*, 116(August): 901–932.

Bound, John, David Jaeger, and Regina Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association*, 90(430): 443–450.

Brunstrom, Jeffrey M., Jeremy F. Burn, Nicola R. Sell, Jane M. Collingwood, Peter J. Rogers, Laura L. Wilkinson, Elanor C. Hinton, Olivia M. Maynard, and Danielle Ferriday. 2012. "Episodic Memory and Appetite Regulation in Humans." *PLoS ONE*, 7(12): e50707.

Buchheim, Lukas, and Thomas Kolaska. 2015. "Projection Bias with Salient State-Dependent Utility." *Working Paper*, 15.

Bushman, B. J., C. N. DeWall, R. S. Pond, and M. D. Hanus. 2014. "Low glucose relates to greater aggression in married couples." *Proceedings of the National Academy of Sciences*, 111(17): 6254–6257.

Bushong, Benjamin, and Tristan Gagnon-bartsch. In Progress. "Learning with Misattribution of Reference Dependence." 1–9.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2009. "Reinforcement Learning and Savings Behaviors." *Journal of Finance*, 64(6): 2515–2534.

Cohen, Joel B, Michel T Pham, and Eduardo B Andrade. 2008. "The Nature and Role of Affect in Consumer Behavior." *Handbook of Consumer Psychology*, , (352): 297–348.

Cole, Shawn, Andrew Healy, and Eric Werker. 2012. "Do voters demand responsive governments? Evidence from Indian disaster relief." *Journal of Development Economics*, 97(2): 167–181.

Cunningham, Tom. 2016. "Hierarchical Aggregation of Information and Decision-Making." *Working Paper*, 1–48.

Davies, Neil M, Von Hinke Kessler, Helmut Farbmacher, Stephen Burgess, and George Davey. 2015. "The many weak instruments problem and Mendelian randomization." *Statistics in Medicine*, 34(3): 454–468.

Dutton, Donald, and Arthur Aron. 1974. "Some evidence for heightened sexual attraction under conditions of high anxiety." *Journal of Personality and Social Psychology*, 30(4): 510–517.

Garbinsky, Emily N., Carey K. Morewedge, and Baba Shiv. 2014*a*. "Does liking or wanting determine repeat consumption delay?" *Appetite*, 72: 59–65.

Garbinsky, E. N., C. K. Morewedge, and B. Shiv. 2014*b*. "Interference of the End: Why Recency Bias in Memory Determines When a Food Is Consumed Again." *Psychological Science*, 25(7): 1466–1474.

Garcia, J ., D . J . Kimeldorf, and R . A . Koellino. 1955. "Conditioned Aversion to Saccharin Resulting from Exposure to Gamma Radiation." *Science*, 122(3160): 157–158.

Gasper, John T., and Andrew Reeves. 2011. "Make It Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters." *American Journal of Political Science*, 55(2): 340–355.

Gervais, Simon, Simon Gervais, Terrance Odean, and Terrance Odean. 2001. "Learning To Be Overconfident." *The Review of Financial Studies*, 14(1): 1–27.

Gilchrist, Duncan Sheppard, and Emily Glassberg Sands. 2015. "Something to Talk About: Social Spillovers in Movie Consumption." *Journal of Political Economy*, , (Forthcoming).

Gneezy, Uri, John List, and Michael K. Price. 2012. "Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments." *National Bureau of Economic Research Working Paper Series*, No. 17855.

Gorn, Gerald J., Marvin E. Goldberg, and Kunal Basu. 1993. "Mood, Awareness, and Product Evaluation." *Journal of Consumer Psychology*, 2(3): 237–256.

Griffin, Dale, and Amos Tversky. 1992. "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology*, 24(3): 411–435.

Hahn, Jinyong, and Jerry Hausman. 2003. "Weak Instruments: Diagnosis and Cures in Empirical Econometrics." *American Economic Review*, 93(2): 118–124.

Hansen, Christian, Jerry Hausman, and Whitney Newey. 2008. "Estimation With Many Instrumental Variables." *Journal of Business & Economic Statistics*, 26(4): 398–422.

Hayes, Rosa C., Masami Imai, and Cameron a. Shelton. 2015. "Attribution Error in Economic Voting: Evidence From Trade Shocks." *Economic Inquiry*, 53: 258–275.

Hirshleifer, David, and Tyler Shumway. 2003. "Good day sunshine: Stock returns and the weather." *Journal of Finance*, 58(3): 1009–1032.

Hitczenko, Marcin. 2013. "Modeling Anchoring Effects in Sequential Likert Scale Questions." *Federal Reserve Bank of Boston Working Paper*, , (13).

Ifcher, J, and H Zarghamee. 2011. "Happiness and time preference: The effect of positive affect in a random-assignment experiment." *American Economic Review*, forthcoming.

Isen, Alice M, and Nehemia Geva. 1987. "The influence of positive affect on acceptable level of risk: The person with a large canoe has a large worry." *Organizational Behavior and Human Decision Processes*, 39(2): 145–154.

Jones, Edward E., and Richard E. Nisbett. 1972. "The actor and the observer: Divergent perceptions of the causes of behavior."

Kellaris, James J., and Anthony D. Cox. 1989. "The Effects of Background Music in Advertising: A Reassessment." *Journal of Consumer Research*, 16(June): 113.

Keltner, Dacher, Kenneth Locke, and Paul Audrain. 1993. "The Influence of Attributions on the Relevance of Negative Feelings to Personal Satisfaction." *Personality and Social Psychology Bulletin*, 19(1): 21–29.

Kirchsteiger, Georg, Luca Rigotti, and Aldo Rustichini. 2006. "Your morals might be your moods." *Journal of Economic Behavior & Organization*, 59(2): 155–172.

Lerner, Jennifer S, Deborah a Small, and George Loewenstein. 2004. "Heart strings and purse strings: Carryover effects of emotions on economic decisions." *Psychological Science*, 15(5): 337–41.

Levy, Matthew. 2010. "An Empirical Analysis of Biases in Cigarette Addiction." *Working Paper.*

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics*, 118(November): 1209–1248.

Loewenstein, G F, E U Weber, C K Hsee, and N Welch. 2001. "Risk as feelings." *Psychological Bulletin*, 127(2): 267–286.

Lucas, Richard E, and Nicole M Lawless. 2013. "Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments." *Journal of Personality and Social Psychology*, 104(5): 872–84.

Malle, Bertram F. 2011. "Attribution Theories: How People Make Sense of Behavior." *Theories in Social Psychology.*, 72–95.

Mandel, Naomi, and EricJ. Johnson. 2002. "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices." *Journal of Consumer Research*, 29(2): 235–245.

Morewedge, Carey K. 2014. "Utility: Anticipated, Experienced, and Remembered." *Blackwell Handbook of Judgment and Decision-Making*, 1–83.

Murdock, Bennet B. 1962. "The serial position effect of free recall." *Journal of Experimental Psychology*, 64(5): 482–488.

Nisbett, Richard E., and Timothy D. Wilson. 1977. "The halo effect: Evidence for unconscious alteration of judgments." *Journal of Personality and Social Psychology*, 35(4): 250–256.

Payne, B. K., D. L. Hall, C. D. Cameron, and A. J. Bishara. 2010. "A Process Model of Affect Misattribution." *Personality and Social Psychology Bulletin*, 36(10): 1397–1408.

Podsakoff, Philip M, Scott B Mackenzie, and Nathan P Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology*, 63: 539–569.

Poi, Brian. 2006. "Jackknife instrumental variable estimation in Stata." *The Stata Journal*, 6(3): 364–376.

Pope, Nolan G. 2016. "How the Time of Day Affects Productivity: Evidence from School Schedules." *Review of Economics and Statistics*, 98(1): 1–11.

Ross, Lee. 1977. "The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process." *Advances in Experimental Social Psychology*, 10: 173–220.

Schachter, Stanley, and Jerome Singer. 1962. "Cognitive, Social, and Physiological Determinants of Emotional State." *Psychological Review*, 69(5): 379–399.

Schwartzstein, Joshua. 2014. "Selective attention and learning." *Journal of the European Economic Association*, , (April): 1–30.

Schwarz, Norbert. 2012. "Feelings-as-information theory." *Handbook of Theories of Social Psychology*, 1(January): 289–308.

Schwarz, Norbert, and Gerald L. Clore. 1983. "Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States." *Journal of Personality and Social Psychology*, 45(3): 513–523.

Shapiro, Teny Maghakian, and Kevin M. Williams. 2015. "Academic Achievement Across the Day: Evidence from Randomized Class Schedules." *Working Paper*.

Simonsohn, Uri. 2006. “Clouds Make Nerds Look Good: Field Evidence of the Impact of Incidental Factors on Decision Making.” *Journal of Behavioral Decision Making*.

Simonsohn, Uri. 2009. “Weather to go to College.” *Economic Journal*, 120: 270–280.

Strack, Fritz, Leonard L Martin, and Norbert Schwarz. 1988. “Priming and communication: Social determinants of information use in judgments of life satisfaction.” *European Journal of Social Psychology*, 18(5): 429–442.

Strahilevitz, Michal Ann, Terrance Odean, and Brad M Barber. 2011. “Once Burned, Twice Shy: How Naive Learning, Counterfactuals, and Regret Affect the Repurchase of Stocks Previously Sold.” *Journal of Marketing Research*, 48(SPL): S102–S120.

Weijters, Bert, Elke Cabooter, and Niels Schillewaert. 2010. “The effect of rating scale format on response styles: The number of response categories and response category labels.” *International Journal of Research in Marketing*, 27(3): 236–247.

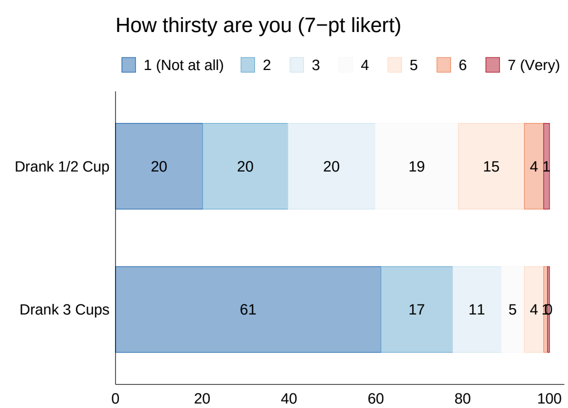
Weiner, Bernard. 2000. “Attributional Thoughts about Consumer Behavior.” *Journal of Consumer Research*, 27(3): 382–387.

Wolfers, Justin. 2007. “Are Voters Rational? Evidence from Gubernatorial Elections.” *Working Paper*, , (215).

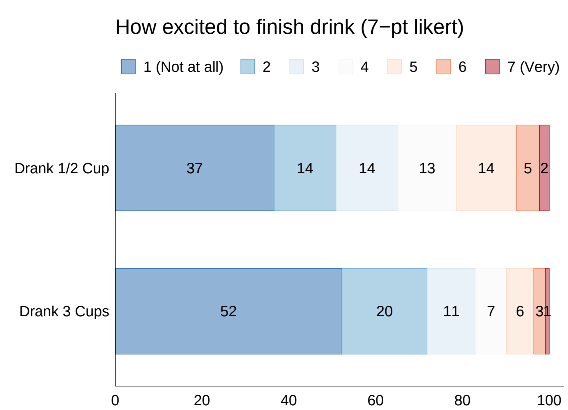
Figures

Figure 1: Baseline Likert Scales

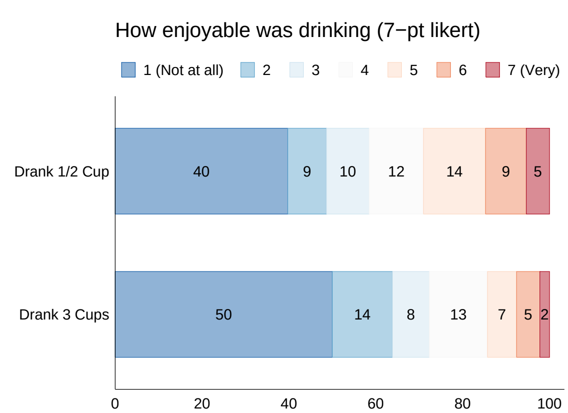
(a) Manipulation Check



(b) After Drinking Half of Mixed Drink

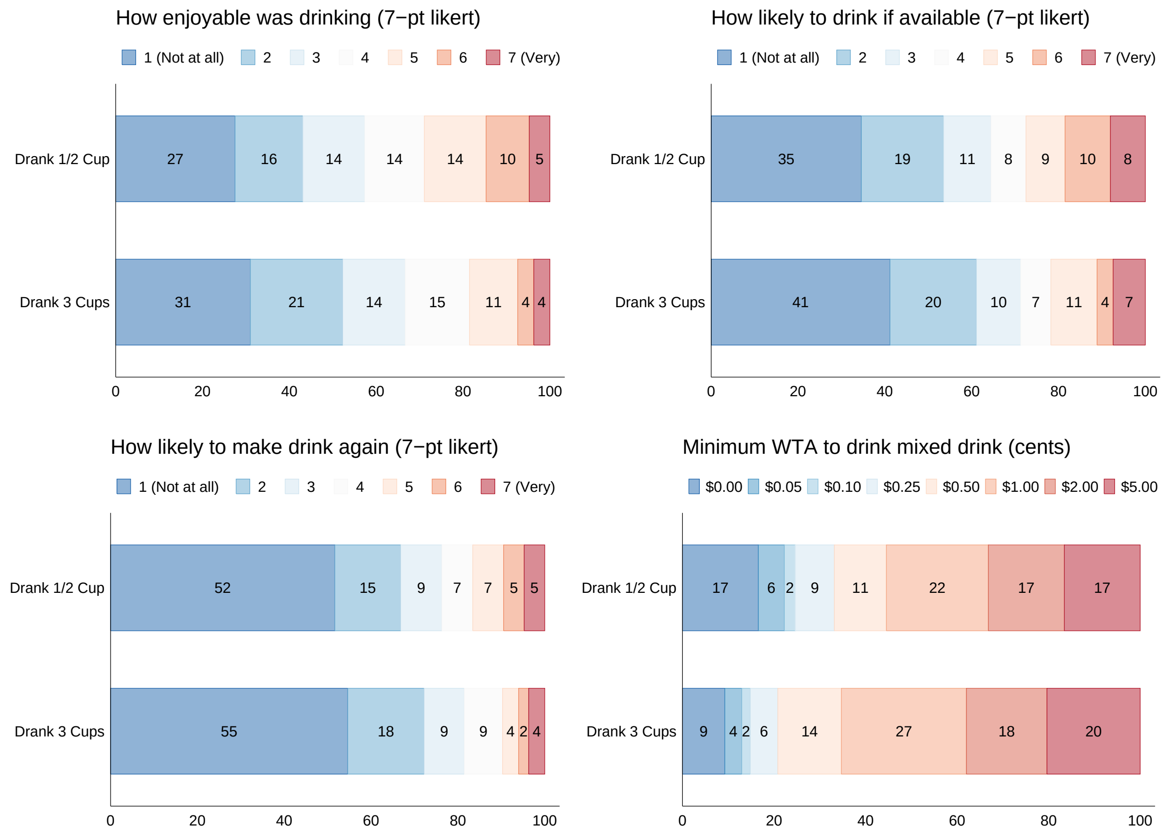


(c) After Drinking Full Mixed Drink



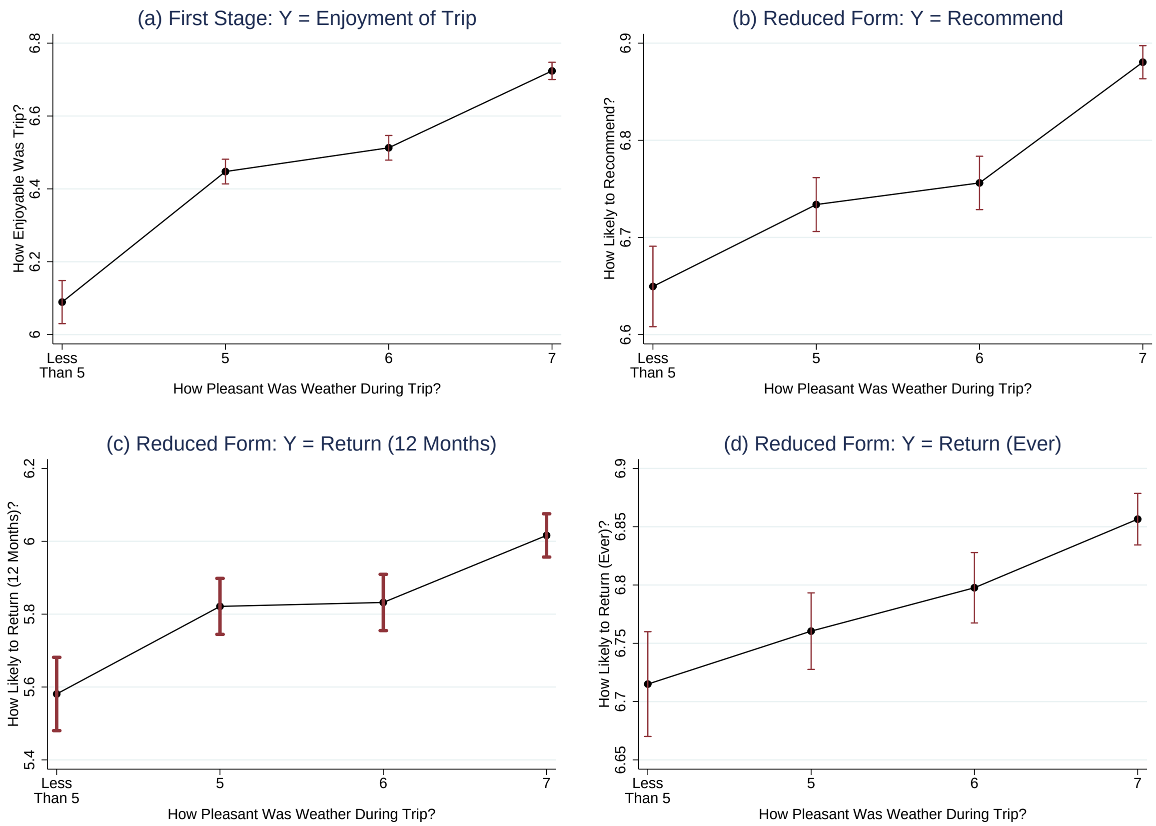
Notes: Full baseline sample (N=448). Horizontal axis is the percentage of responses corresponding to each category within a group (Drank 1/2 Cup or Drank 3 Cups). The survey questions were: (a) *"On a scale from 1 (not at all thirsty) to 7 (very thirsty), how thirsty are you right now?"* (b) *"On a scale from 1 (not at all excited) to 7 (very excited), how excited are you to drink the second half of the drink?"* (c) *"On a scale from 1 (not at all enjoyable) to 7 (very enjoyable), how enjoyable was drinking the mixed drink?"*. Question (a) was asked after drinking either $\frac{1}{2}$ cup of water or 3 cups of water, but before drinking any of the mixed drink. Question (b) was asked after drinking half of the mixed drink. Question (c) was asked after fully drinking the mixed drink.

Figure 2: Follow-Up Likert Scales



Notes: Full follow-up sample (N=427). Horizontal axis is the percentage of responses corresponding to each category within a group (Drank 1/2 Cup or Drank 3 Cups). The survey questions were: (a) *"On a scale from 1 (not at all enjoyable) to 7 (very enjoyable), how enjoyable was drinking the mixed drink we asked you to consume during our last survey?"* (b) *"If you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now, how likely would you be to drink it?"* (c) *"How likely are you to mix and consume the same mixed drink we asked you to prepare in the last survey?"* (d) *"Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so."* Question order was randomized.

Figure 3: Amusement Park Study: First-Stage and Reduced-Form



Notes: The Y-axes change across panels and do not start at zero. 95% confident intervals are reported. All survey questions are on 7-point likert scales (1=Not at all, 7 = Very). "How Pleasant Was Weather During Trip?" corresponds to the survey question *"How pleasant was the weather while visiting theme parks during your last trip to Orlando?"*; we collapse the bottom 4 categories into "Less Than 5" for this figure (the sample sizes for "1", "2", "3", and "4" are 76, 122, 309, and 760 respectively). "Recommend" corresponds to *"How likely are you to recommend an Orlando theme park vacation to your friends and family?"*. "Return (12 Months)" corresponds to *"How likely are you to do another Orlando theme park vacation in the next 12 months?"*. "Return (Ever)" corresponds to *"How likely are you to do another Orlando theme park vacation ever again?"*. "Enjoyment of Trip" corresponds to *"How enjoyable was your theme park experience during your last trip to Orlando?"*.

Tables

Table 1: Summary Statistics and Balance Tests

	(1) Drank 1/2 Cup	(2) Drank 3 Cups	(3) Total	(4) Diff. [(1)-(2)]	(5) P-Val
Baseline Survey					
Female	0.60	0.57	0.59	0.03	0.50
	(0.49)	(0.50)	(0.49)		
Age	33.61	34.57	34.09	-0.96	0.29
	(9.88)	(9.48)	(9.68)		
Weight (pounds)	178.42	174.75	176.58	3.67	0.43
	(51.36)	(47.42)	(49.41)		
Height (inches)	67.44	66.81	67.12	0.64	0.26
	(3.96)	(7.40)	(5.94)		
Usual daily water consumption (glasses)	5.08	5.29	5.19	-0.21	0.42
	(2.85)	(2.65)	(2.75)		
Usual daily liquid consumption (glasses)	6.72	6.81	6.76	-0.09	0.76
	(2.91)	(3.14)	(3.02)		
Time since last drank (categorical)	3.26	3.15	3.20	0.11	0.55
	(1.78)	(2.12)	(1.95)		
Baseline survey length (minutes)	26.48	27.82	27.15	-1.34	0.37
	(15.61)	(16.13)	(15.87)		
Took follow-up survey	0.94	0.96	0.95	-0.02	0.26
	(0.23)	(0.19)	(0.21)		
Observations	224	224	448		
Follow-Up Survey					
Time between surveys (hours)	53.83	53.46	53.64	0.37	0.83
	(17.52)	(18.43)	(17.97)		
Follow-up survey length (minutes)	1.80	2.08	1.94	-0.28	0.16
	(1.79)	(2.33)	(2.08)		
Time between email and follow-up start (hours)	4.04	3.68	3.86	0.37	0.58
	(7.20)	(6.50)	(6.85)		
How thirsty are you (7-pt likert)	3.86	3.82	3.84	0.03	0.84
	(1.73)	(1.66)	(1.69)		
Observations	211	216	427		

Notes: Standard deviations in parentheses. “Drank 3 Cups” corresponds to subjects that were randomly assigned to drink 3 cups of water before answering questions and consuming the mixed drink, while the “Drank $\frac{1}{2}$ Cup” group was instead assigned to drink $\frac{1}{2}$ cup of water. “Time since last drank (categorical)” is the answer to the question, “When is the last time you drank anything prior to starting this survey?” (with categories 1=“0-30 min ago”, 2=“30-60 min ago”, 3=“1-2 hours ago”, 4=“2-3 hours ago”,...,13=“11-12 hours ago”, 14=“More than 12 hours ago”). “Time between email and follow-up start” measures the number of hours elapsed between the follow-up survey’s announcement and the subject’s survey start time. “Baseline survey length” truncates 52 surveys to 60 minutes.

Table 2: Outcome Variable Summary Statistics and T-Tests

	(1) Drank 1/2 Cup	(2) Drank 3 Cups	(3) Total	(4) Diff. [(1)-(2)]	(5) P-Val
Baseline Survey					
How thirsty are you (7-pt likert)	3.08	1.80	2.44	1.29***	0.00
	(1.56)	(1.25)	(1.55)		
How excited to finish drink (7-pt likert)	2.79	2.07	2.43	0.71***	0.00
	(1.76)	(1.46)	(1.66)		
How enjoyable was drinking (7-pt likert)	3.02	2.38	2.70	0.64***	0.00
	(2.02)	(1.73)	(1.91)		
Observations	224	224	448		
Follow-Up Survey					
How enjoyable was drinking (7-pt likert)	3.20	2.80	3.00	0.41**	0.02
	(1.90)	(1.72)	(1.82)		
How likely to drink if available (7-pt likert)	3.01	2.67	2.84	0.35*	0.07
	(2.06)	(1.94)	(2.01)		
How likely to make drink again (7-pt likert)	2.36	2.11	2.23	0.25	0.14
	(1.83)	(1.62)	(1.73)		
Minimum WTA to drink mixed drink (cents)	146.73	173.17	160.11	-26.44	0.12
	(171.12)	(176.74)	(174.29)		
Willing to drink for $0.00	0.17	0.09	0.13	0.07**	0.02
	(0.37)	(0.29)	(0.34)		
Willing to drink for $0.10	0.23	0.14	0.19	0.09**	0.02
	(0.42)	(0.35)	(0.39)		
Willing to drink for $0.25	0.33	0.20	0.26	0.12***	0.00
	(0.47)	(0.40)	(0.44)		
Willing to drink for $0.50	0.44	0.33	0.38	0.10**	0.03
	(0.50)	(0.47)	(0.49)		
Willing to drink for $1.00	0.66	0.61	0.63	0.05	0.31
	(0.48)	(0.49)	(0.48)		
Willing to drink for $2.00	0.82	0.77	0.80	0.05	0.19
	(0.38)	(0.42)	(0.40)		
Willing to drink for $5.00	0.96	0.94	0.95	0.01	0.54
	(0.20)	(0.23)	(0.22)		
Observations	211	216	427		

Notes: Standard deviations in parentheses. All variables from "Minimum WTA to drink mixed drink (cents)" to the bottom of the table are constructed from the survey question: "Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so.". Subjects were required to select "Yes I Would Drink It" or "No I Would NOT Drink It" for each of a series of different amounts ($0.00, $0.05, $0.10, $0.25, $0.50, $1.00, $2.00, $5.00).

Table 3: Thirst Experiment Outcomes: Reduced-Form and IV

	Would Drink		Would Make		Minimum WTA	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Reduced-Form (OLS of Treatment Indicator)						
	b/se	b/se	b/se	b/se	b/se	b/se
3 Cup Treatment	-0.35*	-0.38*	-0.25	-0.25	26.44	26.92
	(0.19)	(0.20)	(0.17)	(0.17)	(16.83)	(17.15)
N	427	427	427	427	427	427
R^2	0.01	0.05	0.01	0.05	0.01	0.02
DepVarMean	3.01	3.01	2.36	2.36	146.73	146.73
Controls		X		X		X
Panel B: IV Using Treatment as Instrument						
Baseline Enjoyment	0.55**	0.57***	0.40*	0.38*	-41.97*	-40.32*
	(0.22)	(0.21)	(0.22)	(0.21)	(23.99)	(22.65)
First-Stage R^2	0.03	0.08	0.03	0.08	0.03	0.08
First-Stage F-Stat	11.93	13.40	11.93	13.40	11.93	13.40
DepVarMean	3.01	3.01	2.36	2.36	146.73	146.73
Controls		X		X		X
Panel C: IV Using Baseline Thirst as Instrument						
Baseline Enjoyment	0.73***	0.68***	0.77***	0.69***	-31.16**	-21.14
	(0.14)	(0.14)	(0.16)	(0.16)	(14.97)	(15.84)
First-Stage R^2	0.06	0.10	0.06	0.10	0.06	0.10
First-Stage F-Stat	26.69	25.44	26.69	25.44	26.69	25.44
DepVarMean	3.01	3.01	2.36	2.36	146.73	146.73

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors are presented in parentheses. Even columns include the full set of variables listed in Table 1 as controls. "How Enjoyable" corresponds to the survey question *"On a scale from 1 (not at all enjoyable) to 7 (very enjoyable), how enjoyable was drinking the mixed drink we asked you to consume during our last survey?"*. "Would Drink" corresponds to *"If you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now, how likely would you be to drink it?"*. "Would Make" corresponds to *"How likely are you to mix and consume the same mixed drink we asked you to prepare in the last survey?"*. "Minimum WTA" is the lowest amount (from a list of amounts) for which the subject answered "yes" to *"Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so."*. *DepVarMean* is the mean of the dependent variable in the "Drank $\frac{1}{2}$ Cup" sample.

Table 4: Comparing IV and OLS Results

	Would Drink		Would Make		Minimum WTA	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: IV Using Treatment as Instrument						
Baseline Enjoyment	0.55**	0.57***	0.40*	0.38*	-41.97*	-40.32*
	(0.22)	(0.21)	(0.22)	(0.21)	(23.99)	(22.65)
First-Stage R^2	0.03	0.08	0.03	0.08	0.03	0.08
First-Stage F-Stat	11.93	13.40	11.93	13.40	11.93	13.40
DepVarMean	3.01	3.01	2.36	2.36	146.73	146.73
Panel B: OLS (Full Sample)						
Baseline Enjoyment	0.76***	0.76***	0.50***	0.49***	-40.38***	-41.23***
	(0.04)	(0.04)	(0.04)	(0.04)	(3.56)	(3.58)
N	427	427	427	427	427	427
R^2	0.53	0.54	0.30	0.33	0.19	0.21
Panel C: OLS (Control Group Only)						
Baseline Enjoyment	0.80***	0.81***	0.57***	0.57***	-39.78***	-41.54***
	(0.04)	(0.04)	(0.05)	(0.05)	(4.55)	(4.61)
N	211	211	211	211	211	211
R^2	0.61	0.64	0.39	0.44	0.22	0.29
Panel D: OLS (Control Group, With Baseline Thirst Control)						
Baseline Enjoyment	0.80***	0.80***	0.54***	0.55***	-39.32***	-41.53***
	(0.04)	(0.04)	(0.05)	(0.05)	(4.56)	(4.58)
Baseline Thirst	0.02	0.01	0.16**	0.13*	-3.52	-0.11
	(0.06)	(0.06)	(0.06)	(0.07)	(5.97)	(6.71)
N	211	211	211	211	211	211
R^2	0.62	0.64	0.41	0.45	0.22	0.29

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors are presented in parentheses. Even columns include the full set of variables listed in Table 1 as controls. "How Enjoyable" corresponds to the survey question *"On a scale from 1 (not at all enjoyable) to 7 (very enjoyable), how enjoyable was drinking the mixed drink we asked you to consume during our last survey?"*. "Would Drink" corresponds to *"If you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now, how likely would you be to drink it?"*. "Would Make" corresponds to *"How likely are you to mix and consume the same mixed drink we asked you to prepare in the last survey?"*. "Minimum WTA" is the lowest amount (from a list of amounts) for which the subject answered "yes" to *"Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so."*. *DepVarMean* is the mean of the dependent variable in the "Drank $\frac{1}{2}$ Cup" sample.

Table 5: Projection Bias vs. Attribution Bias

	Would Drink		Would Make		Min WTA	
	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Baseline Thirst	0.19***	0.17*	0.21***	0.25***	-7.72	-10.04
	(0.07)	(0.10)	(0.06)	(0.09)	(5.35)	(7.77)
Follow-Up Thirst	0.12**	0.06	0.08	0.07	-6.33	-4.74
	(0.06)	(0.09)	(0.05)	(0.08)	(5.29)	(7.67)
N	427	211	427	211	427	211
R^2	0.04	0.02	0.05	0.06	0.01	0.01
DepVarMean	3.01	3.01	2.36	2.36	146.73	146.73
Sample	Full	Control	Full	Control	Full	Control

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors are presented in parentheses. "How Enjoyable" corresponds to the survey question *"On a scale from 1 (not at all enjoyable) to 7 (very enjoyable), how enjoyable was drinking the mixed drink we asked you to consume during our last survey?"*. "Would Drink" corresponds to *"If you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now, how likely would you be to drink it?"*. "Would Make" corresponds to *"How likely are you to mix and consume the same mixed drink we asked you to prepare in the last survey?"*. "Minimum WTA" is the lowest amount (from a list of amounts) for which the subject answered "yes" to *"Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so."*. *DepVarMean* is the mean of the dependent variable in the "Drank $\frac{1}{2}$ Cup" sample.

Table 6: Summary Statistics for Amusement Park Survey

	(1) N	(2) Mean	(3) Standard Deviation
Survey Variables			
Number of trips to Orlando	9,281	8.45	4.85
Trip length (days)	9,285	4.94	3.61
Recalled a date for last trip	9,340	0.88	0.32
Number of days since last trip	8,224	349	982
Number of parks visited during last trip	9,340	3.20	1.84
How pleasant was weather? (1 to 7)	9,332	5.70	1.34
How much rain was there? (0 to 3)	9,332	0.92	0.81
How enjoyable was trip? (1 to 7)	9,331	6.51	0.88
How likely to recommend? (1 to 7)	9,334	6.78	0.63
How likely to return (12 months)? (1 to 7)	9,334	5.86	1.81
How likely to return (ever)? (1 to 7)	9,334	6.80	0.73
Demographic Variables			
Age	8,409	43	12
Female	9,323	0.58	0.49
Live outside Florida	9,323	0.70	0.46
Household size	8,409	3.09	1.32
Finished college	9,323	0.52	0.50
Income over 100k	9,323	0.34	0.47
Annual pass holder	9,340	0.22	0.42

Notes: Demographics were provided by the Amusement Park from earlier records. "Number of trips to Orlando" is substantially truncated, with 3,333 responses in the "14+" category (coded as 14). "Trip length (days)" is also truncated, with 575 responses in the "14+" category. "Number of days since last trip" is coded by subtracting "Trip length (days)" from the date of the last day of trip, among the 88% of respondents that were able to recall that exact date. "Live outside Florida" includes 1,992 respondents that live outside the United States of America.

Table 7: Reduced-Form and IV (Instrument = How Pleasant Was Weather, 1 to 7)

	Would Recommend			Will Return (12 Months)			Will Return (Ever)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Reduced Form (X = How Pleasant Was Weather)									
Pleasant Weather	0.061***	0.061***	0.051***	0.120***	0.106***	0.089***	0.039***	0.035***	0.039***
	(0.006)	(0.007)	(0.008)	(0.017)	(0.017)	(0.022)	(0.007)	(0.007)	(0.009)
N	8,223	7,463	3,743	8,223	7,463	3,743	8,223	7,463	3,743
R^2	0.017	0.039	0.049	0.008	0.080	0.120	0.006	0.014	0.045
DepVarMean	6.79	6.79	6.80	5.89	5.89	5.89	6.82	6.82	6.83
Demographics & Time FE		X	X		X	X		X	X
Info Treatment			X			X			X
Panel B: IV (Instrument = How Pleasant Was Weather)									
Enjoyment on Trip	0.334***	0.321***	0.252***	0.661***	0.437***	0.441***	0.215***	0.186***	0.195***
	(0.030)	(0.031)	(0.036)	(0.097)	(0.090)	(0.113)	(0.035)	(0.036)	(0.043)
N	8,223	7,463	3,743	8,223	7,463	3,743	8,223	7,463	3,743
First-Stage F-Stat	399.4	372.4	197.7	399.4	372.4	197.7	399.4	372.4	197.7
DepVarMean	6.79	6.79	6.80	5.89	5.89	5.89	6.82	6.82	6.83
Demographics & Time FE		X	X		X	X		X	X
Info Treatment			X			X			X
Panel C: OLS (X = How Enjoyable Was Trip)									
Enjoyment on Trip	0.329***	0.330***	0.317***	0.230***	0.215***	0.226***	0.194***	0.188***	0.194***
	(0.015)	(0.016)	(0.022)	(0.031)	(0.028)	(0.034)	(0.018)	(0.016)	(0.023)
Pleasant Weather	0.001	-0.002	-0.013*	0.078***	0.069***	0.043*	0.004	0.002	0.000
	(0.006)	(0.006)	(0.007)	(0.018)	(0.017)	(0.023)	(0.006)	(0.006)	(0.008)
N	8,223	7,463	3,743	8,223	7,463	3,743	8,223	7,463	3,743
R^2	0.222	0.241	0.242	0.020	0.090	0.131	0.061	0.067	0.104
DepVarMean	6.79	6.79	6.80	5.86	5.89	5.89	6.82	6.82	6.83
Demographics & Time FE		X	X		X	X		X	X
Info Treatment			X			X			X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered at the date level (self-reported last day of most recent trip). All survey questions are on 7-point likert scales (1=Not at all, 7 = Very). "Pleasant Weather" corresponds to the survey question *"How pleasant was the weather while visiting theme parks during your last trip to Orlando?"*. "Would Recommend" corresponds to *"How likely are you to recommend an Orlando theme park vacation to your friends and family?"*. "Will Return (12 Months)" corresponds to *"How likely are you to do another Orlando theme park vacation in the next 12 months?"*. "Will Return (Ever)" corresponds to *"How likely are you to do another Orlando theme park vacation ever again?"*. "Enjoyment on Trip" corresponds to *"How enjoyable was your theme park experience during your last trip to Orlando?"*. *DepVarMean* is the mean of the dependent variable in the full sample. "Demographics" are age, female, household size, whether respondent finished college, whether the respondent earns over $100,000/year, and whether the respondent has an annual pass; "Time FE" are week of the year and year fixed effects (for self-reported last day of most recent trip).

Table 8: IV Heterogeneity: Display Order, Location, and Experience

	Would Recommend			Will Return (12 Months)			Will Return (Ever)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: IV Split by Display Order (Trip Evaluation or Weather Evaluation First)									
Enjoyment on Trip	0.283***	0.376***		0.316***	0.585***		0.110**	0.288***	
	(0.038)	(0.053)		(0.107)	(0.142)		(0.044)	(0.061)	
N	3,705	3,758		3,705	3,758		3,705	3,758	
First-Stage F-Stat	239.0	134.4		239.0	134.4		239.0	134.4	
DepVarMean	6.79	6.79		5.87	5.91		6.81	6.83	
Trip Eval Asked First?	No	Yes		No	Yes		No	Yes	
Demographics & Time FE	X	X		X	X		X	X	
Panel B: IV Split by Location of Residence (Florida vs. Outside Florida)									
Enjoyment on Trip	0.327***	0.293***		0.530***	0.178*		0.165***	0.174***	
	(0.040)	(0.053)		(0.115)	(0.095)		(0.043)	(0.064)	
N	5,175	2,249		5,175	2,249		5,175	2,249	
First-Stage F-Stat	226.7	97.8		226.7	97.8		226.7	97.8	
DepVarMean	6.81	6.76		5.59	6.58		6.81	6.83	
Florida Resident?	No	Yes		No	Yes		No	Yes	
Demographics & Time FE	X	X		X	X		X	X	
Panel C: IV Split by Number of Past Trips to Orlando									
Enjoyment on Trip	0.759***	0.350***	0.269***	1.905**	0.526***	0.342***	0.970***	0.211***	0.120***
	(0.269)	(0.053)	(0.037)	(0.795)	(0.174)	(0.086)	(0.351)	(0.079)	(0.035)
N	415	2,280	4,729	415	2,280	4,729	415	2,280	4,729
First-Stage F-Stat	8.4	96.9	247.6	8.4	96.9	247.6	8.4	96.9	247.6
DepVarMean	6.68	6.74	6.83	4.96	5.26	6.26	6.50	6.73	6.89
Number of Past Trips	1	2 to 5	6+	1	2 to 5	6+	1	2 to 5	6+
Demographics & Time FE	X	X	X	X	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered at the date level (self-reported last day of most recent trip). All survey questions are on 7-point likert scales (1=Not at all, 7 = Very). "Would Recommend" corresponds to *"How likely are you to recommend an Orlando theme park vacation to your friends and family?"*. "Will Return (12 Months)" corresponds to *"How likely are you to do another Orlando theme park vacation in the next 12 months?"*. "Will Return (Ever)" corresponds to *"How likely are you to do another Orlando theme park vacation ever again?"*. "Enjoyment on Trip" corresponds to *"How enjoyable was your theme park experience during your last trip to Orlando?"*. *DepVarMean* is the mean of the dependent variable in the full sample. "Demographics" are age, female, household size, whether respondent finished college, whether the respondent earns over $100,000/year, and whether the respondent has an annual pass; "Time FE" are week-of-the-year and year fixed effects (for self-reported last day of most recent trip). Panel B additionally controls for the number of past trips to Orlando and Panel C additionally controls for whether the respondent is a Florida resident.

A Appendix: Additional Tables and Figures

Table A.1: Interval Regression and Ordered Probit for Minimum WTA

	OLS		Interval Regression		Ordered Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
3 Cup Treatment	26.44	26.92	36.47*	38.16*	0.23**	0.25**
	(16.83)	(17.15)	(21.67)	(21.73)	(0.10)	(0.10)
N	427	427	427	427	427	427
DepVarMean	146.73	146.73	146.73	146.73	146.73	146.73
Controls		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Even columns include the full set of variables listed in Table 1 as controls. "Minimum WTA" is the lowest amount (from a list of amounts) for which the subject answered "yes" to *"Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so."*. For the interval regression, we code the left of the interval as the lowest amount for which the subject answers "Yes", and the right of the interval as the next highest option (i.e. the lowest amount for which the subject answer "No"). *DepVarMean* is the mean of the dependent variable in the "Drank $\frac{1}{2}$ Cup" sample.

Table A.2: IV Using Merged Weather Variables (2SLS, LIML, & JIVE)

		Objective Weather Data: Alternative Instrumental Variable Sets						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Y: Would Recommend								
2SLS	$\hat{\beta}$	0.828***	1.168	0.734***	0.497***	0.258	0.577***	0.459***
	SE	(0.182)	(0.752)	(0.131)	(0.062)	(0.175)	(0.098)	(0.056)
LIML	$\hat{\beta}$	0.828***	1.845	0.848***	0.704***	0.052	0.764***	0.654***
	SE	(0.157)	(1.599)	(0.157)	(0.141)	(0.637)	(0.165)	(0.144)
JIVE	$\hat{\beta}$	0.877***	-1.346	0.815***	0.796***	0.582	0.777***	0.762***
	SE	(0.207)	(2.752)	(0.167)	(0.203)	(0.680)	(0.200)	(0.215)
Y: Will Return (12 Months)								
2SLS	$\hat{\beta}$	-0.426	-0.524	-0.456	-0.198	0.236	-0.306	-0.158
	SE	(0.458)	(1.570)	(0.366)	(0.202)	(0.524)	(0.296)	(0.183)
LIML	$\hat{\beta}$	-0.426	-0.524	-0.638	-2.311*	0.085	-0.695	-2.129*
	SE	(0.410)	(1.450)	(0.423)	(1.211)	(6.005)	(0.479)	(1.129)
JIVE	$\hat{\beta}$	-0.494	2.097	-0.631	-0.939	0.440	-0.803	-1.032
	SE	(0.481)	(4.203)	(0.436)	(0.622)	(2.118)	(0.574)	(0.685)
Y: Will Return (Ever)								
2SLS	$\hat{\beta}$	0.457**	0.286	0.471***	0.196***	0.007	0.357***	0.188***
	SE	(0.193)	(0.640)	(0.140)	(0.077)	(0.229)	(0.113)	(0.070)
LIML	$\hat{\beta}$	0.457***	0.289	0.495***	0.305	-0.615	0.464***	0.046
	SE	(0.160)	(0.545)	(0.142)	(1.867)	(0.973)	(0.176)	(2.709)
JIVE	$\hat{\beta}$	0.487**	-0.064	0.538***	0.278	0.819	0.507**	0.275
	SE	(0.217)	(1.398)	(0.181)	(0.204)	(0.983)	(0.219)	(0.221)
N		7,376	7,376	7,376	7,376	7,404	7,376	7,376
DepVarMean		6.818	6.818	6.818	6.818	6.819	6.818	6.818
1^{st}−Stage F-Stat		21.733	1.060	5.666	2.303	1.463	2.520	2.324
# of Instruments		1	2	6	72	18	24	90
Instruments:		Max: MaxTemp	Mean: MaxTemp Rainfall	All Trip Summaries	All Trip Summaries X Month	Last Day: Rainfall MaxTemp	(3) & (5)	(4) & (5)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered at the date level (self-reported last day of most recent trip) for 2SLS, Bekker standard errors for LIML, and robust standard errors for JIVE. All survey questions are on 7-point likert scales (1=Not at all, 7 = Very). Weather variables are taken from the Orlando International Airport weather station (as reported by Weather Underground). Trip summaries (mean, max, min) are taken over the last day of the most recent trip and the preceding X days of the trip (where X is the response to "how many days did you spend visiting theme parks"). Rainfall is in inches and temperature is in Fahrenheit. All variables are residuals from first regressing on demographics and week and year fixed effects. Instrument sets are:
(Column 1) [Trip Summary] Maximum temperature on the hottest day of the trip (i.e. Max(MaxTemp)).
(Column 2) [Trip Summary] Means of the maximum temperature and rainfalls on each day of the trip (i.e. Mean(MaxTemp), Mean(Rainfall)).
(Column 3) [Trip Summary] Maximum temperature on the hottest day of the trip, minimum temperature on the coldest day of the trip, rainfall on the wettest day of the trip, means for maximum temperature and rainfall, and the fraction of days on which it rained at least a quarter of an inch (Max(MaxTemp), Min(MinTemp), Max(Rainfall), Mean(Maxtemp), Mean(Rainfall),Frac(1[Rainfall> 0.25])).
(Column 4) [Trip Summary] Each variable in (3) interacted with dummies for the week of the year.
(Column 5) [Last Day of the Trip] For just the last day of the trip, we include dummy variables for intervals of the maximum temperature (Under 55, 55 to 59, ..., 90 to 94, 95 and above) and for amounts of rainfall (Under 0.25inches, 0.25 to 0.49,..., 2.00 to 2.24, 2.25 and above).
(Column 6) The union of (3) and (5).
(Column 7) The union of (4) and (5).

Table A.3: Reduced-Form, IV, and First-Stage (Instrument = How Much Rain)

	Would Recommend			Will Return (12 Months)			Will Return (Ever)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Reduced Form (X = How Much Rain (Omitted: None))									
Rain: Small	0.015	0.031*	0.052**	-0.258***	-0.167***	-0.174***	0.001	0.017	0.026
	(0.014)	(0.016)	(0.022)	(0.051)	(0.046)	(0.066)	(0.017)	(0.017)	(0.025)
Rain: Medium	0.001	0.019	0.062**	-0.447***	-0.327***	-0.273***	-0.033	-0.019	-0.026
	(0.021)	(0.024)	(0.029)	(0.077)	(0.070)	(0.100)	(0.027)	(0.027)	(0.045)
Rain: Large	-0.089**	-0.084*	-0.008	-0.383***	-0.260**	-0.147	-0.075*	-0.078*	0.014
	(0.041)	(0.046)	(0.051)	(0.116)	(0.112)	(0.146)	(0.043)	(0.047)	(0.056)
N	8,223	7,463	3,743	8,223	7,463	3,743	8,223	7,463	3,743
R^2	0.001	0.025	0.040	0.008	0.078	0.118	0.001	0.010	0.040
DepVarMean	6.79	6.79	6.80	5.89	5.89	5.89	6.82	6.82	6.83
Demographics & Time FE		X	X		X	X		X	X
Info Treatment			X			X			X
Panel B: IV (Instrument = How Much Rain (Omitted: None))									
Enjoyment on Trip	0.257***	0.305***	0.210*	0.384	0.013	-0.321	0.199**	0.229**	0.016
	(0.087)	(0.095)	(0.122)	(0.275)	(0.283)	(0.373)	(0.100)	(0.110)	(0.128)
N	8,223	7,463	3,743	8,223	7,463	3,743	8,223	7,463	3,743
First Stage F-Stat	14.5	12.9	7.6	14.5	12.9	7.6	14.5	12.9	7.6
DepVarMean	6.79	6.79	6.80	5.89	5.89	5.89	6.82	6.82	6.83
Demographics & Time FE		X	X		X	X		X	X
Info Treatment			X			X			X
Panel C: First-Stage (X = How Much Rain (Omitted: None))									
Rain: Small	0.048**	0.053**	0.050	0.048**	0.058***	0.050	0.048**	0.058***	0.050
	(0.022)	(0.024)	(0.035)	(0.022)	(0.022)	(0.035)	(0.022)	(0.022)	(0.035)
Rain: Medium	-0.025	-0.010	0.039	-0.025	-0.003	0.039	-0.025	-0.003	0.039
	(0.032)	(0.035)	(0.046)	(0.032)	(0.031)	(0.046)	(0.032)	(0.031)	(0.046)
Rain: Large	-0.353***	-0.334***	-0.335***	-0.353***	-0.317***	-0.335***	-0.353***	-0.317***	-0.335***
	(0.065)	(0.067)	(0.087)	(0.065)	(0.065)	(0.087)	(0.065)	(0.065)	(0.087)
N	8,223	7,463	3,743	8,223	7,463	3,743	8,223	7,463	3,743
R^2	0.009	0.027	0.036	0.009	0.010	0.036	0.009	0.010	0.036
DepVarMean	6.51	6.51	6.51	6.51	6.51	6.51	6.51	6.51	6.51
Demographics & Time FE		X	X		X	X		X	X
Info Treatment			X			X			X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered at the date level (self-reported last day of most recent trip). All survey questions are on 7-point likert scales (1=Not at all, 7 = Very). Dummies for "Rain: Small", "Rain: Medium", and "Rain: Large" correspond to categorical responses ("A small amount", "A medium amount", and "A large amount") to the survey question *"How much did it rain while visiting theme parks during your last trip to Orlando?"* ("Not at all" is the omitted category). "Would Recommend" corresponds to *"How likely are you to recommend an Orlando theme park vacation to your friends and family?"*. "Will Return (12 Months)" corresponds to *"How likely are you to do another Orlando theme park vacation in the next 12 months?"*. "Will Return (Ever)" corresponds to *"How likely are you to do another Orlando theme park vacation ever again?"*. "Enjoyment on Trip" corresponds to *"How enjoyable was your theme park experience during your last trip to Orlando?"*. *DepVarMean* is the mean of the dependent variable in the full sample. "Demographics" are age, female, household size, whether respondent finished college, whether the respondent earns over $100,000/year, and whether the respondent has an annual pass; "Time FE" are week-of-the-year and year fixed effects (for self-reported last day of most recent trip).

B Appendix: Experiment Materials

Figure B.1: Baseline Survey: Image Upload Page

CHICAGO BOOTH

We now request that you take and upload a photo of all of the ingredients:

1. Measure 3 cups of water and place in one (or more) glass.
2. Measure 1/2 cup of water and place in a separate glass.
3. Measure 1 cup of milk and place it in a third glass. You will be asked to combine the remaining 2 ingredients into this cup later in the survey, but keep them separate for now.
4. Measure 1 tablespoon of sugar and 1/3 cup of orange juice and place each in separate cups, bowls, or other containers.
5. Place all of the bowls and glasses on top of a piece of paper on which you have written your Mechanical Turk Worker ID (handwritten and visible).

Now, take a photo of the above items and upload it below. Try to match the following photo as closely as possible:

If you have a webcam, you may go to Cameroid (http://www.cameroid.com) to take a snapshot from your web browser, which you may then download and upload below.

IMPORTANT: Please **DO NOT DRINK/EAT** any of the ingredients at this time. Also, please remember to not eat or drink anything else during the survey unless prompted to do so. We will later ask you to consume the ingredients in a specific way.

Choose File No file chosen

Figure B.2: Follow-Up Survey: WTA Question

Imagine you had the mixed drink you made in the last survey prepared and ready to drink in front of you right now. For all of the listed amounts of money below, please indicate whether or not you would drink the mixed drink if you were paid that amount of money to do so.

	No I Would NOT Drink It	Yes I Would Drink It
	No	Yes
$0.00	○	○
$0.05	○	○
$0.10	○	○
$0.25	○	○
$0.50	○	○
$1.00	○	○
$2.00	○	○
$5.00	○	○

Figure B.3: Amusement Park Survey: Information Treatment

As a short detour from the survey, we would like to provide you with some information on the kind of weather you can typically expect in Orlando. Here is the monthly average high temperature, low temperature, precipitation (inches), and the number of days with at least 0.01 inches of rain:

Month	High (F)	Low (F)	Precipitation (Inches)	Days with Rain
January	71	49	2.35	6
February	74	52	2.47	6
March	78	56	3.77	7
April	83	60	2.68	5
May	88	66	3.45	8
June	91	72	7.58	14
July	92	74	7.27	16
August	92	74	7.13	16
September	90	73	6.06	13
October	85	66	3.31	8
November	78	59	2.17	5
December	73	52	2.58	6
Year	83	63	50.82	110

Source: Temperature averages (http://www.currentresults.com/Weather/Florida/Places/orlando-temperatures-by-month-average.php). Precipitation averages (http://goflorida.about.com/od/floridaweathe1/qt/Orlando-Weather.htm). Days with at least 0.01 inches of rain (https://www.sercc.com/climateinfo/historical/meanprecip.html).