



Università
Ca' Foscari
Venezia

**Dottorato di ricerca in Economia Aziendale
Scuola Superiore di Economia
Ciclo 23 (A.A. 2010 - 2011)**

Four Essays in Consumer Decision-Making

**SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: SECS-P/08
Tesi di dottorato di Gabriele Paolacci, matricola 955412**

Coordinatore del Dottorato

Prof. Massimo Warglien

Tutore del dottorando

Prof. Massimo Warglien

Co-tutore del dottorando

Prof. Katherine Burson

Acknowledgments

I owe much to my two advisors. Massimo Warglien has provided guidance throughout the dissertation process. Massimo is a very eclectic yet profound scholar, who always sees a bigger picture than yours. I will keep his example in mind when I find myself running ANOVAs without remembering why. Since I joined University of Michigan as a Visiting Student, Katherine Burson has been an exceptionally supportive mentor. The tough part with Katherine has always been to find novel words to thank her, and so it is right now. I wish I could email her and ask for advice.

Other people have helped along the way. I learned a lot by working with Scott Rick. I think Katherine, Scott, and I form a great team and I hope our collaboration will continue for years to come. Rick Bagozzi initially provided me with the opportunity to visit Ross School of Business, and since then has never stopped being generous and supportive. My time in Ann Arbor changed my life as a researcher and made me richer as a person.

Thanks to Fondazione Coin in Venice for the graduate research fellowship, and to Frank Ascione at the College of Pharmacy, University of Michigan, for funding part of my stay in Ann Arbor.

Personal relationships in Venice resisted physical distance, and I hope they'll continue to do so. They may not even know it, but I received more than a little help from my friends in surviving the Ph.D. If home feels "home" it is also because of them. See you at Galliano's.

I am grateful to my aunt Federica for always treating me as her favorite nephew, and not just like her only one.

Special thanks to Marta, for her love, support, and help. Our journey has been tough sometimes, but I can't think of a better reward than a future together.

Finally, I want to thank my parents Consuelo and Ennio, to whom this dissertation is dedicated. Mom and dad: If you are reading this, it means you learned some English. Your happiness when I had good news has been one more reason to work hard. Having you by my side during rough times has been vital. What I've been able to accomplish, I owe it to your love and unconditional support.

Table of Contents

Introduction to the dissertation	1
Inherent, constructed, revealed preferences: Guidelines for marketers	2
The Intermediate Alternative Effect: Considering a small tradeoff increases subsequent willingness to make large tradeoffs	28
Same world, different perceptions: Systems of measurement affect judgments	51
Online experimentation: Amazon Mechanical Turk	59

Introduction to the dissertation

This dissertation includes the research I conducted in the field of consumer decision-making. It is composed of four essays:

Essay 1 addresses the concept that lies at the very heart of Behavioral Decision Theory: Consumer preferences. I provide an original analysis of how contexts can determine the expression of preference. Finally, I suggest how marketers can measure consumers' inherent preferences, explore context effects, and utilize the resulting learnings to help consumers in their decision-making. The essay also serves as a general preamble to the empirical work that follows.

Essay 2 deals with the endowment effect, which is the tendency among consumers to keep their endowment when facing the opportunity to trade it for a “target” alternative. Across three studies, we show that this normatively irrational propensity is reduced when decision-makers are first provided with a choice between keeping their endowment and trading it for an intermediate alternative, i.e., an option that shares some characteristics with the endowment and some characteristics with the target alternative.

Essay 3 builds on research that showed that people tend to ignore unit size in evaluation. Investigating perceptions of gasoline efficiency among participants from the United States and Italy, we show that citizens of different countries react differently to the same objective variation because they use scales with different numbers of units.

Essay 4 puts together the work I conducted on Amazon Mechanical Turk as a subject pool for running online behavioral experiments. We show empirically that Mechanical Turk is a reliable source of experimental data. Further, we discuss its practical and theoretical advantages and highlight the challenges it puts in place for academic researchers.

Inherent, Constructed, Revealed Preferences:

Guidelines for Marketers

Gabriele Paolacci

Ca' Foscari University of Venice

Abstract

A theory of preferences was recently proposed on how consumers' choices depend on both stable "inherent" likings and sensitivity to framing, task, and context effects (Simonson, 2008). Managers can employ this approach to overcome the contradictions exhibited so far in their treatment of preferences, and ultimately improve the outcome of the marketing process. The essay proposes some practical methods to discover chronic dispositions and preference constructability, and suggests how to strategize over the inherent and the constructed components of consumer preferences.

Keywords: Consumer preferences, Behavioral Decision Theory, Framing effects, Task effects, Context effects.

1. Introduction

Although the concept of preference is central in economics and related fields, social scientists have been typically tacit about the very nature of preferences. In particular, economists typically conceived consumers as “rational” decision makers, who choose between options by retrieving well-defined preferences from memory. However, since the work of Tversky and Kahneman (1974; 1981; Kahneman and Tversky, 1979), researchers in behavioral decision theory have challenged this overly simplistic view and identified many heuristics and biases that consumers exhibit during the choice process. This “constructive” view of choice advocates that preferences are not well-defined, but malleable and contingent on the features of the problem at hand (Bettman, Luce, and Payne, 1998).

The marketing discipline echoes both the “rational” and the “constructive” view of preferences. For example, managers often let market research guide the initial stages of the marketing process (e.g., product design), thus presuming that consumers’ actual choices will reflect preferences that are stable over time and contexts. On the other hand, companies recognize that preferences depend on transient influences, for instance when they strategize over alternative distributive options (e.g., selling fine wine in specialty stores rather than in supermarkets). Although heterogeneous theoretical approaches have certainly enriched marketing practices, by the same token the absence of an integrated perspective on preferences may have ultimately undermined their overall effectiveness. In this essay, I outline how employing a recent theory on preferences, based on the concept of “inherent preferences” (Simonson, 2008), can help managers to provide the marketing process with internal consistency. Such theory posits that consumers have stable dispositions to like or dislike object components (including still unfamiliar or non-existing objects), which are neither determined nor affected by transient factors, e.g., the

composition of the choice set. Such factors operate “on the top” of inherent preferences, and contribute to determine the preferences that consumers ultimately reveal through choices. Whereas this theory was originally introduced as descriptive, the present essay points out its prescriptive implications. Given the evidence on the nature of preferences, I point out how managers should rethink about their preference-related decisions, and provide some advice to put the “inherent preferences” approach at work. I suggest original guidelines for marketers to employ in order to discover inherent preferences, explore how such preferences might be modified in the marketplace, and manage revealed preferences.

The essay is organized as follows. Section 2 provides a brief account of how marketing practices treated consumer preferences inconsistently, and presents Simonson’s theoretical approach as a potential way forward. Section 3 offers a perspective on the marketing process as implied by the existence of inherent and constructed preferences, and schematizes how managers should think about consumer choice. Section 4 suggests some methods to discover inherent preferences and preference constructability. Section 5 provides extensive advice to effectively manage consumer preferences. Section 6 concludes.

2. Consumer preferences in marketing

2.1. Rational versus constructed choices

Notwithstanding the importance of preferences in classic theories about consumption, economists were almost silent about their origin and essence. Mainstream economics depicts consumers as rational decision makers: When facing a choice, the consumer symbolically retrieves from memory his preferences for every available option, and chooses the option with the highest utility. Choices, within this framework, are “revealed preferences” (Samuelson, 1948). As advocated by Vilfredo Pareto in a 1987

letter, economists relied “as little as possible on the domain of psychology” and for the sake of generality and tractability they described preferences as well-defined and not dependent on the setting in which they are revealed, i.e., in which choices are made (Bruni and Sugden, 2007).

Over the last thirty years, behavioral decision theorists consistently reported instances of “preference reversals” (Lichtenstein and Slovic, 1971; 2006): situations in which, contrary to the predictions of rational choice theory, A is preferred to B under one circumstance and B is preferred to A under another, seemingly equivalent, circumstance. There is now large evidence that preferences are “constructed” (Bettman, Luce, and Payne, 1998), which means malleable and conditional on the features of the choice problem. Therefore, choices are potentially inconsistent across different settings. In particular, research has explored the effects on choice exerted by the description of the options (framing), the method of preference elicitation (task), and the composition of the choice set (context). I briefly describe each of these influences. (i) Framing effects. The description of the features of the choice options has a strong influence on preferences and choices (Tversky and Kahneman, 1981). In a frequently cited study, Levin and Gaeth (1988) framed ground beef as either 25% fat or 75% lean. Participants rated the taste of the beef significantly higher in the positive-“lean” condition than in the negative-“fat” condition, both with the framing preceding the tasting and vice versa. (ii) Task effects. Evaluating options jointly versus separately can lead to different choice outcomes (Hsee, 1996). For instance, consumers are more likely to prefer a low-price, low-perceived-quality brand when the choice task facilitates (versus impedes) the comparison with a high-price, high-perceived-quality brand (Nowlis and Simonson, 1997). (iii) Context effects. Preferences depend on the composition of the choice set (Huber and Puto, 1983; Pettibone and Wedell, 2000; Simonson, 1989). For instance, the presence (versus absence) of a “decoy” option which is clearly inferior to only one of the two other options in the choice set increases the

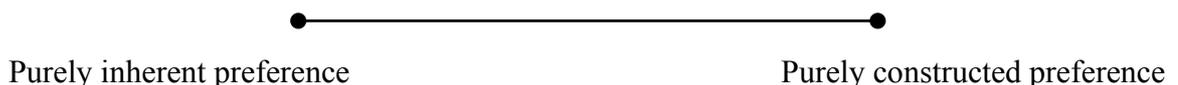
consumer's attraction for the dominating option with respect to the non-dominating option (Huber, Payne, and Puto, 1982).

So far, the marketing discipline has often dealt contradictorily with consumer preferences. On the one hand, some of the standard frameworks and methods of preference prediction, such as expectancy-value theory (Fishbein and Ajzen, 1975) and conjoint analysis (Green and Srinivasan, 1978) traditionally neglected any role for transient influences in shaping preferences. More in general, marketers implicitly regard preferences as predictable, and assume that preferences elicited during market research would correspond to preferences subsequently revealed by consumers through choice. On the other hand, marketers are also sensitive to the implications of behavioral decision research. Decoy options are a frequently used retailing tool, and products that are relatively superior on comparable dimensions (e.g., private labels, which have a lower price) are often strategically placed next to the relatively inferior options (Simonson, 1999). At best, this incongruence will result in potentially suboptimal results, e.g., because neglecting the existence of contextual factors during market research led to misaligned predictions about market outcomes. However, the lack of a clearly defined perspective on preferences can engender more dramatic consequences: For instance, if the expected performance of a product designed according to market surveys relies on a choice context that cannot be reproduced, the actual performance would be extremely poor, and in addition nobody would be clearly accountable for it. In sum, the disagreement between different theoretical perspectives on preferences gave rise to inconsistencies that, albeit often difficultly noticeable, may have affected the marketing process. An integrated view would help managers to deal coherently with consumers' preferences.

2.2. Resolving the dispute: Inherent preferences

Simonson recently proposed that people have inherent preferences for things they have or have not yet experienced (2008). Inherent preferences are relatively stable dispositions to like that are not determined by frames, tasks, or contexts. These factors, in turn, operate during the choice process and influence the choice outcome. Inherent preferences may be undiscovered (“dormant”), e.g., because of the unavailability of options that offer or communicate the preference objects (e.g., an ice cream flavor that still does not exist), or because of improper tests of the preference objects (e.g., a superficial test of an ice cream flavor). This theoretical proposition seems to hold empirically (Simonson and Sela, 2011), and is capable of a reconciliation of the notions of preference in rational choice theory and in behavioral decision theory. Every preference that is “revealed” through choice, according to Simonson’s approach, has an “inherent” component and a “constructive” component. Figure 1 represents revealed preferences as a continuum between “purely inherent” preferences, which like in rational choice theory do not depend on frames, tasks, and contexts, and “purely constructed” preferences, which are entirely determined by the influence of constructive factors. Moving from the left to the right side of the line, the relative weight of the constructed component is greater, resulting in a higher likelihood of preference reversals across choice settings.

Figure 1: Revealed Preferences



To see the argument applied, consider Martha and Rick, who are hanging out at a friend’s place and are offered an ice cream. They can choose between a chocolate ice cream and a lemon ice cream, and both Martha and Rick say they would prefer the

chocolate ice cream with equal conviction. That is, they reveal a similar preference for chocolate ice cream over lemon ice cream. At this point, the friend remembers he also has a strawberry ice cream, and he adds it to the choice set. While Martha still prefers the chocolate ice cream, Rick now decides to opt for the lemon ice cream. Therefore, whereas Martha displayed a stable preference for chocolate ice cream over lemon ice cream, the presence of strawberry ice cream in the choice set reversed Rick's previously revealed preference. This might have happened for a number of reasons: For instance, the presence of another fruit-flavored, low-calories option might have made Rick more sensitive to these characteristics, thus inducing him to choose its favorite ice cream between lemon and strawberry. On the other hand, Martha, who grew up eating a lot of ice cream, is very confident about her tastes, and in particular about her preference for chocolate over lemon (and strawberry). Therefore, the same act of revealing a preference had different degrees of malleability, and in particular the weight of the "inherent" component was probably higher for Martha than for Rick.

The previous example demonstrates the descriptive power of the "inherent preferences" approach, which lies in its capacity to synthesize the roles of stable dispositions and transient influences on choice (see also Kivetz et al., 2004; Tversky and Simonson, 1993). In the remainder of the essay, I will stress its prescriptive implications: The behavioral decision theorists' advice to avoid simplistic inferences about preferences across choice settings remains valid; however, I will argue that marketers need not to give up on the goal of predicting preferences. In fact, striving for a deeper interpretation of consumer preferences—that encompasses both their predictable "inherent" component and their transient "constructed" component—may allow managers to achieve consistency along the marketing process, and ultimately superior results.

3. A marketing perspective on preferences

3.1. Marketing in a nutshell

Simonson's theory of preferences can provide managers with a fresh perspective on the marketing process. The marketer's goal, in a hypothetical jargon of preferences, is the maximization of favorable revealed preference orderings, i.e., of choices of the company's products over competitors'. Since revealed preferences have an inherent and a constructed component, it follows that this goal needs to be accomplished through an optimal leverage of both: Put simply, marketers need to identify the product configurations that optimize consumers' "true" dispositions, and market them within frames, tasks, and contexts that maximize the consumer's likelihood to choose the product.

The inherent component of preferences is mostly exogenous to the company: It follows that results, in general, are highly dependent on an accurate assessment of inherent preferences. Therefore, at the early stages of the marketing process, managers should strive to discover inherent preferences, i.e., predict those likings and desires that do not depend on frames, tasks, and contexts. The core of the marketing strategy, in turn, should interpret at best the predicted inherent preferences. Later on along the process, managerial decisions mostly concern the frame, the task, and the context in which the product is marketed (e.g., place in the store, promotions). The company, therefore, should act in order to exploit the product's constructive potential: That means, marketers should try to increase the choice likelihood that would be implied by the consumers' baseline dispositions. This sketch of the marketing process is seemingly isomorphic to existing practices. Critically, however, marketers so far have not tried to predict inherent preferences, but revealed preferences in and of themselves, that by definition are contaminated by constructive factors. The outlined process, on the other hand, treats preferences consistently with their nature, with

benefits in terms of fit between otherwise disjointed actions, accountability of the personnel employed in the process, and –as I shall argue– potential results.

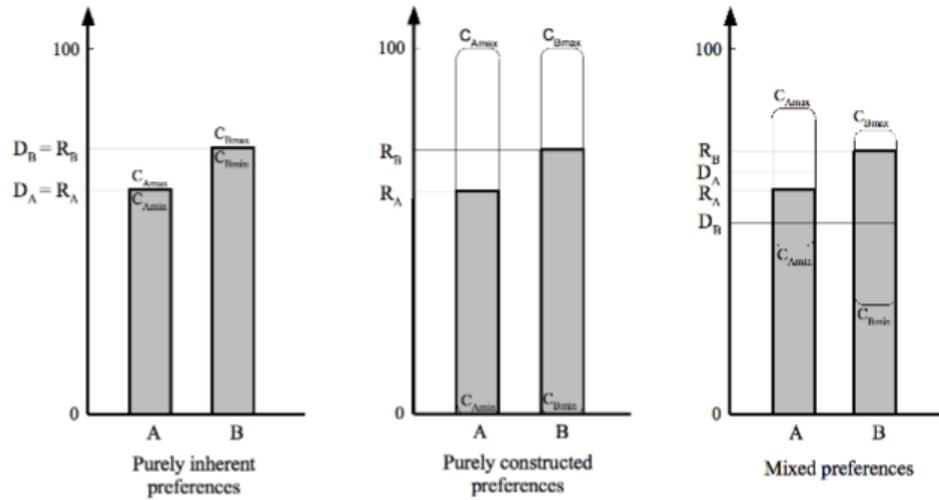
At this point, I suggested how a stereotyped marketing process that directly results from the “inherent preferences approach” could resolve the contradictory assumptions on preferences made de facto by marketing management. However, as we will see, the rule of thumb “first maximize inherent preferences, then construct at best” is only one of the available strategies implied by Simonson’s theory. Using a higher level of detail and sophistication, I now provide an original analysis of inherent, constructed, and revealed preferences, which directly supports the prescriptive implications discussed in the following sections.

3.2. Understanding preferences

In order to manage preferences effectively, marketers need to understand the role of their inherent and constructed components in a systematic manner. In general, the inherent liking for a certain product may have a high or a low degree of malleability, which expresses how much the consumer’s baseline disposition can be influenced by frames, tasks, or contexts, e.g., by the presence of a decoy option in the choice set. Figure 2 represents two options A and B; the consumer’s preference levels for such options can vary between two arbitrary values 0 and 100. In particular:

- D_A and D_B represent the consumer’s baseline dispositions for A and B;
- $[C_{Amin}, C_{Amax}]$ and $[C_{Bmin}, C_{Bmax}]$ represent the preference ranges in which D_A and D_B can be pulled by constructive influences, and they are marked as slightly rounded rectangles;
- R_A and R_B represent the consumer’s “final” preference levels for A and B. The ordering of R_A and R_B determines choice, i.e., the consumer chooses A if $R_A > R_B$, and chooses B if $R_B > R_A$.

Figure 2: Actual Choice and Underlying Preferences



In all the three panels in Figure 2, $R_B > R_A$, hence the consumer chooses B.

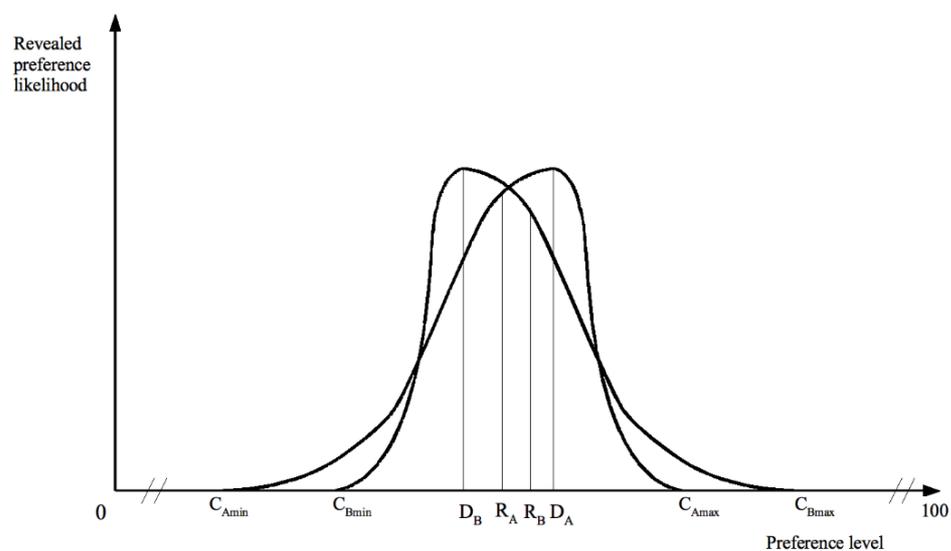
However, similarly to the ice cream example, the underlying preference configurations are different across panels: In the left panel, both intervals $[C_{Amin}, C_{Amax}]$ and $[C_{Bmin}, C_{Bmax}]$ are null, that means revealed preferences are “purely inherent” ($D_A = R_A$ and $D_B = R_B$). Constructive factors do not play any role in the determination of choice, and the consumer will choose B over A within any frame, task, or context. For instance, the likelihood of choosing chocolate ice cream over lemon ice cream would be unaffected by the presence (versus absence) of strawberry ice cream.

In the central panel, there exist no inherent likings D_A and D_B , and the final preference levels for A and B flow within the largest possible range ($[0, 100]$). Revealed preferences are “purely constructed”, such that the consumer is highly likely to prefer A to B within some circumstances, and B to A within some other circumstances. For instance, the choice of chocolate ice cream over lemon ice cream within a certain choice set (e.g., not including strawberry ice cream) would not be at all diagnostic of the same consumer’s choice within a different choice set (e.g., including strawberry ice cream).

The right panel depicts a stereotypical circumstance, in which revealed preferences

have both an inherent and a constructed component. The consumer has an inherent preference for option A over option B ($D_A > D_B$); however, the options' sensitivity to the constructive forces ($[C_{Amin}, C_{Amax}]$ and $[C_{Bmin}, C_{Bmax}]$) provides a good deal of uncertainty about what the consumer will actually choose; in fact, in the specific circumstance presented, constructive factors reverse inherent likings, making B more appealing than A. Even if the consumer inherently prefers chocolate ice cream to lemon ice cream, the presence of strawberry ice cream shifts his preference towards the latter alternative, ultimately making choice reverse baseline dispositions.

Figure 3: Asymmetry in sensitivity to constructive influences



So far, I described preference malleability as the degree to which inherent preferences can be pulled by constructive factors. However, managers might need to assess products' sensitivity to constructive factors in a more sophisticated fashion. In fact, products might respond asymmetrically to constructive influences: for instance, the liking for an option (e.g., chocolate ice cream, relative to lemon ice cream) might be more frequently boosted than decreased by decoy options (other ice cream flavors), e.g., because of its characteristics, or because of the unavailability of detrimental decoys. Similarly to

the right panel in Figure 2, Figure 3 presents a circumstance in which preferences are somewhat malleable. Given the consumer's baseline dispositions (D_A and D_B) and the ranges in which such dispositions can be pulled ($[C_{Amin}, C_{Amax}]$ and $[C_{Bmin}, C_{Bmax}]$), the vertical axis represents the likelihood that certain preference levels for A and for B will be ultimately achieved. As can be seen in the horizontal axis, the consumer's inherent liking for A is greater than his inherent liking for B ($D_A > D_B$). Notwithstanding that preference ranges for A and B are equal ($C_{Amax} - C_{Amin} = C_{Bmax} - C_{Bmin}$), the probability of a positive shift from inherent likings is higher for B than for A. This follows from the different skewness of the likelihood distributions of A's and B's final preference levels: This distribution is negatively skewed for A, that means A is more likely to suffer from constructive forces than it is to benefit from them; it is positively skewed for B, that means B is more likely to benefit than to suffer from framing, task, or context effects. Typically, such configuration of inherent and constructed preferences results in a choice of B over A ($R_B > R_A$): the constructive forces contrasted the relation between baseline dispositions to the point of reversing it. Managers shall strive to understand the likelihood of positive (vs. negative) shifts from inherent likings. A "statistical" assessment of products' susceptibility to constructive forces does not substitute a case-by-case prediction of the influence of specific frames, tasks, and contexts; however, as I will argue, regularities at the "constructed level" of preferences might be an important source of opportunities and threats.

In general, the present analysis suggests that marketers need to be aware that a high sensitivity to framing, task, and context effects can hugely affect the preference ordering that results from inherent likings. However, when placed within the described theory, context-dependence does not imply that context matching is the only solution to predict preferences (Bettman, Luce, and Payne, 1998): For instance, if inherent preferences were robust enough, choices would be consistent across frames, tasks, and contexts; moreover,

successfully predicting the effect of constructive factors at the beginning of the process would allow marketers to build effective strategies upon them.

4. Discovering inherent preferences and preference constructability

In Section 3, given the characteristics of preferences as described by Simonson (2008), I sketched a prototype of marketing process which is consistent with such theory, and described what information managers shall try to obtain in order to make such process effective and internally consistent. In this and in the following section, I focus on how managers should practically operate in order to reach this goal. How can marketers obtain a body of knowledge about inherent preferences and preference constructability that is, if not perfect, at least sufficient to inform adequately strategy makers? Given this knowledge, how should they try to maximize revealed preferences?

The following paragraphs suggest some methods that marketers may employ in order to (a) discover inherent preferences, i.e., those likings that are likely to be highly stable over time, and (b) discover preference constructability, i.e., how such likings might be shifted in the marketplace. This analysis represents the first, necessary step in the marketing process, and its results should help managers to choose the strategy that maximizes revealed preferences, and ultimately help the company reach its goal.

4.1. Discovering inherent preferences

In order to elicit inherent preferences, managers need to exclude any constructive influence from the elicitation process. Accordingly, traditional marketing techniques, such as focus groups, or surveys in which the respondent ranks some arbitrarily described options, are inadequate. Discovering inherent preferences, especially when dormant, might be extremely hard without inducing a direct experience with the preference objects.

Unfortunately, it is often complicated for marketers to make consumers try out the company's products before these are launched into the market. I suggest two "non-standard" methods to obtain data that are diagnostic of inherent preferences.

Online Configurators. If the company is interested in understanding the consumer's inherent preferences for some aesthetic features of a product, online product configurators might be effective instruments. Product configurators are self-design tools provided to customers in order for them to tie the product to their requirements and tastes. Deng and Hutchinson (2009) argue that product configurators "allow consumers to explore the design space in search of inherent preferences, experiencing their personal aesthetic reactions to each potential design". In other words, they provide "revealing experiences" prior to purchase. Product configurators can be often created easily and in fact are already integrated within many e-commerce websites, e.g., in the apparel industry. Managers should consider using this tool at earlier stages in the marketing process, to let consumers explore the range of feasible (and even unfeasible) possibilities for a product. The results might offer unique insights on what consumers would inherently prefer. Since even brainstorming is constrained by existing knowledge and mindsets, product configurators could be the only efficient substitute for the actual experience of a novel product, and help marketers discover dormant preferences. Because product configurators can be uploaded online, marketers can also obtain cheaply large amounts of heterogeneous respondents, using online labor markets such as Amazon Mechanical Turk (Horton, Rand, & Zeckhauser, 2011).

Implicit Association Test. Marketers can obtain insights into inherent preferences by studying consumers' automatic associations. The Implicit Association Test (IAT; Greenwald, McGhee, and Schwartz, 1998) is a task developed to study the strength of

automatic associations between concepts (e.g., racial groups) and attributes (e.g., good or bad), which are supposed to reveal implicit associations (e.g., racial prejudices). Typically, participants who are administered the IAT are given two sets of images representing concepts, and two lists of words: one with positive attributes and the other with negative attributes. Images and words appear randomly on a computer screen, and the computer records the time participants spend in order to assign concepts to attributes. When people are asked to associate a concept and an attribute that are not paired in their attitudes, they use more time to make the link, because they need to overcome an existing attitude. In this way, one can estimate participants' implicit attitudes towards concepts. The IAT can be used to measure implicit consumption-related constructs in an associative consumer social knowledge structure (Brunel, Tietje, and Greenwald, 2004). In particular, Forehand and Perkins (2001) showed that the IAT can reliably measure associations that occur outside of conscious awareness. Thus, IAT-based techniques might be useful for companies to discover inherent preferences that are dormant, or regularly overwhelmed by constructive influences. Marketers should identify concepts that might be tied to product features, and test the implicit attitudes of different segments of consumers towards these concepts. Any concept can be submitted to participants, including shapes, famous persons, or places. The IAT software is available for free, fully and easily modifiable, and can be administered online; therefore, as for product configurators, companies can exploit the Internet and online labor markets to obtain reliable data in a cheap and timely fashion.

Note that the proposed methods are not mutually exclusive; in fact, managers may want to employ them in parallel and cross-check the obtained results. Moreover, even if inherent preferences tend to be stable over time, marketers should keep in mind that baseline dispositions can evolve based on factors such as changes in lifestyle, priorities, and new information (Simonson, 2008). Reprising the earlier example, Martha may come

to prefer lemon ice cream to chocolate ice cream with no contextual influence, for instance because she perceives the former as more compatible with her increasing health-consciousness. Accordingly, inherent preferences should be investigated periodically.

4.2. Discovering preference constructability

Inherent preferences should be assessed jointly with their degree of sensitivity to constructive influences. The two analyses determine the scope of the strategic opportunities and threats that the company's product will face (see next section). Ideally, in order to predict preference constructability, managers would examine the influence of each combination of frames, tasks, and contexts on the consumer's baseline dispositions. Such operation, however, is virtually impossible due to its costs and complexity. I suggest two methods, drawn from academic research in decision-making, which might help managers to accomplish the task of discovering preference constructability.

Field experimentation. Researchers in decision-making usually explore framing, task, and context effects by running experiments in controlled environments like the laboratory, and defining choice options with clearly defined attributes. If companies were to do the same, they would probably obtain results that are scarcely transferable to their specific market. However, managers should not necessarily renounce to the benefits of experimentation. In fact, a high number of experiments in marketing have been recently carried out in the field (e.g., Aaker, Fournier, and Brasel, 2004; Kivetz, 2005). Running experiments in naturally occurring settings, like the marketplace, allows for a mixture of control and realism usually not achieved in the lab or with uncontrolled data (Levitt and List, 2009). Field experimentation can often be a viable strategy for marketers. For instance, it is straightforward to test whether the sales of a new edition of a book respond to whether its extra-length is framed in terms of absolute versus relative number of pages.

Similarly, companies' owned retailers can manipulate choice sets and observe how responsive consumers' preferences are to different tasks and contexts. Even when marketers have little control over the task and the context that consumers face (e.g., when the choice set necessarily include competitors' products), they can still benefit from the huge number of experiments that naturally occur in the marketplace: Managers need to pay attention to seemingly irrelevant changes in the purchase setting, that might affect (or not) consumers' preferences. By doing so, they would accumulate a great deal of evidence regarding products' sensitivity to constructive forces. In particular, marketers may be able to analyze how malleable preferences are, and whether malleability is inherently favorable or unfavorable for the company's product.

Genetics Studies. Researchers in a variety of fields, including economics and management (e.g., Cesarini et al., 2009; Nicolaou et al., 2008), are now conducting studies that link genetics and individual behavior and decisions. Recently, Simonson and Sela (2011) employed a classic twins study design to identify heritable effects on a broad range of judgment and choice phenomena, including likings for specific products and experiences. Comparing monozygotic twins and dizygotic twins, they found evidence of heritable individual differences for some preferences (e.g., dark chocolate, horror movies) and not for others (e.g., ketchup, tattoos). Although baseline dispositions do not depend exclusively on genetics, as the authors say, "the most inherent "inherent preferences" are probably those that consumers inherit." Therefore, the degree of constructability of a preference for an existing or potential product component might be partially explored by looking at preferences of monozygotic and dizygotic twins, who are usually recruited through public registries of twin pairs. More in general, managers should devote attention to the trend in marketing and decision science of investigating the role of genetics in determining individual "innate" attitudes, including attitudes toward products.

5. Managing revealed preferences

5.1. Strategizing over inherent and constructed preferences

Relying on a solid knowledge about consumers' inherent preferences and their constructability, managers can choose the optimal strategy to make the company's product most likely to be chosen by the consumer. By "strategy", I mean a high-level proposition for how the inherent and the constructed component of preference shall contribute to the achievement of an adequate liking on the marketplace. Such strategy would guide the marketing process and determine all decisions in terms of product, promotion, etc.

The strategy pursued by marketers would be placed somewhere in the continuum between two prototypical alternatives, based on the identified level of preference malleability. On one extreme, a very low sensitivity to frames, tasks, and contexts entails that these factors cannot alter significantly the preference relations determined by the consumer's dispositions to like; accordingly, marketers would need to aim at product configurations that are inherently preferred to most of the other feasible configurations; tying the product to suboptimal baseline dispositions would almost certainly imply poor results in terms of choices. This is not to say that managers should neglect other important decisions, such as those related to the context in which the product is sold. Rather, they would employ the previously explained paradigm "first maximize inherent preferences, then construct at best." However, results would be substantially dependent on the marketing actors in charge of translating the elicited baseline dispositions into product features. Given the importance of inherent preferences for this strategy, marketing intelligence would need to make sure that such relations actually remain stable over time.

The other prototypical situation is the one characterized by a very high sensitivity to constructive factors. If preferences are unstable, consumers are likely to be influenced by seemingly irrelevant characteristics of the choice setting. Although preference

malleability adds uncertainty to marketing results, it is not intrinsically bad for companies: For instance, notwithstanding inferior baseline dispositions with respect to the competitor's products, marketers might be able to overwhelm such preference relation using constructive factors. Indeed, Williams-Sonoma increased the sales of its \$275 bread machine by introducing to the catalogue a second, larger model at a price of just over \$400 (Simonson and Tversky, 1992). On the other hand, even if the company's product benefits from an inherent advantage over competitors', a high degree of malleability might put that advantage in danger. Marketers need to acquire the ability to understand when constructive factors, besides representing "tactical tools," can become crucial in the marketing strategy. For instance, the diffusion of websites that provide explicit comparisons of electronic equipment, notwithstanding high preference malleability among consumers, may stably favor brands with relatively good alignable features (Nowlis and Simonson, 1997). In such situations, managers might deliberately configure a product such that inherent likings for it are suboptimal, but which would largely and stably benefit from the influence of some constructive factors, e.g., a competing product whose constant presence in the choice set highlights the benefits of the company's product. Managers need to be particularly careful at deploying this strategy: Preferences that are revealed as favorable because of constructive factors might be potentially weak (Yoon and Simonson, 2008), and without a strong "true" underlying liking, the context might be decreasingly effective in a repeated choice setting. Moreover, the strategy needs to be shared with all the actors along the marketing process, otherwise some of them (e.g., retailers who are responsible for maintaining a favorable context) may refuse to take responsibility for unsatisfactory results. Finally, if companies do not have complete control over the context that benefits the product, marketing intelligence needs to be particularly efficient at monitoring the evolution of choice settings.

5.2. *Managing constructed preferences*

No matter the strategy pursued by the company, constructed preferences are an important component of choice. With no claim of being exhaustive, I suggest some ways in which marketers could harness the power of framing, task, and context effects and boost consumers' dispositions. The idea that managers could systematically "manage" consumers' preferences points to the debate about ethics within marketing practices. Although it is beyond the scope of this essay to discuss the ethical implications of the proposed guidelines, note that companies must market their products within certain frames, tasks, and contexts. In other words, even if they wanted to, managers could not elude decisions such as those described in the present section.

Frames. Marketers have a fair degree of autonomy in framing the company's offer. One important decision concerns how to frame the advantages the consumer would obtain by choosing the product: Persuasive messages could be "positive" (e.g., highlighting the benefits gained by choosing the product) or "negative" (e.g., highlighting the benefits foregone by not choosing the product). Whereas positively framed messages are more persuasive when consumers do not process them in detail, negatively framed messages are more persuasive when consumers approach them more analytically (Maheshwaran and Meyers-Levy, 1990). It follows that managers should try to boost the consumers' dispositions by achieving consistency between the level of involvement inherent in the product domain and the valence of the persuasive messages. Moreover, no matter the valence of the frame, message frames should emphasize the presence of behavior outcomes (i.e., benefits in positive frames, costs in negative frames) as opposed to the absence of behavior outcomes (i.e., costs in positive frames, benefits in negative frames; Zhao and Pechmann, 2006).

Apart from their valence, how should product attributes be presented and

quantified? When the product has more than one desirable attribute, the consumer should evaluate each of these attributes separately in order for its perceived importance to be increased. On the other hand, attributes that are perceived as losses should be integrated whenever possible (Thaler, 1985). Recently, it has been shown that attribute differences appear larger on scales with a higher number of units (Pandelaere, Briers, and Lembregts, 2011). Therefore, in those cases where it is desirable for an attribute to be large (e.g., warranty length), managers should favor expanded scales (months) over contracted scales (years). Moreover, whenever possible, managers should privilege the use of numbers that are naturally fluent (such as round numbers), which have been shown to induce positive affective reactions (King and Janiszewski, 2011). For similar reasons, the prints used to present the attributes can be manipulated to encourage or discourage an analytic processing of such attributes (Alter et al., 2007).

Product names in and of themselves can also contribute to frame the company's offer. For instance, a product name which is difficult to pronounce is perceived a signal of risk, which might be inherently good or bad depending on whether risk is a desirable feature within the product domain (Song and Schwarz, 2009). Making it explicit that a product has "balanced attributes" might induce consumers to think about the product as a "compromise," even when it is not actually a compromise in the choice set (Chernev, 2005).

All in all, framing effects might be more or less important, depending on the degree of preference malleability inherent in the situation. However, note that the opportunities offered by framing effects come with little if any cost.

Tasks and contexts. Contrary to the case of frames, managers may not have complete control over the tasks and the contexts in which the product is marketed, e.g., because competing products are also part of the choice set faced by consumers. In case of

substantial preference malleability, this entails a systematic and undesirable component of uncertainty in product performance that managers should try to reduce. Companies that cannot operate directly in the market should try to obtain the collaboration of intermediaries, thus allowing marketers to fully harness the opportunities of task and context effects.

Managers should try to market products within retail environments that promote the best task for the target product. If this is preferred when evaluated separately, retailers may physically isolate it from the other products (Nowlis and Simonson, 1997); vice versa, the environment should facilitate the most appropriate comparisons, stimulating favorable context effects: Since options in the choice set generally benefit from being perceived as “dominating” or “compromise” (Huber, Payne, and Puto, 1982; Simonson, 1989), the target product could be placed close to those options that induce such perceptions. In particular, because price is a product attribute that is easily comparable across options, it shall be used as a relevant dimension when designing product displays. However, depending on the product domain, there might be many attributes that are inherently desirable or undesirable, and can serve effectively as assortment display criteria.

Even when it is difficult to manipulate the choice set directly, managers may be nonetheless able to exploit task and context effects. For instance, companies can generally choose where and how to distribute their products: If a product is evaluated more favorably when considered separately from other similar products, distributing only one of its versions (e.g., high price, high quality) to each retailer might make consumers less likely to engage in joint comparisons with some of the most salient alternatives (products in the same line). Moreover, companies can select the retailers in which they want to market their product according to their assortments. Since choosing from larger assortments tends to be more difficult, when there is abundance of options consumers are more likely to choose hedonic products (as opposed to utilitarian; Sela and Berger, 2009) and products that excel

on attributes that are quicker and easier to assess (Lenton and Francesconi, 2010). It follows that companies that market these types of product should privilege retailers with large assortments, and vice versa. Once a product is distributed by a certain retailer, marketers can use promotion sales strategically: If consumers are accustomed to a certain price for the company's product (as compared to competitors' products), putting the product on sale will create a favorable contrast on the price dimension (Simonson and Tversky, 1992).

Most importantly, note that the influence of the context does not operate only at the retail level. Consumers often structure their own consideration sets well before getting directly in contact with products. The company's website, catalogues, advertising, and communication tools in general should promote a conception of the product assortment that is in line with the goals assigned to the target product, e.g., compromise option. With a careful choice of product names, marketers might also be able to leverage third parties' communication tools. Consider again those websites that provide explicit comparisons of electronic equipment. Using certain names (e.g., specifically ordered alphabetically), marketers may succeed in favoring certain comparisons and hinder certain other comparisons, thus recreating a favorable micro-context inside the website. In general, managers should consider the likely position of new products in the choice sets from the very stage of product design. Indeed, one of the advantages of the "inherent preferences" approach is that it allows constructive factors to become strategic means, rather than mere nuisance to be dealt with once the marketing process has reached its final stages.

6. Conclusion

Simonson's recent theory of consumer preferences attempted to synthesize the views of rational choice theory and behavioral decision theory. In the present essay, I

argued that marketing management, by employing an approach to preferences based on such theory, could overcome the contradictions exhibited so far in the marketing process, and ultimately improve its contribution to the company's goals. I elaborated on how choices in the marketplace may result from various combinations of "stable" dispositions and "on-the-spot" reactions to different frames, tasks, and contexts. Managers should acquire the capabilities to handle strategically these two preference components, in order to maximize the products' likelihood to be chosen by the consumer. I proposed some methods that may integrate the traditional managerial toolkit, suggested some high-level strategies that marketers might pursue, and highlighted specific marketing actions that can improve the product performance.

References

- Aaker, J., Fournier, S. & Brasel, S. (2004). When good brands do bad. *Journal of Consumer Research*, 31, 1: 1-17.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning. *Journal of Experimental Psychology: General*, 136, 4: 569-576.
- Bettman, J., Luce, M., & Payne, J. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25, 3: 187-217.
- Brunel, F., Tietje, B., & Greenwald, A. (2004). Is the Implicit Association Test a valid and valuable measure of implicit consumer social cognition?. *Journal of Consumer Psychology*, 14, 4: 385-404.
- Bruni, L. & Sugden, R. (2007). The road not taken: how psychology was removed from economics, and how it might be brought back. *Economic Journal*, 117, 516: 146-173.
- Cesarini, D., Dawes, C., Johannesson, L., Lichtenstein, P., & Johannesson, M. (2009). Genetic variation in preferences for giving and risk-taking. *Quarterly Journal of Economics*, 124, 2: 809-842.
- Chernev, A. (2005). Context effects without a context: Attribute balance as a reason for choice. *Journal of Consumer Research*, 32, 2: 213-223.
- Deng, X. & Hutchinson, J. (2009). Aesthetic self-design: Just do it yourself. Working Paper, Wharton School, University of Pennsylvania, Philadelphia.
- Dhar, R. & Novemsky, N. (2008). Beyond rationality: The content of preferences. *Journal of Consumer Psychology*, 18, 3: 175-178.
- Fishbein, M. & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction

to theory and research. Reading: Addison-Wesley.

- Forehand, M. & Perkins, A. (2001). Unconscious processing of spokesperson information: The influence of implicit cognition. *Proceedings of the Society for Consumer Psychology Conference* eds. Susan E. Heckler and Stewart Shapiro American Psychological Association 123-126.
- Green, P. & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 5, 2: 103-123.
- Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 6: 1464-1480.
- Hsee, C. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67, 3: 247-257.
- Horton, J.J., Rand, D.G., & Zeckhauser, R.J. (2011). The Online Laboratory: Conducting Experiments in a Real Labor Market. *Experimental Economics*, 4, 399-42
- Huber, J., Payne, J., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9, 1: 90-98.
- Huber, J. & Puto, C. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research*, 10, 1: 31-44.
- Kahneman, D. & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, 47, 2: 263-91.
- King, D. & Janiszewski, C. (2011). The sources and consequences of the fluent processing of numbers. *Journal of Marketing Research*, 48, 2: 327-341
- Kivetz, R. (2005). Promotion reactance: The role of effort-reward congruity. *Journal of Consumer Research*, 31, 4: 725-36.
- Kivetz, R., Netzer, O., & Srinivasan, V. (2004). Alternative models for capturing the compromise effect. *Journal of Marketing Research*, 41, 3: 237-257.
- Lenton, A. P. & Francesconi, M. (2010). How humans cognitively manage an abundance of mate options. *Psychological Science*, 21, 4: 528-533.
- Levin, I. & Gaeth, G. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15: 374-378.
- Levitt, S. & List, J. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53, 1: 1-18.
- Lichtenstein, S. & Slovic, P. (1971). Reversal of preference between bids and choices in gambling decisions, *Journal of Experimental Psychology*, 89, 1: 46-55.
- Lichtenstein, S. & Slovic, P. (2006). *The construction of preference*. New York: Cambridge University Press.
- Maheswaran, D. & Meyers-Levy, J. (1990). The influence of message framing and issue involvement. *Journal of Marketing Research*, 27, 3: 361-367.
- Nicolaou, N., Cherkas, L., Hankin, J., & Spector, T.D. (2008). Is the tendency to engage in entrepreneurship genetic?. *Management Science*, 54, 1: 167-79.

- Nowlis, S. & Simonson, I. (1997). Attribute-task compatibility as a determinant of consumer preference reversals. *Journal of Marketing Research*, 34, 2: 205-218.
- Pandelaere, M., Briers, B., & Lembregts, C. (2011). How to make a 29% increase look bigger: The unit effect in option comparisons. *Journal of Consumer Research*, 38(2), 308-322.
- Pettibone, J. & Wedell, D. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, 81, 2: 300-328.
- Samuelson, P. (1948). Consumption theory in terms of revealed preference. *Economica*, 15, 60: 243-253.
- Sela, A., Berger, J., & Liu, W. (2009). Variety, vice, and virtue: How assortment size influences option choice. *Journal of Consumer Research*, 35, 6: 941-951.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16, 2: 158-174.
- Simonson, I. (1999). The effect of product assortment on buyer preferences. *Journal of Retailing*, 75, 3: 347-70.
- Simonson, I. (2008). Will I like a “medium” pillow? Another look at constructed and inherent preferences. *Journal of Consumer Psychology*, 18, 3: 155-169.
- Simonson, I. & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29, 2: 281-95.
- Song, H. & Schwarz, N. (2009). If It’s Difficult to Pronounce, It Must Be Risky. *Psychological Science*, 20, 2: 135-138.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 4157: 1124-31.
- Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211, 4481: 453-58.
- Tversky, A. & Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39, 10: 1179-1189.
- Yoon, S. & Simonson, I. (2008). Choice set configuration as a determinant of preference attribution and strength. *Journal of Consumer Research*, 35, 2: 324-336.
- Zhao, G. & Pechmann, C. (2006). Regulatory focus, feature positive effect, and message framing. *Advances in Consumer Research*, 33, 100.

**The Intermediate Alternative Effect:
Considering a Small Tradeoff Increases Subsequent Willingness to Make
Large Tradeoffs**

Gabriele Paolacci	Katherine Burson	Scott Rick
<i>Ca' Foscari University of Venice</i>	<i>University of Michigan</i>	<i>University of Michigan</i>

Abstract

Prior research has consistently demonstrated that people are reluctant to trade a good they own for an alternative good, particularly when the alternative (or "target") represents a substantial departure from the "endowment." We demonstrate that the endowment effect can be reduced by first making participants consider trading their endowment for an intermediate alternative (which shares some characteristics of the endowment and some characteristics of the target). We find that this "intermediate alternative effect" operates primarily by shifting one's reference point in the direction of the target alternative. Even when the intermediate alternative is not adopted, the extent to which one's endowment is treated as a reference point is weakened, which can also facilitate subsequent trading.

Keywords: Endowment Effect, Loss Aversion, Decision Making, Reference Point, Prospect Theory

1. Introduction

Countless studies have demonstrated that ownership of a good makes decision-makers reluctant to trade it for an alternative (e.g., Kahneman, Knetsch, & Thaler, 1990). This “endowment effect” (Thaler, 1980) implies that consumers may make suboptimally conservative choices, such as keeping their current allotment (e.g., an inexpensive drug that has some side effects) although they might prefer an alternative (e.g., a more expensive drug that has no side effects) when asked to choose between the two. Endowment effects are most often interpreted as a manifestation of reference-dependent preferences that exhibit loss aversion—the tendency for losses to have greater hedonic impact than comparable gains (e.g., Thaler, 1980; Tversky & Kahneman, 1981, 1986). Loss aversion contributes to the widespread (and often costly) influence of defaults on decision-making (e.g., Johnson & Goldstein, 2003; Thaler & Sunstein, 2008).

We demonstrate a straightforward method to reduce the common reluctance to trade an endowment for a “target alternative.” We propose that the process of deciding whether or not to trade the endowment for an “intermediate alternative” (an option that possesses characteristics of both the endowment and the target alternative) subsequently reduces the extent to which trading for the target alternative is viewed as a loss, which in turn stimulates trading. We propose that this “intermediate alternative effect” operates primarily by shifting one’s reference point in the direction of the target alternative (among people who actually adopt the intermediate). However, even when the intermediate alternative is not adopted, the extent to which one’s endowment is treated as a reference point is weakened, which may also facilitate subsequent trading.

Theoretically, we build upon work on reference dependence and on the moderating effect of similarity on the endowment effect. We anticipate that adopting the intermediate alternative will shift one’s reference point toward the target alternative, which should make

the prospect of trading for the target alternative feel like less of a loss. This intuition stems from research suggesting that the similarity of alternatives may moderate the endowment effect (e.g., Chapman, 1998; Ortona & Scacciati, 1992; Tversky & Kahneman, 1991; Van Dijk & Van Knippenberg, 1998). In all such studies, the number of trades from an endowment to a target alternative increased with the similarity or substitutability between the endowment and the target. For example, Samuelson and Zeckhauser (1988) found that Harvard employees were more likely to choose a slightly modified medical plan than to choose a new plan as a substitute for their current plan. Reference dependence and loss aversion together explain this moderation: When the consumer examines an alternative to her endowment, its relative attractiveness depends on the magnitude of the entailed tradeoff such that small tradeoffs are usually preferred to large tradeoffs. More formally, in prospect theory terms (Kahneman & Tversky, 1979), it is generally the case that $v(x) + v(-x) > v(2x) + v(-2x)$. Thus, people who refuse to “leap” from their endowment to a distant target alternative might instead “walk” to it. Other work also supports the notion that people will be willing to make what they perceive as small tradeoffs. For example, there is some evidence that small gains actually loom larger than small losses (e.g., Harinck, Van Dijk, Van Beest, & Mersmann, 2007).

Intermediate alternatives may help to facilitate subsequent trade even when they are rejected. Knetsch and Wong (2009) recently argued that endowment effects are largely dependent on the extent to which the endowment conveys the reference state. In particular, they showed that the consideration of counterfactual endowments may undermine the strength of the individual’s reference point (i.e., the extent to which one’s endowment serves as the salient reference point against which potential trades are evaluated). This is consistent with recent work that has demonstrated that individuals with extensive trading experience, such as professional sports card dealers (List, 2003) and real estate investors (Genesove & Mayer, 2001), are less reluctant to part with their endowments than are those

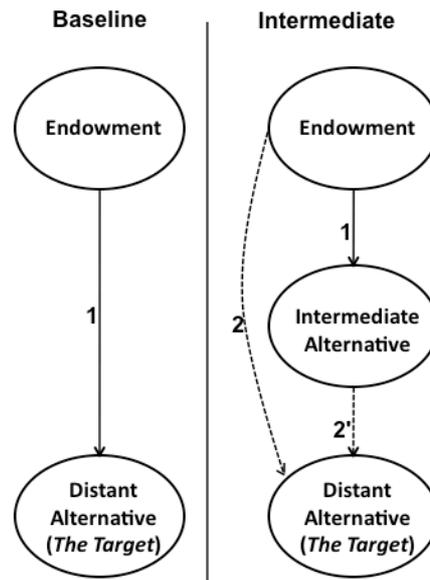
with less experience. Similarly, putting participants in a less emotional trading mindset has been shown to reduce loss aversion (Sokol-Hessner et al., 2009). Reconciling such evidence with a reference point account of the endowment effect (cf. Knetsch & Wong, 2009), We maintain that facing an opportunity to trade one's endowment for an intermediate alternative weakens the extent to which the endowment is treated as the reference point. This, in turn, should make the prospect of trading away one's endowment feel like less of a loss. In essence, rejecting the intermediate may subsequently make the decision whether to exchange one's endowment for the target alternative feel more like a choice between receiving the endowment and receiving the target alternative (i.e., help to turn "owners" into "choosers").

Of course, even if the reference point is weakened, it is unlikely that intermediate-rejectors will completely re-set their reference point (i.e., behave purely as choosers who have no endowment), since they still physically retain their endowment. However, intermediate-adopters, whose reference point has shifted toward the target, must endure only a small loss to obtain it. Moreover, if the loss is indeed perceived as small, the prospect of gaining the target might be weighted more heavily than the prospect of losing the intermediate (cf. Harinck et al., 2007), further increasing the likelihood of trading. Accordingly, WE anticipate that the overall intermediate alternative effect will be driven primarily by intermediate-adopters.

Taken together, the reasoning above leads to the following hypotheses:

- H1:** People who consider trading their endowment for an intermediate alternative will subsequently be more likely to trade for a distant (target) alternative than people who do not initially consider trading their endowment for an intermediate alternative.
- H1a:** Intermediate-adopters will be more likely to trade for the target than intermediate-rejectors.

Figure 1. The Basic Experimental Paradigm



Note that each arrow represents a single choice. There is one choice in the Baseline condition and two choices in the Intermediate condition. The second choice in the Intermediate condition depends on whether the intermediate alternative is adopted. If the intermediate alternative is rejected, the second choice is between the endowment and the target (link 2). If the intermediate is adopted, the second choice is between the intermediate alternative and the target (link 2').

We begin by examining the intermediate alternative effect and its underlying reference point dynamics using the basic paradigm illustrated in Figure 1. Imagine a consumer who is reluctant to trade her endowment for a distant target alternative even though she would prefer such an alternative in a simple binary choice. We propose that this endowment effect can be overcome by employing a two-step trading process from the endowment to the target alternative (see Intermediate panel of Figure 1). Specifically, consumers who would refuse to trade their endowment for the target alternative (path 1 in the Baseline panel of Figure 1) may ultimately do so by first considering trades for intermediate alternatives (path 1-then-2' or path 2 in the Intermediate panel). We argue that considering whether or not to trade one's endowment for an intermediate alternative subsequently reduces the extent to which trading for the target alternative is viewed as a loss, either by weakening the reference state conveyed by the endowment (if the

intermediate is rejected) or by shifting it toward the target alternative (if the intermediate is adopted). This will, in turn, increase the likelihood of trading to the target alternative.

2. Study 1: Trading Theater Seats

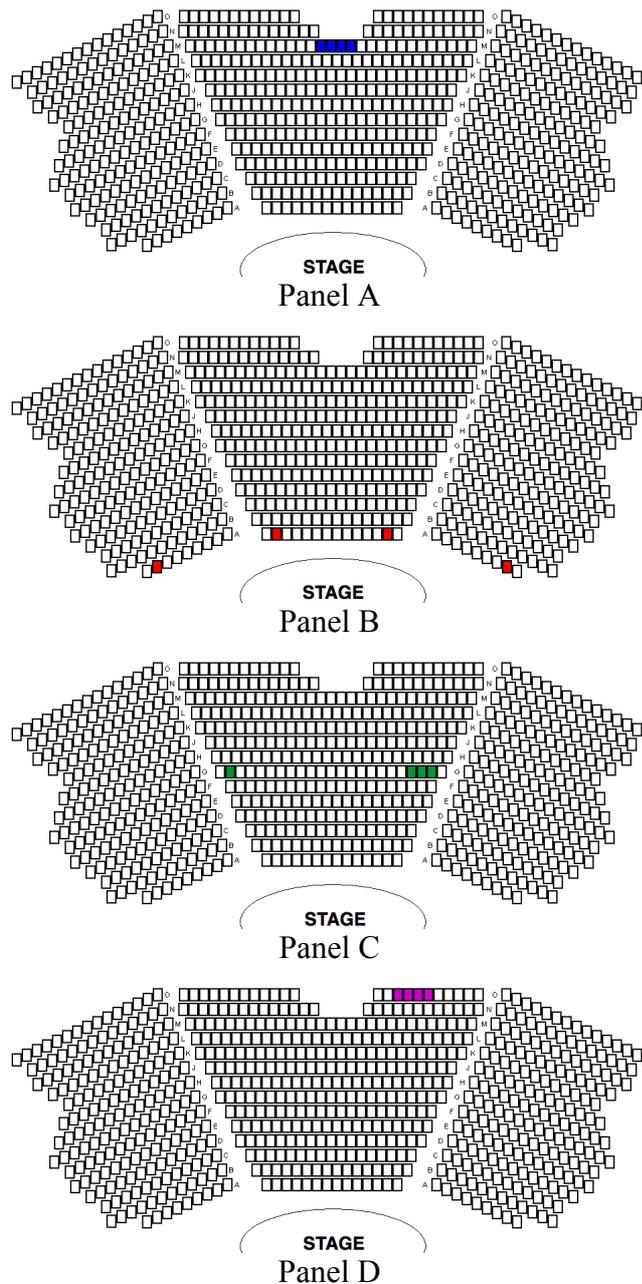
In Study 1 we examine whether there is an intermediate alternative effect by comparing trading rates (the proportion of people who ultimately adopt the target alternative) between a Baseline condition and an Intermediate condition (see Figure 1). In a hypothetical concert scenario, participants were endowed with seats in a theater and decided whether to trade for alternative seats characterized by tradeoffs (proximity to friends and proximity to the stage). Regardless of the features of the original endowment, we predicted that participants who had previously encountered an opportunity to trade the endowment for an intermediate alternative would be more likely to trade to the distant target alternative than those who had not encountered the intermediate alternative.

2.1. Procedure

One hundred twenty-two adults (mean age = 33.4, 59% female) participated via Amazon Mechanical Turk, an online service validated as a survey website (Horton, Rand, & Zeckhauser, 2011). Participants were asked to imagine that they and three of their friends would soon attend a concert of their favorite artist. Each participant was then assigned the hypothetical task of buying all four tickets based on two criteria: 1) the four seats should be as close as possible to the stage in order to have the best possible view; and 2) the four seats should be as close as possible to each other in order to enjoy the concert together. The cover story reported that at the time of purchase there was only one set of tickets available, and we counterbalanced across participants which endowment they received: either four seats that were superior on proximity to each other (Figure 2, Panel A,

henceforth referred to as Figure 2a), or four seats that were superior on proximity to the stage (Figure 2, Panel B, henceforth referred to as Figure 2b). (Complete instructions and stimuli are available from the authors upon request.)

Figure 2. Study 1 and Study 3 Theater Maps



We randomly assigned participants to either a Baseline condition or an Intermediate condition. Participants in the Baseline condition were simply provided with

the opportunity to keep their endowed set of seats or trade them for a set of seats that ostensibly just became available, namely the target alternative (those who were endowed with the seats in Figure 2a were offered the seats in Figure 2b, and vice versa). Participants in the Intermediate condition were initially given the opportunity to trade their endowed set of seats for an intermediate set of seats (Figure 2, Panel C, henceforth referred to as Figure 2c). The seats in Figure 2c are intermediate on both the “proximity to stage” dimension (closer to the stage than the Figure 2a seats, but farther than the Figure 2b seats) and the “proximity to friends” dimension (less clustered than the Figure 2a seats, but more clustered than the Figure 2b seats). All Intermediate participants (whether or not they decided to trade for the intermediate alternative) were then asked whether they wanted to trade their current seats for the target alternative.¹

Comparing trading rates in the Baseline and the Intermediate condition reveals whether or not consideration of an intermediate alternative attenuates the endowment effect. Regardless of whether participants naturally prioritize proximity to stage or proximity to friends, because we counterbalance whether participants are endowed with one option or the other, the classical economic prediction is that half of the participants in the Baseline condition should trade away their endowed set of tickets. Trading rates below 50% are indicative of an endowment effect.²

2.2. Results and Discussion

Consistent with prior research, we observed an endowment effect in the Baseline

¹ To make the task as clear as possible, participants were reminded of the position of their current seats throughout the experiment. In particular, while making their choices, participants could observe each alternative in a distinct map similar to the panels in Figure 2.

² To explain this prediction formally, imagine that a share p of all participants prefer the close-to-stage seats to the clustered ones. Then, assuming no reference dependence, in the condition where participants are endowed with close-to-stage seats, a share $1 - p$ should exchange the close-to-stage seats for the clustered seats; in the condition where participants are endowed with clustered seats, a share p should exchange clustered seats for close-to-stage seats. If the number of participants is the same across conditions, this would produce an average trading rate of $(p + 1 - p)/2 = 1/2$. Therefore, a trading rate lower than 50% is evidence of an endowment effect.

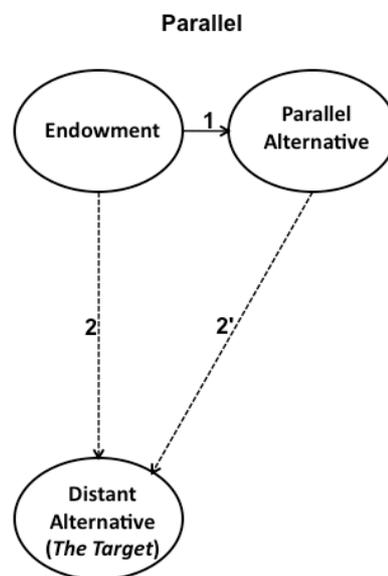
condition. Only 29.6% of participants in the Baseline condition chose to trade away their endowed set of seats for the target alternative seats, a proportion significantly lower than the normative benchmark of 50% ($\chi^2(1) = 8.96, p < .01$). However, consistent with H1, participants who were presented with an intermediate alternative were significantly more likely to trade. Among participants in the Intermediate condition, 47.1% ultimately traded their endowed set of seats for the target alternative seats. This trading rate is significantly higher than the 29.6% trading rate in the Baseline condition ($\chi^2(1) = 3.83, p = .05$). Moreover, 47.1% is not significantly different from the 50% normative benchmark ($\chi^2(1) = .24, p = .63$), meaning that the endowment effect is statistically non-existent in the Intermediate condition. In addition, in the Intermediate condition, participants who adopted the intermediate alternative were more likely to trade for the target than people who rejected the intermediate (66.7% vs. 41.5%, $\chi^2(1) = 2.97, p = .08$). This provides support for H1a and is consistent with our proposition that adopting an intermediate alternative does indeed shift the reference point toward the target, whereas rejecting an intermediate alternative can only weaken it.

3. Study 2: Trading Chocolate

The difference in subsequent trading between intermediate-adopters and intermediate-rejectors observed in Study 1 suggests that reference point shifting is the critical component of the intermediate alternative effect. However, it must be acknowledged that two things differ between intermediate-adopters and intermediate-rejectors: Unlike intermediate-rejectors, (i) intermediate-adopters have parted with the endowment and (ii) have obtained the intermediate alternative. The current results cannot pinpoint whether the difference between intermediate-adopters and intermediate-rejectors is primarily driven by (i) parting with the endowment or (ii) obtaining the intermediate. If

the trading-rate difference between intermediate-adopters and intermediate-rejectors is primarily driven by parting with the endowment, then reference point shifting is not the essential component of the intermediate alternative effect. Instead, that might suggest that the effect is driven primarily by reference point weakening (and that parting with the endowment by accepting the intermediate simply does a better job of reference point weakening than rejecting the intermediate).

Figure 3. The Parallel Alternative Condition



There are two choices in the Parallel condition. The second choice in the Parallel condition depends on whether the parallel alternative is adopted. If the parallel alternative is rejected, the second choice is between the endowment and the target (link 2). If the parallel alternative is adopted, the second choice is between the parallel alternative and the target (link 2').

Study 2 examines the relative contributions of parting with the endowment and obtaining the intermediate by introducing a Parallel condition (see Figure 3). Specifically, we examine the ability of a “parallel alternative” (which differs only superficially from the endowment) to increase subsequent trading. Notice that the act of adopting the parallel alternative requires parting with the endowment, but it cannot shift the reference point toward the target. If the intermediate alternative effect is primarily driven by the reference

point shifting toward the target when that intermediate is adopted, the ultimate trading rates should not increase when a parallel alternative is considered, since it cannot shift one's reference point toward the target, even if adopted. Indeed, if reference point shifting is the primary component of the intermediate alternative effect, the following predictions can be made:

- H2:** Relative to the Baseline condition, trading rates will not increase when a parallel alternative is considered instead of an intermediate alternative.
- H2a:** People who initially adopt an intermediate alternative will be more likely to trade for the target alternative than people who initially adopt a parallel alternative.

In addition to testing these hypotheses, Study 2 examines whether the intermediate alternative effect persists when participants must make real choices between goods. Specifically, participants were confronted with real choices between bags of customized milk chocolate M&M's. Participants traded between bags of M&M's that were characterized by tradeoffs (number of bags and aesthetic appeal).

3.1. Procedure

One hundred and twelve female participants at University of Michigan completed this study in exchange for a \$10 participation fee. They were randomly assigned to one condition of a 2 (endowment: university M&M's or spider M&M's) \times 3 (path to target: baseline, intermediate, or parallel alternative) between-participants design. Participants were initially endowed with either two small bags of M&M's customized with the logo of the university or with four small bags of M&M's customized with an unpleasant picture of a spider (see Figure 4). These two options were thus characterized by a relatively large tradeoff in terms of the quantity of chocolate and the aesthetic appeal of the candies (a pretest indicated that females found university M&M's more attractive than spider

M&M's). To reinforce participants' feeling of ownership for the M&M's they initially received, we first asked them to estimate the weight of a single bag (all bags weighed 1.15 oz.), and paid them \$1 at the end of the study if their estimate fell within ± 0.25 ounces of the actual weight.

Figure 4. Sample Stimuli from Study 2



From left to right: Solid university M&M's, outlined university M&M's, solid spider M&M's, outlined spider M&M's.

Participants in the Baseline condition saw no intermediate alternative and were simply asked if they wanted to trade their two bags of university M&M's for four bags of spider M&M's, or vice versa. Participants in the Intermediate condition were first given the opportunity to trade their endowment for an intermediate alternative composed of one bag of university M&M's and two bags of spider M&M's. After this choice, participants were offered the opportunity to trade their current allotment for the target alternative (two bags of university M&M's if they began with four bags of spider M&M's, or four bags of spider M&M's if they began with two bags of university M&M's). Participants in the Parallel condition were not initially offered an intermediate, but rather a slightly modified

version of their initial endowment. These parallel alternatives were identical to the initial endowment except that the image was slightly different. For instance, a participant endowed with four bags of solid brown spider M&M's was given the opportunity to trade for four bags of brown-outlined spider M&M's (Figure 4). Given its trivial difference from the endowment, there is little reason to suspect that adopting the parallel alternative would shift the reference point toward the target. If intermediate-adopters trade more for the target than intermediate-rejectors because their reference point is shifted, intermediate-adopters should also be more likely to trade for the target than parallel-adopters.

3.2. Results and Discussion

Consistent with prior research, we observe an endowment effect in the Baseline condition. Participants in the Baseline condition traded their endowed set of M&M's 25.0% of the time, indicative of a significant endowment effect ($\chi^2(1) = 9.00, p < .01$). By contrast, 47.2% of participants in the Intermediate condition traded their M&M's for the target alternative. This trading rate is significantly higher than the trading rate in the Baseline condition ($\chi^2(1) = 3.85, p < .05$), illustrating once again that considering an intermediate alternative substantially increases trading and eliminates the endowment effect (47.2% vs. 25.0%, $\chi^2(1) = 10.24, p = .002$).

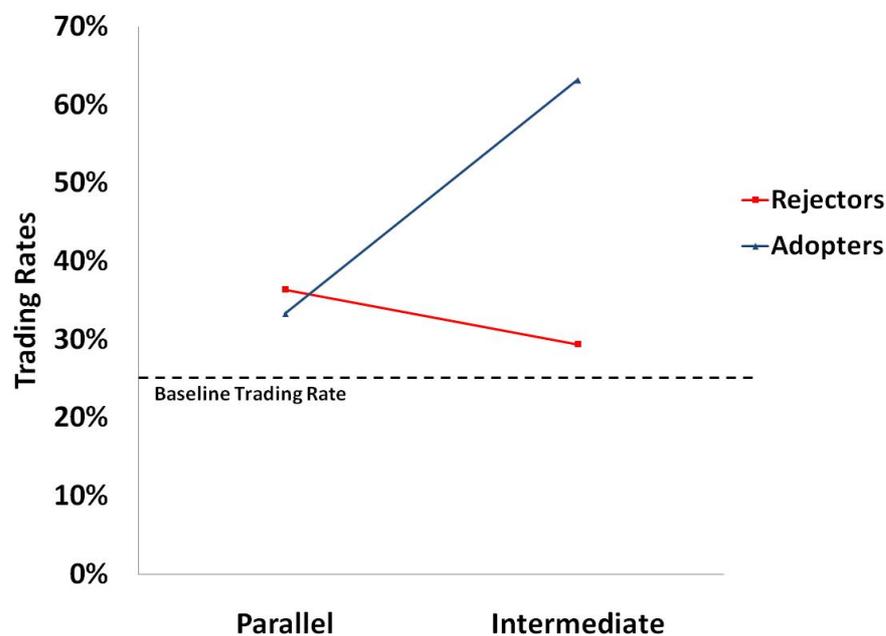
We next examined the pattern of trading for the Parallel condition. There, 35.0% of participants traded from the endowment to the target alternative. Although the difference in trading rates between the Parallel condition and the Baseline condition (35.0% vs. 25.0%) is directionally consistent with the idea that considering trading away the endowment weakens the reference point, this difference was not significant ($\chi^2(1) = 0.90, p = .34$). Indeed, consistent with H2, participants in the Parallel condition also exhibited an endowment effect (35.0% vs. 50.0%, $\chi^2(1) = 3.60, p = .06$).

To provide a focused test of H2a, we compare the behavior of adopters and

rejectors in the Intermediate and Parallel conditions. Figure 5 summarizes the results.

Consistent with Study 1, intermediate-adopters were significantly more likely to trade for the target than intermediate-rejectors (63.2% vs. 29.4%, $\chi^2(1) = 4.10, p < .05$). By contrast, parallel-adopters were no more likely to trade for the target than parallel-rejectors (33.3% vs. 36.4%, $\chi^2(1) = 0.04, p = .84$). Most importantly, whereas intermediate-rejectors and parallel-rejectors did not significantly differ in the extent to which they ultimately traded for the target (29.4% vs. 36.4%, $\chi^2(1) = 0.21, p = .65$), intermediate-adopters were nearly twice as likely as parallel-adopters to subsequently trade for the target (63.2% vs. 33.3%, $\chi^2(1) = 3.29, p = .07$). Because adopting the parallel alternative can only weaken one's reference point, but not shift it in the direction of the target, this large difference suggests that reference point shifting is the critical component of the intermediate alternative effect.

Figure 5. Trading Rates for the Target by Condition in Study 2



Given the demonstrated importance of reference point shifting via intermediate-adoption, it is worth considering whether, and to what extent, reference point weakening via intermediate-rejection actually contributes to the intermediate alternative effect.

Though we have shown that reference point shifting is critical, it is difficult to precisely measure the relative contributions of shifting versus weakening. It is tempting to compare trading rates for the target among intermediate-rejectors and Baseline participants. Across studies, intermediate-rejectors were about 4-12 percent more likely to trade for the target than Baseline participants (41.5% vs. 29.6% in Study 1; 29.4% vs. 25.0% in Study 2). However, this is essentially an apples-to-oranges comparison, because we know that all intermediate-rejectors prefer the endowment to the intermediate, but some Baseline participants would undoubtedly have preferred the intermediate to the endowment. Arguably, then, this 4-12 percent estimate understates the extent to which reference point weakening contributes to the overall effect. (The appropriate, but impossible comparison would be between intermediate-rejectors and Baseline participants who would have rejected the intermediate.)

The Parallel condition is another relevant source of data. Since the parallel alternative differs only slightly from the endowment, accepting or rejecting it is unlikely to do anything other than weaken one's reference point. (Indeed, the difference in trading rates for the target between parallel-adopters and parallel-rejectors does not even approach significance; $p = .84$.) The overall 10% difference in trading rates for the target between Parallel participants and Baseline participants (35% vs. 25%) provides another estimate of the benefits of reference point weakening. Thus, the results suggest that weakening plays a small, but non-negligible role in the overall effect.

Study 2 served several purposes. Using real products, we replicated the intermediate alternative effect and, as in Study 1, found that it was primarily driven by intermediate-adopters rather than intermediate-rejectors. In addition, by introducing a Parallel condition, we were able to determine whether the intermediate alternative effect is primarily driven by reference point shifting or by reference point weakening. The large trading difference between intermediate-adopters and parallel-adopters suggests that effect

is primarily driven by reference point shifting.

One limitation of Studies 1 and 2, however, is that they measure only participants' behavior. The trading data is consistent with our proposed process of reference point weakening and shifting, but does not directly measure those reference points. Therefore, in Study 3 we use a more sensitive measure of participants' reference points: We measure continuous satisfaction ratings that likely correspond more closely to reference points than binary trading decisions.

4. Study 3: Judging Theater Seats

In Study 3 we used a concert scenario similar to Study 1 to provide a more sensitive test of the precise influence of intermediate alternatives on reference points. We manipulated whether or not participants were endowed with a set of seats, and among people who were endowed with seats, we manipulated whether or not they were offered an alternative that represented a small tradeoff (for a total of three between-participants conditions). We then asked all participants how satisfied they would be with a set of seats that was dominated by the endowment. Degree of satisfaction with the dominated alternative arguably reflects the perceived distance between a participant's current state and the dominated alternative (such that a larger distance between the two would lead to lower satisfaction with the dominated alternative). If reference points are shifted by adopting the intermediate alternative, dissatisfaction with a dominated option should be greatest among intermediate-adopters. If rejecting an intermediate weakens one's reference point, the satisfaction ratings of intermediate-rejectors should be similar to the satisfaction ratings of non-endowed participants.

4.1. Procedure

Two hundred adults (mean age = 35.6, 61% female) recruited on Amazon Mechanical Turk were randomly assigned to one of three conditions: Baseline, Intermediate, and No Endowment. In the Baseline condition, participants were endowed with a set of side-by-side seats that were moderately far from the stage (Figure 2a). They were then asked to imagine that they had to move to a set of dominated side-by-side seats in the back row (even further back than the endowed set; Figure 2, Panel D), and were asked how satisfied they would be with those seats on a 1-7 scale (1 = extremely unsatisfied, 7 = extremely satisfied). In the Intermediate condition, participants were first asked whether they wanted to move to a set of less clustered seats in the middle of the theater (the intermediate alternative in Study 1; Figure 2c), and were then asked how satisfied they would be with the dominated seats. In the No Endowment condition, participants were simply asked how satisfied they would be with the dominated seats.

4.2. Results and Discussion

Consistent with reference dependence, participants in the Baseline condition were significantly less satisfied with the dominated seats than participants in the No Endowment condition (3.85 vs. 4.90, $t(123) = 2.11, p < .05$). However, there was no significant difference between participants who considered and rejected the intermediate alternative and participants in the No Endowment condition (4.37 vs. 4.90, $t(120) = 1.13, p = .26$). These patterns suggest that, as predicted, rejecting the intermediate weakened the reference point (though not completely re-setting it to a state like that before they were endowed). In addition, participants who considered and adopted the intermediate alternative were less satisfied with the dominated seats than Baseline participants (2.83 vs. 3.85, $t(78) = 1.87, p = .07$), suggesting their reference point had shifted as a result of the trade.

These results fully support our weakening and shifting reference point account of

the intermediate alternative effect. Rejecting an intermediate alternative to one's endowment weakens the reference state instilled by the endowment, and adopting such an alternative shifts the reference point. Taken together with the results from Studies 1 and 2, the results strongly suggest that intermediate alternatives reduce the endowment effect by weakening reference points (if rejected) and shifting reference points (if adopted).

5. General Discussion

Too often, consumers are stuck with choices they inherited from the past, rejecting favorable alternatives to their endowments because the losses they perceive are too difficult to ignore (e.g., Okada, 2001). We demonstrate that reluctance to trade can be reduced by first providing the consumer with an opportunity to trade the endowment for a smaller tradeoff option. Through consideration of that intermediate alternative, participants abandoned their endowments and adopted alternatives they likely would have rejected if forced to trade for them directly.

This essay joins a growing body of literature that identifies context-based boundaries of the endowment effect (e.g., Lerner, Small, & Loewenstein, 2004) and contributes to our understanding of the endowment effect by showing that an incremental presentation of alternatives attenuates it: In our studies, participants who considered owning an intermediate alternative before considering the target alternative showed no endowment effect. We maintain that this pattern results from a reference point mechanism that makes the trade for the target alternative feel like less of a loss: Rejecting the intermediate alternative weakens the reference state conveyed by the endowment, and adopting the intermediate shifts the reference point from the endowment toward the target.

This work also joins a growing body of research that attempts to shed light on how reference points develop and change over time (e.g., Heffetz & List, 2011; Köszegi &

Rabin, 2006). For example, the results of Study 3 suggest that reference points are not strictly defined by ownership (cf. Ericson & Fuster, in press; Peck & Shu, 2009).

Some readers may see parallels between the current research and findings on context-dependent choices (Huber, Payne, & Puto, 1982). The present findings do not contradict existing research on the compromise effect (Simonson, 1989): Whereas the compromise effect posits that an intermediate alternative benefits from the presence of an extreme alternative (in terms of choice share), we demonstrate that introducing an intermediate before introducing an extreme alternative can subsequently benefit the extreme alternative. Similarly, these findings are not a replication of the attraction effect: Whereas the attraction effect illustrates how the addition of a dominated option can pull choice share to an extreme option, we always use a non-dominated intermediate option.

Readers familiar with the classic foot-in-the-door (FITD) phenomenon, whereby willingness to comply with a large request is increased when preceded by a small request, may also see overlap between this effect and the intermediate alternative effect. However, a closer look reveals several distinctions. Note that the intermediate alternative effect is a way to increase willingness to make tradeoffs between losses and gains, whereas FITD is a way to increase willingness to accept losses. Specifically, while we present participants with the opportunity to trade something desirable for something they may or may not find more desirable, FITD researchers attempt to coerce participants into behavior that they find unpleasant, such as allowing strangers into their home (Freedman & Fraser, 1966).

In addition, FITD researchers do not distinguish between people who comply and do not comply with the initial request (adopters and rejectors, in the language of this essay). When it is reported, the proportion of participants who comply with the first request is typically very high (e.g., 100%, Burger & Petty, 1981, p. 497). Indeed, in a meta-analysis of FITD studies, Burger (1999, p. 305) found only one study in which the majority of participants did not comply with the initial request. High compliance with the

initial request is not surprising given the lack of anonymity in FITD paradigms (requests are typically made in person or over the phone) and the fact that complying with the request does not require parting with a salient reference state (unlike owners deciding whether to trade away an endowment). Moreover, all participants, whether they comply with the small request or not, are lumped together and included in all analyses (e.g., Burger, 1999; Freedman & Fraser, 1966). By contrast, we demonstrate that intermediate alternatives operate differently depending on whether or not they are adopted.

Finally, the FITD and intermediate alternative effects appear to be driven by different processes. Whereas FITD is typically attributed to self-perception, psychological reactance, a need to remain consistent, and several other factors (Burger, 1999), those factors are not needed to explain the effectiveness of intermediate alternatives. Instead, intermediate alternatives appear to work by shifting or weakening reference points.

Future research should delve more deeply into the process by which adopting an intermediate alternative influences one's reference point. Our current data suggest that adopting the intermediate shifts the reference point. It is possible, however, that the reference point shift is actually a more nuanced process of initial weakening due to mere consideration of the trade and then shifting via adoption of the intermediate alternative. Comparing a Baseline condition like those utilized in these studies to an Intermediate-to-Target condition, in which participants start with an intermediate alternative and decide whether to trade for the target, would be informative. If trading from the endowment to the intermediate both shifts and weakens one's reference point, subsequent trading to the target should be greater among intermediate-adopters in the Baseline condition than among participants in the Intermediate-to-Target condition.

Our findings have practical relevance for marketers and policy makers. We suggest that when consumers are reluctant to undertake big changes, "smoothing" the transition process to the final state might help. There are a number of products and services that can

be introduced and adopted (or not) in an incremental fashion, the most salient example being digital goods: Unless alternatives are very compelling, websites and software manufacturers should consider introducing new versions of their existing offers through incremental updates to minimize the net impact of perceived losses.

For policy makers, this proposal might be utilized as a “nudge” (Thaler & Sunstein, 2008) to overcome the problem of suboptimally conservative choices. Offering people small trades eases the path to a potentially beneficial but controversial alternative without forcing that trade. Consider, for instance, the problem of insufficient saving rates among American households. Since households are accustomed to a certain level of disposable income, they may be reluctant to save more money (even when they know they should) because they do not want to experience a “cut” in take-home pay (Benartzi & Thaler, 2004). Our results suggest that firms could implement successive contribution plans that only slightly increase employees’ saving rates but eventually lead to more aggressive saving.

In conclusion, although much is understood about loss aversion and the endowment effect, many open questions remain (cf. Johnson, Gachter, & Herrmann, 2006; McGraw, Larsen, Kahneman, & Schkade, 2010; Rick, 2011). Our work contributes to answering those questions by explicating the role of weakened versus shifted reference points in overcoming endowment effects. Practically, this essay demonstrates a simple way to reduce the endowment effect. Consumers who consider trading their endowment for an intermediate alternative subsequently demonstrate no endowment effect when deciding whether to trade for a more distant alternative. Thus, to encourage tradeoffs, marketers and policy makers are advised to help people walk toward the desired outcome rather than trying to force them to leap.

References

- Benartzi, S., & Thaler, R. H. (2004). Save More Tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(1), S164-S187.
- Burger, J. M. (1999). The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and Social Psychology Review*, 3(4), 303-325.
- Burger, J. M., & Petty, R. E. (1981). The low-ball compliance technique: Task or person commitment? *Journal of Personality and Social Psychology*, 40(3), 492-500.
- Chapman, G. B. (1998). Similarity and reluctance to trade. *Journal of Behavioral Decision Making*, 11(1), 47-58.
- Ericson, K. M., & Fuster, A. (in press). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *Quarterly Journal of Economics*.
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4, 195-202.
- Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics*, 116(4), 1233-1260.
- Harinck, F., Van Dijk, E., Van Beest, I., & Mersmann, P. (2007). When gains loom larger than losses. *Psychological Science*, 18(12), 1099-1105.
- Heffetz, O., & List, J. A. (2011). Is the endowment effect a reference effect? *NBER Working Paper No. w16715*.
- Horton, J. J., Rand, D.G., and Zeckhauser, R.J. (2011). The Online Laboratory: Conducting Experiments in a Real Labor Market. *Experimental Economics*, 4, 399-425
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1), 90-98.
- Johnson, E. J., Gachter, S., & Herrmann, A. (2006). Exploring the nature of loss aversion. Working paper, Columbia University.
- Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649), 1338-1339.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98(6), 1325-1348.
- Knetsch, J. L., & Wong, W. (2009). The endowment effect and the reference state: Evidence and manipulations. *Journal of Economic Behavior & Organization*, 71(2), 407-413.
- Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4), 1133-1165.
- Lerner, J., Small, D., & Loewenstein, G. (2004). Heart strings and purse strings: Effects of specific emotions on economic transactions. *Psychological Science*, 15(5), 337-341.

- List, J. A. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, 118(1), 41-71.
- McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science*, 21(10), 1438-1445.
- Okada, E. M. (2001). Trade-ins, mental accounting, and product replacement decisions. *Journal of Consumer Research*, 27(4), 433-446.
- Ortona, G., & Scacciati, F. (1992). New experiments on the endowment effect. *Journal of Economic Psychology*, 13(2), 277-296.
- Peck, J., & Shu, S. (2009). The effect of mere touch on perceived ownership. *Journal of Consumer Research*, 36(3), 434-447.
- Rick, S. (2011). Losses, gains, and brains: Neuroeconomics can help to answer open questions about loss aversion. *Journal of Consumer Psychology*, 21 (4), 453-463.
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7-59.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2), 158-174.
- Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences*, 106(13), 5035-5040.
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39-60.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4), S251-S278.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, 106(4), 1039-1061.
- van Dijk, E., & van Knippenberg, D. (1998). Trading wine: On the endowment effect, loss aversion, and the comparability of consumer goods. *Journal of Economic Psychology*, 19(4), 485-495.

**Same World, Different Perceptions:
Systems of Measurement Affect Judgments**

Gabriele Paolacci
*Ca' Foscari University
of Venice*

Enrico Rubaltelli
University of Padua

Teresa Gavaruzzi
Leeds University

Abstract

Recent research has shown that people tend to ignore unit size in evaluation. We show that citizens of different countries react differently to the same objective variation because they use scales with different numbers of units. We investigated perceptions of gasoline efficiency among participants from the United States and Italy. By comparing the country's divergence in satisfaction judgments provided in the presence versus the absence of quantitative information, we isolated the marginal effect of systems of measurement on judgments. Consistent with predictions, the metric system induced significantly higher satisfaction ratings regarding efficiency improvements than the American system.

Keywords: Size Discrimination, Systems of measurement, Cross cultural differences, Numerosity heuristic, Unit effect

1. Introduction

Recent research has explored how numbers and scales used to describe quantities affect related judgments and decisions (Larrick & Soll, 2008; Wong & Kwong, 2000). There is substantial evidence that semantically equivalent information can be perceived differently depending on the scales used to describe such information. For instance, Furlong and Opfer (2009) found that players' decisions in a prisoner's dilemma game are affected by the numeric representation of the economic rewards. Representing the reward of cooperation as 300¢ leads to higher cooperation rates relative to representing the reward as \$3. Wertenbroch, Soman, and Chattopadhyay (2007) showed that participants perceived more expensive name brands as carrying a smaller price premium relative to corresponding private labels when both were priced in less numerous currency (euros) than when both were priced in more numerous currency (pesetas). Burson, Larrick, and Lynch (2009) showed that people grant more weight to an attribute when it is expressed on "expanded" scales as opposed to "contracted" scales. Participants evaluated two movie-rental plans with tradeoffs in terms of availability of new movies and price. Participants were more likely to choose the cheaper plan when the number of new movies was presented on a contracted scale (new movies per week) and were more likely to choose the more expensive plan when the number of new movies was presented on an expanded scale (new movies per year).

Pandelaere, Briers, and Lembregts (2011) provided more definitive evidence on the process underlying what they call the "unit effect": Consistent with the demonstrated tendency to rely on numerosity as a cue for judging magnitudes (Pelham, Sumarta, & Myaskovsky, 1994), people tend to ignore unit sizes and thus overinfer quantities from the number of units. Accordingly, the same differences in an attribute appear larger on scales

with many units than on scales with fewer units, which can affect judgments and preferences related to that attribute. For instance, the authors find that expressing the energy content of snacks in kilojoules rather than kilocalories (1 kcal = 4.18 kJ) increases the choice of a healthy snack. In sum, unless individuals are induced to consider counterfactual units, they will generally react more strongly to an attribute change when such change is described on an expanded scale relative to a contracted scale.

In the current research, we document a fundamental consequence of people's tendency to ignore unit size in evaluation. We show that judgments depend on the culturally relevant system of measurement. Countries differ in the standard units commonly used to convey quantitative information: While citizens in the United States consistently use United States customary units ("American system"), most of the world relies on the International System of Units ("metric system"). Because of the conversion rate between scales, equal levels or variations in distance, volume, etc., are described using numbers of different magnitudes. For instance, a change in distance is described by a larger numerical change when using kilometers than miles, and a change in volume is described by a larger numerical change when using liters than gallons. We hypothesize that this will cause citizens of different countries to react differently to the same objective variation because they use different systems of measurement. In particular, since scales in the metric system can generally be considered as expanded scales relative to scales in the American system, we predict that judgments made by people who use the metric system (e.g., continental Europeans) will be more reactive to changes than judgments made by people who use the American system (e.g., Americans).

We conducted a cross-cultural investigation about perceptions of gasoline efficiency with participants from the United States and Italy. By comparing the country's divergence in judgments provided in the presence versus the absence of quantitative information, we isolated the marginal effect of systems of measurement on judgments.

Specifically, we varied the system of measurement of volume (Study 1a) and distance (Study 1b). When such variations were described with the assistance of units, American participants considered them in the American system (gallons, miles) while Italians used the metric system (liters, kilometers). Intrinsically, “liters-volume” and “kilometers-distance” are expanded “miles-distance” and “gallons-volume” scales, therefore changes should seem larger to Italian participants than to American participants. Accordingly, whereas we expected no differences in satisfaction judgments regarding efficiency improvements in the absence of quantitative information, we hypothesized that satisfaction ratings made in the presence of quantitative information would be significantly higher for Italians than for Americans.

2. Study 1a

2.1. Method

One hundred fifty-nine participants (mean age = 34.7, 59% female) recruited from online panels in the US and Italy completed this study. In a 2 (country: US vs. Italy) x 2 (quantitative information: present vs. absent) design, each participant was shown a 200-word essay about ways to reduce gasoline consumption. The essay described some behavioral changes (e.g., checking tires more often) that, if implemented, would help car drivers to reduce gasoline consumption. In the control conditions, no quantitative information about the potential reduction was presented and Americans and Italians read identical essays. When quantitative information was presented, the volume benefit (gasoline saved per week) was described using either gallons or liters, depending on the standard units of the participants' countries. Therefore, the reduction in gasoline consumption was described by a larger numerical interval for Italians (34.5 to 30.7) than for Americans (9.1 to 8.1). After reading the essay, participants indicated their satisfaction

with the potential benefit on an 18-point scale (1 = *Not at all satisfied*, 18 = *Very satisfied*).

2.2. Results

As predicted, systems of measurement affected judgments of satisfaction with the reduction in gasoline consumption (Table 1). There was no interaction between the presence of quantitative information and country, $F(1,159) = .63, p = .429$. However, planned contrasts revealed that whereas in absence of quantitative information satisfaction ratings did not differ by country, $t(155) = .89, p = .373$, in the presence of quantitative information satisfaction ratings were significantly higher for Italians, who evaluated volume in liters, than for Americans, who evaluated volume in gallons, $t(155) = 2.27, p = .025, d = 0.36$. In other words, systems of measurement affected participants' judgments by making Italians more satisfied than Americans with the same reduction in gasoline consumption.

Although these results support our hypothesis, we observed surprisingly large standard deviations in the conditions providing no quantitative information. This might suggest that participants were confused by the description and thus unreliable in their satisfaction judgments. Though we regarded this as unlikely, we conducted a new study to confirm Study 1a results.

Table 1: Means (and standard deviations) of Satisfaction Ratings

Condition	Study 1a (Reduced Consumption)		Study 1b (Increased Distance)	
	Americans	Italians	Americans	Italians
Presence of Units	14.7 (3.37)	16.4 (2.59)	14.0 (4.09)	17.4 (1.20)
Absence of Units	14.0 (4.40)	14.8 (4.01)	16.2 (2.49)	16.4 (2.34)

3. Study 1b

3.1. Method

One hundred sixty-five online panelists (mean age = 33.3, 67% female) recruited from online panels in the US and Italy completed this 2 (country: US vs. Italy) x 2 (quantitative information: present vs. absent) study. Participants read an essay identical to the essay in Study 1a except that the described benefit was an increase in the distance covered by the car on a full tank. In the presence of quantitative information, such an increase was numerically larger for Italians (419.7 to 1387.7 kilometers) than for Americans (260.8 to 862.3 miles). Participants indicated their satisfaction with the benefit on an 18-point scale (1 = *Not at all satisfied*, 18 = *Very satisfied*).

3.2. Results

The results confirmed our prediction (Table 1). There was a significant interaction between country and presence of quantitative information, $F(1, 165) = 13.40, p < .001, \eta^2 = .07$. Once again, contrasts revealed that whereas in the control conditions participants' satisfaction with the increase in distance did not differ between countries, $t(71.51) = -.31, p = .760$, in the presence of quantitative information satisfaction ratings were significantly higher for Italians than for Americans, $t(52.23) = 5.28, p < .001, d = 1.46$. These results confirm that systems of measurement affect people's evaluation of semantically equivalent information.

4. General discussion

The results suggest that systems of measurement are not neutral carriers of information, but affect people's judgments in a predictable manner. Specifically, changes

in distance and volume are perceived differently depending on the culturally relevant standard units: Whereas Italians and Americans exhibited the same reactions to the essays when no quantitative information was presented, judgments were more extreme for Italians than Americans when participants were presented with quantitative information. Although changes were semantically equivalent, the numeric intervals used to convey them were not, making changes appear larger to Italians than to Americans.

The consistent use of different standard units might ultimately have a role in shaping cross-cultural divergences in how the public thinks and acts on global economic and environmental issues. For instance, Americans might read about the dieback of the Amazon rainforest or the decay of the Greenland ice sheet and be less struck than Europeans merely because square miles are contracted with respect to square kilometers. Note that incidentally, as it is the case with the present stimuli, numerical variations are generally larger when using the metric system than the American system. Therefore, one could speculate that citizens in most of the world are likely to exhibit stronger reactions than Americans to phenomena that are typically described with measurement units.

Our findings also suggest that cross-cultural researchers in consumer psychology should be careful when comparing the magnitude of the effects obtained using stimuli that are described with different units of measurement. The results from the two studies clearly show that the mere presence of quantitative information introduces a possible confound between the effect of measurement units and the effect hypothesized by the researcher in a cross-cultural study.

Future research should explore the boundary conditions of the effect of systems of measurement on judgments and evaluations. People who are familiar with different systems of measurement might be less likely to use the number of units as a cue for judging quantities. Similarly, presenting individuals with additional quantitative information that are not dependent on systems of measurement (e.g., percentages) might

attenuate the unit effect.

References

- Burson, K. A., Larrick, R. P., & Lynch, J. G. (2009). Six of One, Half Dozen of the Other. *Psychological Science, 20*, 1074-1078.
- Furlong, E. E., & Opfer, J. E. (2009). Cognitive Constraints on How Economic Rewards Affect Cooperation. *Psychological Science, 20*, 11-16.
- Larrick, R. P., & Soll, J. B. (2008). The MPG Illusion. *Science, 320*, 1593-1594.
- Pandelaere, M., Briers, B., & Lembregts, C. (2011). How to Make a 29% Increase Look bigger: The Unit Effect in Option Comparisons. *Journal of Consumer Research, 38*(2), 308-322.
- Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The Easy Path From Many to Much: The Numerosity Heuristic. *Cognitive Psychology, 26*, 103-33.
- Wertenbroch, K., Soman, D., & Chattopadhyay, A. (2007). On the Perceived Value of Money: The Reference Dependence of Currency Numerosity Effects. *Journal of Consumer Research, 34*, 1-10.
- Wong, K. F. E., & Kwong, J. Y. Y. (2000). Is 7300 m Equal to 7.3 km? Same Semantics but Different Anchoring Effects. *Organizational Behavior and Human Decision Processes, 82*, 314-333.

Online Experimentation: Amazon Mechanical Turk

Gabriele Paolacci

Jesse Chandler

Pam Mueller

*Ca' Foscari University
of Venice*

Princeton University

Princeton University

Abstract

Amazon Mechanical Turk has become increasingly popular among social scientists as a source of experimental data. Behavioral scientists have noticed the ease with which online workers can be recruited and paid using crowdsourcing services and have begun using Mechanical Turk intensively. We report the results of a comparative study involving classic experiments in judgment and decision-making, which found no differences in the magnitude of effects obtained using AMT and using traditional subject pools. We further discuss the advantages and drawbacks of AMT, with particular attention to those that seem to be overlooked by the academic community.

Keywords: Online research; Crowdsourcing; Judgment and Decision-Making

1. Introduction

Crowdsourcing is an increasingly popular method of allocating and managing labor. Just as businesses have used the web to outsource labor, a number of websites have been developed to aid specific academic projects (Gaggioli & Riva, 2008). For example, reCaptcha verifies that website users are human by asking them to transcribe distorted images of words, and also helps digitize illegible portions of books (von Ahn et al., 2008). Galaxy Zoo (www.galaxyzoo.org) solicits “citizen-scientists” to view and classify astronomical images. For businesses with smaller or shorter-term projects, a number of companies (e.g., Crowd Cloud, Amazon Mechanical Turk) offer various ways to access large pools of workers to complete more modest tasks. These sites are primarily used by businesses seeking to outsource business tasks, but social scientists have increasingly become interested in crowdsourcing as a viable alternative to traditional methods of participant recruitment. Behavioral researchers have noticed the ease with which workers can be recruited and paid and have begun using them to recruit experimental participants. In particular, approximately half of the respondents to a survey reported below (all belonging to the academic community) reported conducting experiments using AMT as a subject pool.

The present essay reports the research we conducted with the aim of testing the reliability of AMT as a subject pool. The structure of the essay is somewhat unusual, as it covers two distinct waves of research that were conducted in two distinct periods. Section 2 introduces the AMT service and presents the demographics of AMT workers. Section 3 reports the results of a comparative study involving classic experiments in judgment and decision-making, which found no differences in the magnitude of effects obtained using AMT and using traditional subject pools. This research was conducted in the first half of 2010 with the goal of “validating” AMT, when crowdsourcing was not yet diffuse among

social scientists. Since then, work on AMT as a research tool has been intense in the fields of economics, psychology, and computer science. The remainder of the essay critically reviews the most updated evidence on the pros and cons of recruiting experimental participants on AMT, including results from both other researchers' work and from a new wave of original research that was motivated by such review. Both advantages (Section 4) and causes of concern (Section 5) of the use of AMT are discussed, and we highlighted in particular those features that have been overlooked by the academic community. Section 6 concludes.

2. AMT: The service and the demographics

AMT was named after an 18th century chess playing “automaton” that was in fact operated by a concealed person. It is an online labor market where employees (called *workers*) are recruited by employers (called *requesters*) for the execution of tasks (called *HITs*, acronym for Human Intelligence Tasks) in exchange for a wage (called *reward*). Both workers and requesters are anonymous although responses by a unique worker can be linked through an ID provided by Amazon. Requesters post HITs that are only visible to workers who meet predefined criteria (e.g., country of residence or accuracy in previously completed tasks). When workers access the website, they can sort tasks according to various criteria, including size of the reward and maximum time allotted for the completion. Workers can read brief descriptions and see previews of the tasks before accepting to work on them.

Tasks are typically simple enough to require only a few minutes to be completed such as image tagging, audio transcriptions, and survey completion. More complicated tasks are typically decomposed into series of smaller tasks including the checking and validation of other workers' HITs. Once a worker has completed a task, the requester who

supplied that task can pay him. A requester can reward good work with bonuses and punish poor quality work by refusing payment or even blocking a worker from completing future tasks.

Rewards can be as low as \$0.01, and rarely exceed \$1. Translated into an hourly wage, the typical worker is willing to work for about \$1.38 an hour (Horton & Chilton, 2010). It is reasonable to assume that only those in poor countries would be willing to work for such low wages. In fact, the available workforce is composed of hundreds of thousands of individuals in a wide distribution of ages, ethnicities, levels of education and income. Most workers are female (65%), are currently living in the United States or India, and are generally young (Ipeirotis 2010). Reflecting the more general demographics of younger cohorts, they also tend to be more educated but earn less than the average person in their country of origin. Despite these overall patterns, there is considerable heterogeneity. Workers range in age from 18 to more than 70 and the income distribution parallels that of their country of origin (see Ipeirotis, 2010 for worker demographics)

3. Comparing AMT workers and more traditional research participants

We conducted numerous replications of traditional findings in judgment and decision-making on AMT, suggesting that it is reliable (see <http://experimentalturk.wordpress.com>). The study presented in this section extended these findings by directly comparing AMT data with data collected from other sources. Subjects were recruited from three different sources: AMT, a traditional subject pool at University of Michigan, and visitors of online discussion boards. The study (carried out in April and May 2010) provides additional evidence on the consistency between AMT workers and more traditional subjects, with respect to both actual behavior and attention provided to the experimental tasks.

3.1. The survey

Subjects completed three classic experimental tasks drawn from the heuristics and biases literature. The survey was completed using Qualtrics survey software. Questionnaires were identical across conditions with the exception that AMT workers were asked to provide a code automatically generated by Qualtrics at the end of the experiment. The *Asian disease problem* (Tversky & Kahneman, 1981) demonstrates framing effects. Subjects had to choose what action plan between a riskier and a safer one they preferred in order to contrast the outbreak of an unusual disease. In a between-subjects manipulation, the outcomes of the plans were either framed in positive terms (people saved), or in negative terms (people lost). The *Linda problem* (Tversky & Kahneman, 1983) demonstrates the conjunction fallacy, that is the fact that people often fail to regard a combination of events as less probable than a single event in the combination. Respondents read a description of Linda and rated which of two alternative profiles was more likely to describe Linda, with one being more general than the other. The *physician problem* (Baron & Hershey, 1988; experiment 1, cases 1 and 2) demonstrates the outcome bias, the fact that stated judgments of quality of a decision often depend on the valence of the outcome. Subjects rated on a seven-point scale (Stanovich & West, 2008) the quality of a physician's decision to perform an operation on a patient. The operation was described as either a success or a failure in a between-subjects design.

After completing these three tasks, subjects completed the Subjective Numeracy Scale (SNS; Fagerlin et al., 2007). The SNS is an eight-item self-report measure of perceived ability to perform various mathematical tasks and preference for the use of numerical versus prose information. Because of its high correlation with the numeracy measure (Lipkus et al., 2001), the SNS provides a parsimonious measurement of an individual's quantitative abilities. Therefore, evidence of low quantitative score on the

SNS may raise some concerns regarding the actual capacity of workers in AMT to appreciate the magnitude of the wages/effort ratio in listed HITs. Moreover, the SNS provided an ideal context for a catch trial that measured whether participants were attending to the questions. Included with the SNS, subjects read a question that required them to give a precise and obvious answer (“*While watching the television, have you ever had a fatal heart attack?*”). This question employed a six-point scale anchored on “*Never*” and “*Often*” very similar to those in the SNS, thus representing an ideal test of whether subjects paid attention to the survey or not.

3.2. *The samples*

AMT. The task required workers to complete an externally hosted survey in exchange for \$0.10. The HIT was titled “Answer a short decision survey” and described as “Make some choices and judgments in this 5-minutes survey”. The (overestimated) completion time was included in the HIT description in order to provide workers with a rough assessment of the reward/effort ratio. The actual ratio was \$1.71/hour. The HIT was visible only to workers with an acceptance rate greater than 95% and who were residents in the U.S. One hundred thirty-one workers took part in the study.

Lab subject pool. One hundred and forty-one students from an introductory subject pool at University of Michigan completed this study.

Internet Discussion Boards. A link to the survey was posted to several online discussion boards which host online experiments in psychology. The survey has been available online for two weeks, and one hundred thirty-seven visitors took part in the study.

3.3. Results

Subjects' demographics. Overall 318 subjects participated in the study (66.0% female, $M_{\text{age}} = 28.3$). Subjects recruited from the Internet boards were comparable in terms of average age to subjects recruited from AMT (30.6 and 34.3 respectively) and unsurprisingly, both were older than subjects recruited from the lab subject pool (18.8).

Table 1: Subject pools characteristics

Subject Pool	% Females	Avg. Age	Median Age	Subjective Numeracy (SD)	% Failed Catch trial	% Survey completion
Mechanical Turk	75%	34.3	29	4.35 (1.00)	4.17%	91.6%
University of Michigan	68.8%	18.8	19	4.17 (0.81)	6.47%	98.6%
Internet Boards	52.6%	30.6	26	4.25 (1.16)	5.26%	69.3%

Non-response error. We looked at the number of people who accessed to the study but did not complete it entirely. As expected, almost everybody in the lab subject pool completed the study (98.6%). Subjects recruited from online discussion forums were significantly less likely to complete the survey than subjects on AMT (66.7% and 91.6% respectively), $\chi^2(1,268) = 20.915, p < .001$. This suggests that AMT strongly diminishes the potential for non-response error in online research.

Attention. Subjects in the three subject pools did not differ in terms of attention provided to the survey. Subjects in AMT had the lowest catch trial failing rate (defined as the proportion of subjects who did not select “Never” to the question “While watching the television, have you ever had a fatal heart attack?”), although the number of respondents who failed the catch trial is very low and not significantly different across subject pools, $\chi^2(2,301) = .187, p = 0.91$. Subjects who failed the catch trial, or did not reach the page containing the catch trial, were removed from subsequent analyses.

Subjective Numeracy. Participants in the three subject pools did not differ significantly in the SNS score, $F(2,299) = 1.193, p = 0.30$. As the SNS is closely associated with many measures of quantitative ability, this result suggests that workers in AMT are not less able to handle quantitative information (e.g., payments for participation) than more traditional experimental participants.

Experimental Tasks. Table 2 summarizes the results obtained in the experimental tasks. The present tasks, along with their variations, are widely used in judgment and decision-making, and in particular they had already been posted repeatedly on AMT (<http://experimentalturk.wordpress.com>; Horton et al., 2011). Therefore, for each task we excluded from the analysis subjects who declared they previously completed the task.

Table 2: Results on experimental tasks

	Mechanical Turk	University of Michigan	Internet Boards
<i>Asian Disease</i>			
% Risky Positive Frame	17.6%	28.1%	23.7%
% Risky Negative Frame	55.3%	67.7%	63.0%
χ^2	10.833	20.230	13.013
p	< 0.001	< 0.001	< 0.001
Effect size (w)	0.39	0.39	0.39
<i>Linda's problem</i>			
% Conjunction Fallacy	72.2%	78.3%	64.4%
<i>Physician's problem</i>			
Avg. Quality Success (SD)	5.93 (0.81)	5.63 (0.75)	5.73 (0.98)
Avg. Quality Failure (SD)	5.13 (1.24)	4.86 (1.29)	4.93 (1.41)
t	3.70	4.14	2.547
p	< 0.001	< 0.001	0.007
Effect size (d)	0.76	0.73	0.66

In the Asian disease problem, people were significantly more likely to choose the risky course of action when the problem was framed in terms of losses than when it was framed in terms of gains. Effect sizes are exactly the same across samples. Note that subjects on AMT exhibited more risk aversion than subjects in the other subject pools,

although this did not occur in previous tests of the same problem

(<http://experimentalturk.wordpress.com>; Horton et al., 2011).

Respondents in all subject pools exhibited the conjunction fallacy. Large majorities regarded a combination of events (“Linda is a bank teller and is active in the feminist movement”) as more probable than a single event in the combination (Linda is a bank teller”). There were slight differences across samples for this effect, $\chi^2(2,274 = 4.606, p = 0.1$, however this is consistent with the large variability of results in the conjunction fallacy literature (e.g., Charness, Karni, & Levin 2009).

Subjects in all the subject pools showed an outcome bias. In the physician problem, subjects judged the quality of the physician decision to be higher when it was followed by a success than when it was followed by a failure. The result is significant in all the subject pools, and the effect size in AMT is the highest among the three samples.

Overall, the results of the comparative study confirmed that AMT is a reliable source of experimental data in judgment and decision-making. Results obtained in AMT did not substantially differ from results obtained in a subject pool at University of Michigan. Moreover, response error was significantly lower in AMT than in Internet discussion boards.

Most research conducted after our comparative study has continued to emphasize similarities between participants recruited from AMT and participants recruited from traditional sources. This allowed important differences to be overlooked. In fact, some of these differences are cause for concern: participants remain members of crowdsourcing websites for longer than participants remain members of traditional subject pools, increasing the likelihood that they will complete many similar studies. Further, unlike Internet panels in which a company determines which respondents are assigned to complete individual surveys, AMT workers self-select into the tasks they wish to complete. Self-selection is always a problem, but AMT participants compound this by

completing surveys regularly, and by tracking requesters and sharing information about them as they seek the most lucrative and interesting opportunities. The remainder of the essay integrates a review on the extant literature on AMT with the new research we conducted to identify the unexploited opportunities of AMT and its challenges.

4. Advantages of using AMT

4.1. Commonly Known Advantages of Using AMT to Recruit Research Participants

A convenience sample of researchers who belong to the Society for Personality and Social Psychology, and the Society for Judgment and Decision Making mailing lists completed a brief survey on “methods of data collection.” Respondents ($N = 369$; 122 graduate students, 48 post docs, 147 faculty) were asked to select methods that they use to recruit face-to-face and online samples. Approximately half ($N = 182$) reported using crowdsourcing to collect online data, with this method particularly common among graduate student (60%) and relatively less common among associate and full professors (31%). As a point of comparison, about 60% of participants reported using undergraduate subject pools to recruit participants to complete online samples, 40% reported using convenience samples, 26% reported using a purpose built website and 15% reported using online panels.

Respondents were asked to list the advantages and disadvantages of using AMT in a free response format. The most frequently mentioned advantages included the speed and affordability of data collection (58% and 37% respectively), which parallels the emphasis of these attributes in the crowdsourcing literature (Buhrmester, Kwang, & Gosling, 2011; Chilton et al., 2009; Mason & Watts, 2009). Further, empirical investigations suggest that data quality seems to be virtually independent from pay rates (Buhrmester et al., 2011; Marge, Banerjee & Rudnicky, 2010; Mason & Watts, 2009). Payments are easily

distributed to workers en masse and there is no need to maintain payment receipts or other income tax records. AMT allows you to submit an Employer Identification Number and will generate the relevant tax documentation for workers. Thus the payment system is streamlined and administrative burdens are low.

As reported in Section 2, an additional appealing feature of AMT is the relative diversity of the workforce compared to traditional college students (mentioned by 31% of respondents). Likewise, political affiliation and support for various public policies often parallel the distribution observed in high quality panel data, sometimes more closely than other convenience samples and always more so than student samples (Berinsky, et al., 2011). Most importantly, the size of the workforce and the ability to screen workers before hiring them makes it comparatively easy to find nearly any kind of sample, including those that are not Western, educated, industrialized, rich, and Democratic (“WEIRD”; Heinrich, Heine, & Norenzayan, 2010).

Finally, AMT has a number of features that allow researchers to avoid certain threats to experimental validity that are difficult to avoid in other Internet recruiting techniques (see Horton et al., 2011 for a detailed discussion). First, AMT by default prevents the same worker from completing a given HIT more than once – although as will be discussed later, this may not eliminate duplicate workers entirely. Second, selection bias and attrition are less of a concern in samples recruited from AMT than from other online convenience samples. Studies typically find that workers are more concerned with the financial reward of completing HITs than its contents (Horton et al., 2011). As a result, HITs can be structured so that workers must accept them without substantive knowledge of their contents, before receiving additional information on the survey page. If workers drop out after reading the consent form (which they can do without penalty) they will contribute to (measurable) attrition rather than (unmeasurable) selection biases. Thus concerns about self-selection are reduced to i) self-selection into AMT membership, ii) interest in research

studies in general, and iii) willingness to accept a particular level of compensation in return for completing a study (which has not yet shown to be relevant: Buhrmester et al., 2011; Marge et al., 2010; Mason & Watts, 2009). Although these are issues for researchers who care about representative samples, they are at least issues reflected in all studies conducted using AMT uniformly rather than differently depending on the source of the convenience sample.

4.2. Some Overlooked Advantages of Using AMT

Although the ease of finding and paying workers is important, AMT has a number of additional technical advantages that stem from the ability to track individual workers by their Worker ID (unique to individual workers and linked to a credit card). This allows researchers to avoid duplicate respondents within a single study without assuming that participants will not change IP addresses between responses (easily done using a proxy server or resetting their router) to participate multiple times. Worker IDs are of especial importance when trying to exclude workers from an ongoing program of related studies because in this case, simply filtering out responses from the same IP address will miss many duplicate responses from respondents who inadvertently submit different IP addresses at different times (e.g., because their ISP assigns dynamic IP addresses or they use a public WiFi connection).

Finally, although AMT is usually used to collect conventional survey or experimental data, as a platform it offers a number of features that make it possible to run less conventional experiments with ease. For example, the large number of workers makes it easy to recruit participants into real-time experiments on group dynamics (e.g., Suri & Watts, 2011), which is logistically difficult in traditional subject pools. Likewise, the bonus payment system makes it easy to implement designs with real monetary incentives, e.g., game-theoretic designs. Additionally, the unique identification number assigned to

each worker makes it possible to prescreen participants and maintain a pool of workers who fulfill specific criteria or to recontact individual workers participating in longitudinal designs.

5. Lingering Concerns about AMT

Although AMT has a number of appealing features, there are also concerns that are worth considering and as of yet have been overlooked. Concerns noted by researchers in the SPSP and SJDM mailing lists tended to be more heterogeneous than strengths. The most prevalent concerns were that the data provided were low quality (19%) and that participants lacked attention, motivation or effort (25%). Experimenter perceptions of data quality contradict the direct empirical comparisons of AMT and other samples conducted by me and other researchers (Buhrmester et al., 2011; Horton et al., 2011). It is difficult to interpret subjective concerns about data quality without knowing more about what counts as quality data, but concerns about attention and motivation are a tractable problem. Respondents seemed largely unaware of the possibility that workers may participate in conceptually related experiments (mentioned by 4% of respondents) or the possibility that workers share information (mentioned by nobody). However, even a brief review of worker-organized discussion boards suggests that these may be problems.

Below we discuss several different concerns that are often overlooked and present evidence to support whether they are a problem or not. At times it was necessary to collect additional survey data to address some unanswered questions about the habits of AMT workers. These data were collected from a sample ($N = 300$) typical of those recruited for empirical research: Americans who have successfully completed > 95% of all previously submitted tasks. This sample was supplemented with an additional 20 workers who were known to be among the most productive 1% of workers in a sample of previously

completed research HITs (see below). Unless stated otherwise, all claims reported in this section about the AMT population in general that are based on these data use only the original sample of 300 workers, while all findings that compare workers by productivity level include the supplementary sample of especially productive workers.

4.1. Worker attention

Although experimenters can exclude poor-quality workers (as measured by low approval rates for submitted HITs), even high-quality workers are motivated to complete HITs efficiently and may thus miss important instructions or carelessly respond to subjective measures. Like any Internet survey, the environments in which workers complete HITs are heterogeneous and there may be distractions that may lead them to miss experimental manipulations. Our survey revealed that although most workers completed the HIT from home (86%) and alone (73%), they are often doing other activities simultaneously: 18% of them reported watching TV, 14% of them reported listening to music and 6% of them were also instant messaging with at least one other person. If anything, these estimates are too conservative, as workers may be motivated to under report behaviors that call the quality of their data into question (Behrend et al., 2011)

Many techniques have been suggested to measure or increase respondent attention in traditional Internet surveys (see Couper, 2008 for a discussion). Passive techniques can be used to identify problematic workers: catch trials can be included to identify workers who agree with unlikely or even impossible statements, as in our comparative study, and scale responses can be examined for excessive use of the same response category (Johnson, 2005). Participants can also be actively reminded that researchers expect them to attend to questions using “instructional manipulation checks” (see Oppenheimer, Meyvis, & Davidenko, 2009) – although the success of this technique relies on its unfamiliarity among participants and may confuse workers by appearing deceptive.

There are also a number of additional techniques to improve response quality that are unique to AMT. Within AMT, money is only earned for correctly completed tasks. As a result, factual questions may focus workers' attention and lead them to answer with greater care. Supporting this, one study recruited workers and experts to evaluate the quality of Wikipedia pages (Kittur et al., 2008). Worker ratings and expert ratings were uncorrelated, except when workers were also required to include answers to objectively verifiable questions. Moreover, attention can also be directly incentivized in AMT by dividing payment between money for simply completing a task and a bonus payment for correctly answering factual questions about the task they completed.

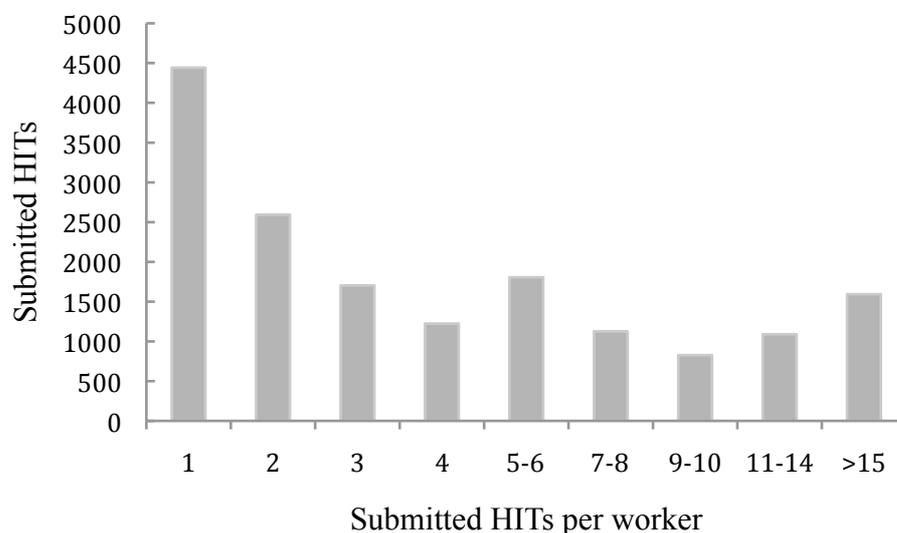
4.2. Participation in conceptually related experiments

Although AMT by default prevents the same worker from completing a single HIT multiple times, some care is required if one wants to prevent workers from being included in conceptually or methodologically similar studies. While undergraduate subject pools are continually replenished with naïve participants, AMT is not. Duplicate respondents are of concern not only because they violate assumptions of statistical independence but also because prior knowledge about the purpose of an experiment, familiarity with an experimental manipulation, or reason to suspect deception influence participant responses (Brock & Beckker, 1966; Edlund et al., 2009; Glinski, Glinski, & Slatin, 1970).

Pooling data from several collaborators and me resulted in a sample of 16,408 HITs distributed across 132 batches. Within this sample we found substantial reason to be concerned about duplicate responses. These HITs had been completed by a total of 7498 workers. The average worker completed 2.24 HITs ($SD = 3.19$), but a very small minority of workers was responsible for submitting most of the HITs. The most prolific 1% of workers from this sample was responsible for completing 11% of the submitted HITs (the highly productive workers we described earlier were recruited from this group), and the

rest of most prolific 10% were responsible for completing 41% of the submitted HITs (see Figure 1). A similar distribution was observed by Berinsky and colleagues (2011) across six experiments conducted by the authors over a period of several months –in their sample 24% of the workers participated in two or more experiments and 1% completed five or more of these experiments. These figures also corroborate earlier worker self-reports that suggested that AMT has a small population of very active workers (Ipeirotis, 2010).

Figure 1. Number of HITs completed by workers of different levels of productivity.



We looked at how frequently workers in the survey submitted experimental HITs and compare frequent versus infrequent submitters. There is nothing particularly special about the demographics of more productive workers although they tend to be somewhat older, more educated and more likely to be White than the general worker population. They also tend to be somewhat more focused than the general pool of workers: when completing the demographic survey they were more likely to be alone and less likely to be engaged in other tasks like listening to music, watching TV, or chatting online (see Table 3), suggesting that they may be particularly suitable for experiments that are sensitive to participant attention (e.g., those that rely on reaction time).

Table 3. Distractedness and Involvement Among Mechanical Turk Workers

	Overall	No productivity information	0-89 th percentile	90-98 th percentile	99 th percentile ^a	M-H χ^2 ^b
With other people	27%	32%	20%	23%	15%	4.95*
Listening to Music	14%	18%	10%	10%	0%	8.85**
Watching TV	18%	24%	12%	14%	15%	3.53†
Chatting online	6%	9%	3%	0%	3%	5.45*
Read Mturk blogs	28%	26%	26%	36%	40%	3.64†
Follow Requesters	55%	43%	68%	71%	72%	19.55***
Follow Academic Requesters	33%	27%	39%	32%	48%	4.93*

Note. Percentages are the proportion of respondents who affirmed that they engaged in this particular behavior. Productivity percentiles were assigned to workers based on the number of HITs completed in a 132 previous samples by 7498 workers).

^aIncludes high productivity workers who completed the initial questionnaire ($N = 13$) and a targeted supplemental sample ($N = 20$) recruited immediately after collection of the initial sample.

^bChi-Square and significance tests for Mantel-Haenszel linear-by-linear association test.

† $p < .06$, * $p < .05$, ** $p < .01$, *** $p < .001$

The most productive workers in the sample of previously completed HITs who also responded to the survey did not report spending more time using AMT than less productive workers during the previous week. However, they were still unusually likely to find and complete the survey. The most prolific 1% of workers comprised 4% of the sample and remainder of the most prolific 10% comprised a further 20% of the sample, significantly more than expected by chance, $\chi^2(1, 300) = 290.69, p < .001$. Further, it is important to note that this survey was posted using a new RequesterID and was only available for 3 days, so reputation effects of the authors do not explain the overrepresentation of productive workers. One possible explanation is that this survey had the keywords “survey”,

“research”, and “experiment” associated with it (as recommended to increase response rates; Chilton et al., 2009) and that the productive workers in our sample were those who seek out social science research HITs.

In short, these findings suggest that there is a small group of professional survey takers who operate on AMT. They complete a disproportionate amount of work. The presence of these workers is something of a mixed blessing. On one hand, they seem less distracted than the average worker. On the other hand, it is likely that they have participated in hundred of psychological experiments already.

In our survey, a substantial proportion of workers reported participating in some of the more common and easily describable experimental paradigms, such as the prisoner’s dilemma or the “trolley problems” commonly used to illustrate moral reasoning. As would be expected, the most productive workers are also the ones who are most likely to have participated in common experimental paradigms (see Table 4). Based on these findings, it seems that, without taking steps to filter or measure non-naïve participants, AMT may not be appropriate for commonly used paradigms. For less used paradigms, researchers can minimize this problem by sharing lists of who has completed specific experimental manipulations and excluding them from future experiments.

For established researchers, the problem of repeated workers across several experiments probably increases: more than half of all workers (55%) reported having a list of favorite requesters that they monitor for available HITs, and 58% of those who followed favorite requesters (about a third of the entire sample) reported that this list included academic researchers. The most productive workers are also especially likely to read blogs about AMT and follow specific requesters (see Table 4) and in fact websites exist that allow workers to track individual requesters and receive notification whenever they have work available. Thus, researchers should be aware that they may have a loyal following

who have completed their experiments in the past, read debriefing materials, and deliberately seek out their future experiments.

Table 4. Previous Exposure to Common Experimental Paradigms

	Overall	No productivity information	0-90 th percentile	90-98 th percentile	99 th percentile ^a	M-H χ^2 ^b
Prisoner's Dilemma	56%	36%	71%	85%	88%	68.71***
Ultimatum Game	52%	32%	65%	78%	94%	69.12***
Dictator Game	0%	22%	51%	64%	76%	64.79***
Trolley Problem	30%	10%	33%	68%	85%	107.95***
p-Beauty contest	7%	5%	10%	10%	9%	6.68**

Note. Percentages are the proportion of respondents who affirmed that they engaged in this particular behavior. Productivity percentiles were assigned to workers based on the number of HITs completed in a 132 previous samples by 7498 workers).

^aIncludes high productivity workers who completed the initial questionnaire ($N = 13$) and a targeted supplemental sample ($N = 20$) recruited immediately after collection of the initial sample.

^bChi-Square and significance tests for Mantel-Haenszel linear-by-linear association test.

** $p < .01$, *** $p < .001$

4.3. Workers may communicate with each other about experiments

A third potential problem is worker crosstalk. AMT workers maintain online forums where they share information and opinions about AMT and links to particularly interesting or lucrative HITs, which could potentially lead to foreknowledge in experimental participants. Empirical research on college undergraduate populations has demonstrated that participants do share information with each other, at least when sufficiently motivated (e.g., when incentives are offered for a correct response; Edlund, et al., 2009).

In the survey, 26% of participants reported knowing someone else who used AMT personally, and 28% reported reading forums and blogs about AMT. However, when asked to rank the frequency with which they discuss or read about various aspects of AMT, the actual purpose or contents of AMT HITs is far less important than pragmatic considerations such as the amount requesters pay or their reputations (see Table 5) suggesting that cross-talk may not be an issue unless the information involved could increase financial reward. Only half of the respondents who actually read blogs (about 13% of the overall population) recalled ever seeing a discussion about the contents of a social science research study online. However, workers do on occasion discuss experiments on discussion boards (accompanied by links to the HIT) and they have been known to inadvertently share details that are a part of the experimental manipulation. Thus, researchers should probably ask workers how they found the HIT at the end of their survey and monitor discussion boards that refer a lot of respondents.

In sum, participant attentiveness is a problem common across electronically recruited populations; however, there are unique options for increasing attentiveness on AMT. A small group of workers contribute a surprisingly large amount of the data, and researchers may want to monitor or exclude them from participation. It goes beyond the goal of the present essay to discuss techniques that can be used to screen out participants on AMT, but the “Qualifications” feature allows doing so in a quite effective manner. Worker cross-talk seems to be a relatively minor problem, as few workers seem to remember discussions about the contents of surveys or experiments, but should still be monitored.

6. Concluding Remarks

In this essay 23 reviewed the effectiveness of crowdsourcing websites to augment traditional empirical research, with specific reference to AMT. There are many practical and some theoretical reasons to use crowdsourcing websites to collect data that have been identified by our work and others'. However, while AMT makes recruiting and running online studies easy, it comes with its own challenges and drawbacks. For instance, workers on MTurk can and sometimes will search specifically for opportunities related to scientific research, ultimately becoming followers of specific requesters and being familiar with paradigmatic experiments such as bargaining games and moral dilemmas. Scientists should consider and report how subjects are screened in light of these tendencies. Moreover, a subject pool which is shared by the entire community must be handled with particular care: While aggressively rejecting work may ensure attentiveness, and deception may increase the external validity for individual experiments, they may also increase demand characteristics and suspicion for future studies. Analogously, experimenters who are slow to pay workers or unresponsive to communications might jeopardize the reputation of the research community. In short, while conducting experiments using AMT workers, researchers may find themselves participants in a larger public goods experiment.

References

- Baron, J. & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54, 569–579.
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, 43(3), 800-813.
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2011). Using Mechanical Turk as a Subject Recruitment Tool for Experimental Research. Unpublished manuscript.
- Brock, T. C., Becker, L. A. (1966). 'Debriefing' and susceptibility to subsequent experimental manipulations, *Journal of Experimental Social Psychology*, 2, 3-5.

- Buhrmester, M. D., Kwang, T., Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.
- Chandler, D. & Kapelner, A. (2010). Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets. *University of Chicago mimeo*.
- Charness, G., Karni, E., & Levin, D. (2009). On the conjunction fallacy in probability judgment: New experimental evidence regarding Linda. *Games and Economic Behavior*, 68, 551–556.
- Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2009). Task search in a human computation market. In Proceedings of the ACM SIGKDD workshop on human computation (1-9). In P. Bennett, R. Chandrasekar, M. Chickering, P. Ipeirotis, E. Law, A. Mityagin, F. Provost & L. von Ahn (Eds.) *HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation* (77–85). New York: ACM.
- Couper, M. (2008). Designing effective web surveys. New York; Cambridge University Press.
- Edlund, J. E., Sagarin, B. J., Skowronski, J. J., Johnson, S. J., Kutter, J. (2009). Whatever happens in the laboratory stays in the laboratory: The prevalence and prevention of participant crosstalk. *Personality and Social Psychology Bulletin*, 35, 635–642.
- Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, 5, 159-163.
- Fagerlin, A., Zikmund-Fisher, B., Ubel, P., Jankovic, A., Derry, H., & Smith, D. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, 27, 672–680.
- Gaggioli & Riva (2008). Working the Crowd. *Science*, 12, 1443.
- Glinski, R.J., Glinski, B.C. and Slatin, G.T., (1970). Nonnaivety contamination in conformity experiments: sources, effects, and implications for control. *Journal of Personality and Social Psychology*, 16, 478–485.
- Gosling, S., Vazire, S., Srivastava, S., & John, O. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, 59, 93–104.
- Heinrich, J., Heine, S., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29.
- Horton, J.J., Rand, D.G., & Zeckhauser, R.J. (2011). The Online Laboratory: Conducting Experiments in a Real Labor Market. *Experimental Economics*, 4, 399-42
- Horton, J., & Chilton, L. (2010). The labor economics of paid crowdsourcing. *Proceedings of the 11th ACM Conference on Electronic Commerce*.
- Johnson, J. A. (2005). Ascertaining the validity of Web-based personality inventories. *Journal of Research in Personality*, 39, 103-129
- Ipeirotis, P. (2010). Demographics of Mechanical Turk. *CeDER-10-01 working paper*, New York University.
- Lipkus, I., Samsa, G., & Rimer, B. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21, 37–44.

- Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the ACM conference on human factors in computing systems* (pp. 453–456). New York: ACM.
- Marge, M.; Banerjee, S.; Rudnicky, A.I. (2010). Using the Amazon Mechanical Turk for transcription of spoken language. *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (5270-5273). Washington: Institute of Electronics and Electrical Engineers.
- Mason, W., & Watts, D. (2009). Financial incentives and the “performance of crowds. In P. Bennett, R. Chandrasekar, M. Chickering, P. Ipeirotis, E. Law, A. Mityagin, F. Provost & L. von Ahn (Eds.) *HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation* (77–85). New York: ACM.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872.
- Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 2863–2872, New York, NY, USA. ACM.
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672-695.
- Suri, S., & Watts, D. J. (2011). A Study of Cooperation and Contagion in Web-Based, Networked Public Goods Experiments, *PLoS ONE 6*(3).
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.
- U.S. Census, http://pubdb3.census.gov/macro/032007/hhinc/new06_000.htm and ComScore Statistics, June 2008
- von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science, 321*, 1465-1468.

Estratto per riassunto della tesi di dottorato

Studente: Gabriele Paolacci
Matricola: 955412
Dottorato: Economia Aziendale
Ciclo: 23
Titolo della tesi: Four Essays in Consumer Decision-Making

Estratto:

La tesi si compone di quattro capitoli. **1. Inherent, constructed, revealed preferences: Guidelines for marketers.** Dopo una rassegna della letteratura, suggerisco modi concreti in cui i manager possono misurare le preferenze innate e sfruttare gli effetti di contesto. **2. The Intermediate Alternative Effect: Considering a small tradeoff increases subsequent willingness to make large tradeoffs.** Dimostriamo che l'“effetto dotazione” può essere ridotto offrendo prima ai consumatori un'opzione che rappresenti un passaggio intermedio dallo status quo all'opzione alternativa. **3. Same world, different perceptions: Systems of measurement affect judgments.** Dimostriamo in due studi sperimentali che cittadini di paesi diversi percepiscono diversamente informazioni equivalenti perché le considerano con diversi sistemi di misurazione. **4. Online experimentation: Amazon Mechanical Turk.** Validiamo Mechanical Turk come subject pool e ne discutiamo in modo esteso vantaggi e rischi.

Abstract:

The dissertation is composed of four essays. **1. Inherent, constructed, revealed Preferences: Guidelines for marketers.** After reviewing the debate on consumer preferences, I suggest concrete steps that marketers can use to measure inherent preferences and exploit context effects. **2. The Intermediate Alternative Effect: Considering a small tradeoff increases subsequent willingness to make large tradeoffs.** We propose and show that the “endowment effect” can be reduced by first providing consumers with the choice between keeping their endowment and trading it of an “intermediate” option. **3. Same world, different perceptions: Systems of measurement affect judgments.** We propose and show that citizens of different countries perceive differently semantically equivalent information because they use different systems of measurement. **4. Online experimentation: Amazon Mechanical Turk.** We validate Mechanical Turk as a subject pool and discuss extensively its advantages and causes of concern.

Firma dello studente

