8

Arbitrarily Coherent Preferences

DAN ARIELY, GEORGE LOEWENSTEIN, AND DRAZEN PRELEC

Economic theories of valuation generally assume that prices of commodities and assets are derived from underlying "fundamental" values. For example, in finance theory, asset prices are believed to reflect the market estimate of the discounted present value of the asset's payoff stream. In labor theory, the supply of labor is established by the tradeoff between the desire for consumption and the displeasure of work. Finally, and most importantly for this work, the economic account of consumer behavior assumes that the demand curves for consumer products—chocolates, CDs, movies, vacations, drugs, etc.—can be ultimately traced to the valuation of pleasures that consumers anticipate receiving from these products.

Because it is difficult, as a rule, to measure fundamental values directly, empirical tests of economic theory typically examine whether the effects of changes in circumstances on valuations are consistent with theoretical prediction—for example, whether labor supply responds appropriately to a change in the wage rate, whether (compensated) demand curves for commodities are downward sloping, or whether stock prices respond in the predicted way to share repurchases. It has often been noted, however, that such "comparative static" relationships are a necessary but not a sufficient condition for fundamental valuation (e.g. Summers 1986). Gary Becker (1962) was perhaps the first to make this point explicitly, when he observed that consumers choosing commodity bundles randomly from their budget set would, nevertheless, produce downward sloping demand curves.

In spite of this ambiguity in the interpretation of demand curves, the intuition that prices must, in some way, derive from fundamental values is still strongly entrenched. Psychological evidence that preferences can be manipulated by normatively irrelevant factors, such as option "framing," changes in the "choice context," or the presence of prior cues or "anchors," is often rationalized by appealing to consumers' lack of information about the options at stake and the weak incentives operating in the experimental setting. From the standpoint of

This work is heavily based on Ariely, Loewenstein, and Prelec. "Coherent arbitrariness: Stable demand curves without stable preferences." *Quarterly Journal of Economics*, 118(1): February 2003.

economic theory, it is easy to admit that consumers might not be very good at predicting the pleasures and pains produced by a purchase, especially if the purchase option is complex and the choice hypothetical. It is harder to accept that consumers might have difficulty establishing how much they value each individual bit of pleasure or pain in a situation where they can experience the full extent of this pleasure or pain just before the pricing decision.

In this chapter we provide evidence that consumers' absolute valuations of simple pains are surprisingly arbitrary. We will try to illustrate that consumers' pricing of a simple hedonic stimulus (an unpleasant sound played over headphones) is not consistent with fundamental valuation in that it is powerfully influenced by non-informative anchor values. However, we also show that when subjects are asked to price sounds of different durations, the relative valuations of noises of different durations are coherent—as if supported by demand curves derived from fundamental preference. Valuations, therefore, display a peculiar combination of arbitrariness and coherence that we refer to as "coherent arbitrariness." Our experiments show how comparative static tests could support predictions that follow from fundamental valuation even when values incorporate a substantial arbitrary component.

At a theoretical level we show that these findings are consistent with an account of preferences that posits that preferences are initially malleable but become "imprinted" (i.e. precisely defined and largely invariant), after the individual is called upon to make an initial decision. Prior to imprinting, preferences are "arbitrary," meaning that they are highly responsive to both normative and non-normative influences. Following imprinting, preferences become "coherent," meaning that they are more precisely defined and largely fixed in subsequent decisions. The model predicts that consumers will respond to changes in conditions in a coherent fashion, as if supported by demand curves derived from fundamental preference, even when their initial valuations are arbitrary.

The material that follows is divided into three sections. Section 1 reviews the relevant psychological literature. Section 2 presents the imprinting model. Section 3 presents the results of a set of lab experiments. Section 4 concludes with a summary and a general discussion of economic implications.

1. RELEVANT PSYCHOLOGICAL RESEARCH

There is a large body of research showing that estimates of unknown quantities can be influenced by exposure to normatively irrelevant information. In a famous early study, Tversky and Kahneman (1974) spun a wheel of fortune with numbers that ranged from 0 to 100, asked subjects whether the fraction of African nations in the United Nations was greater than or less than that number, and then instructed subjects to estimate the actual figure. Estimates were significantly related to the number spun on the wheel (the anchor), even though subjects could clearly see that the number had been generated by a purely chance process. The authors explained this effect in terms of an "anchoring and

adjustment" heuristic according to which subjects began with the number shown on the wheel and then insufficiently adjusted away from that number, based on their own knowledge of the relevant domain.

The large literature on anchoring effects includes several studies that demonstrate the impact of the anchor on valuations (for recent studies of anchoring, see, for example, Jacowitz and Kahneman 1995; Strack and Mussweiler 1997; Chapman and Johnson 1999; Epley and Gilovitch 2001). For example, Johnson and Schkade (1989) have shown that prior anchoring questions (i.e. whether the subject's certainty equivalent is above or below the anchor) affects certainty equivalents for gambles. Higher anchor values lead subsequently to higher stated certainty equivalents. Green et al. (1998), and Kahneman and Knetsch (1993), found the same effect with judgments of willingness to pay (WTP) for public goods; higher values in the initial Yes/No question led to higher subsequent WTP.

In a related experiment we (Ariely, Loewenstein, and Prelec 2003) presented a class of 55 Sloan MBA students with six products (computer accessories, wine bottles, luxury chocolates, and books), which were briefly described without mentioning market price. The average retail price of the items was about $70. After introducing the products, students were asked whether they would buy each good for a dollar figure equal to the last two digits of their social security number. After this Accept/Reject response, they stated their dollar maximum WTP for the product. A random device determined whether the product would in fact be sold on the basis of the first, Accept/Reject response, or the second, WTP response (via the incentive-compatible Becker–Degroot–Marschak procedure, 1964). Subjects understood that both their Accept/Reject response and their WTP response had some chance of being decisive for the purchase, and that they were eligible to purchase at most one product.

In spite of the realism of the products and transaction, the impact of the social security number on stated WTP was significant in every product category. Subjects with above-median social security numbers stated values from 57 to 107 percent greater than did subjects with below-median numbers. The effect is even more striking when examining the valuations by quintiles of the social security number distribution, as shown in Table 8.1. The valuations of the top quintile subjects were typically greater by a factor of three. For example, subjects with social security numbers in the top quintile were willing to pay $56 on average for a cordless computer keyboard, compared with $16 on average for subjects with bottom quintile numbers. Evidently, these subjects did not have, or were unable to retrieve personal values for, ordinary products.

Alongside this volatility of absolute valuation we also observed a marked stability of relative preference. For example, the vast majority of subjects (>95 percent) valued a cordless keyboard more than a trackball, and the highly rated wine more than the lower-rated wine. Subjects, it seems, did not know how much they valued these items, but they did know the relative ordering within the categories of wine and computer accessories.

Table 8.1. *Average stated WTP sorted by quintile of the sample's social security number distribution*

Quintile of SS# distribution	Cordless trackball	Cordless keyboard	Average wine	Rare wine	Design book	Belgian chocolates
1	$8.64	$16.09	$8.64	$11.73	$12.82	$9.55
2	$11.82	$26.82	$14.45	$22.45	$16.18	$10.64
3	$13.45	$29.27	$12.55	$18.09	$15.82	$12.45
4	$21.18	$34.55	$15.45	$24.55	$19.27	$13.27
5	$26.18	$55.64	$27.91	$37.55	$30.00	$20.64
Correlations	0.415	0.516	0.328	0.328	0.319	0.419
	$p=0.0015$	$p<0.0001$	$p=0.014$	$p=0.0153$	$p=0.0172$	$p=0.0013$

Note: The last row indicates the correlations between social security numbers and WTP (and their significance levels).

In another study we offered a group of seventy-seven business professionals, who came to Massachusetts Institute of Technology (MIT) for a day of executive education, to participate in auctions for Godiva chocolates. We conducted three auctions for chocolate boxes, one auction for each box size: Small (4 oz), medium (0.5 lbs), and large (1 lbs). Before participating in each auction, the executives were asked to write down the last two digits of their social security number and then to indicate whether they would pay this sum for the chocolate box that was offered. After answering this anchoring question, the executives were asked to submit a bid for each of the chocolate boxes (the pricing procedure in this study was a standard second price, sealed-bid auction, where the highest bidder gets the chocolate for the price set by the second highest bidder). The results showed the non-surprising pattern that all participants offered higher prices for larger quantities of chocolate. More important for our purpose, the correlations between the last two digits of social security numbers and the bids were 0.29 for large, 0.27 for the medium, and 0.27 for the small (all significant at the 0.01 level). At the end of the auctions all participants were asked whether they thought that the anchors influenced their prices. Fifty-six (72 percent) of the respondents indicated that they were not influenced by the anchors. An analysis of the anchoring effect as a function of reported anchors' influence revealed that while the participants who indicated that the anchor influenced them placed higher bids, the relative influence of the anchors on their bids was not different. Moreover, when we asked the participants to provide a new "clean" estimate (trying to partial out the effect of the anchor) they were unable to do so and the anchoring effect remained unchanged.

Two others research literatures, on the "compromise effect" and on "evaluability," are also relevant to the question of fundamental valuations. The "compromise effect" refers to the finding that people have an unreasonable preference for the intermediate alternative across a series of three-alternative choices (Simonson and Tversky 1992). When the choice set consists of A, B, and C, ordered, say, by price and quality, a majority of consumers will typically choose B, but when it consists of A+, A, and B, a majority opt for A. Such a pattern suggests that consumers do not have a stable tradeoff between price and quality attributes but they do know that they have diminishing marginal utility for one or both attributes, producing a desire to compromise. As Drolet, Simonson, and Tversky (2000) express it, it is as if consumers have convex indifference curves, but those curves "travel with the choice set" (Drolet, Simonson, and Tversky 2000). The research on the compromise effect suggests that individuals will act as if they have sensible (in this case convex) indifference curves, even when they are uncertain about other important properties of those curves, such as their slope or positioning.

"Evaluability" refers to preference reversals that arise when options are evaluated jointly (within subject) and separately (between subject). Chris Hsee et al. (1999) explains these reversals by assuming that it is more difficult to evaluate attributes separately than jointly, and that the difficulty with absolute evaluations is larger for attributes that do not have well-established standards. For example, subjects in one study were asked to assess two political candidates, one who would bring 1000 jobs to the district and the other who would bring 5000 jobs to the district but had a DUI conviction. When the candidates were evaluated separately, the first candidate was judged more favorably, presumably because the employment figure was hard to evaluate. However, when the candidates were compared side-by-side people indicated that the employment difference more than compensated for the DUI conviction, and gave their preference to the second candidate. The research on evaluability suggests that people respond in a sensible fashion to differences they are made aware of, even when their baseline judgments are partly arbitrary. This explains why, for example, the individuals in the auction study sensibly priced the more expensive wine higher than the less expensive wine, even when the strengths of the anchoring effects revealed that their prices for both items were largely arbitrary.

2. THE THEORY: COHERENT ARBITRARINESS

The work we have just described suggests that people might exhibit stable, coherent, relative preferences but unstable or arbitrary absolute (or fundamental) preferences. In this section we will indicate how standard utility theory, which assumes fundamental valuation, could be modified to accommodate this pattern. We use the formal framework of an incomplete preference order, which means that there exist pairs x and y such that neither option is (weakly) preferred to the other one. We describe how this formal structure can be specifically adapted to express our idea of stable relative and undefined absolute preference.

Consider two multi-attribute or multi-commodity alternatives, $x=(x_1,\ldots,x_n)$ and $y=(y_1,\ldots,y_n)$. In the standard microeconomic model, preferences are complete (either x is [weakly] preferred to y, or y to x, or both) and there exists a utility function $u(x)$ such that x is strictly preferred to y, or $x \geq y$, if and only if the utility of x is strictly greater than the utility of y, $u(x) > u(y)$. For the purpose of the illustration, we focus now on the special case where the utility function is additive: $u(x) = \sum_i u_i(x_i)$, and imagine a situation where the contribution of each attribute to total utility is well defined up to an arbitrary multiplicative constant (α_i), one for each attribute (in the non-additive formulation, we would let the utility function depend on a vector of α-parameters, $u(x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n)$). Before the moment of choice, these constants (or weights) are not fixed but are restricted to a set, A, of acceptable vectors of values. Faced with a choice, a person will prefer x to y if and only if for all possible α in A, the utility of x is greater than the utility of y:[1]

$$x \geq_A y \text{ iff } \underset{\alpha \in A}{Min} \sum_i (\alpha_i u_i(x_i) - \alpha_i u_i(y_i)) > 0.$$

The preference relation, $\geq_A$, is subscripted to highlight the fact that it is conditional on the set A. Of course, for many pairs x, y, neither alternative will be strictly preferred according to this definition. In that case, preference is undefined. However, a person will still make a choice. We do not say anything within the model about how the choice in such a case might be made. We only assume that if the situation demands a choice, the person will, in fact, choose. This "foundational" choice then becomes a part of that person's stock of decisional precedents, ready to be invoked the next time a similar choice situation arises.

To take a generic case, suppose that an identical choice presents itself at some later point in time. We predict that the foundational choice will simply be repeated, because from the consumer's perspective this particular "choice problem" has been solved and its solution is known. Even if the new choice differs in some respects, however, the consumer will use whatever available precedents exist (Gilboa and Schmeidler 1995). The impact of a precedent is formally expressed as a restriction on the set of possible attribute weights, A. After the first choice, the set A is reduced to those values that are consistent with preferences "revealed" by that initial choice.

Specifically, we assume that the person follows a choice process as described in the next three steps:

Step 1. If $x \geq_A y$, choose x (and if $y \geq_A x$, choose y).
Step 2. If neither $x \geq_A y$ nor $y \geq_A x$, then choose x or y according to some unspecified criterion.
Step 3. If Step 2 yields a choice of x, then for the next choice, restrict A to the largest subset B of A such that $x \geq_B y$ (if Step 2 yields a choice of y, then restrict A to largest subset C such that $y \geq_C x$).

[1] This definition is closely related to Bewley's (1986) model of inertia in decision-making under uncertainty. In that model, a decision-maker holds multiple priors over states-of-the-world, and will only leave one risk position for another if it is better under all priors. Formally, the α-weights in our model correspond to subjective probabilities in Bewley's representation.

Figure 8.1 shows how this process works at the level of preferences over options defined on two attributes (measured on the horizontal and vertical axes of each panel). In Figure 8.1(a), we see that instead of the traditional non-overlapping indifference curves, there is a family of admissible indifference curves passing through an option, creating a "bow-tie" pattern. These two curves partition the space of possible options into three regions. Options that lie to the northeast of the bow-tie are preferred to x; options that lie to the southwest of the bow-tie are inferior to x (i.e. x is preferred to them); the remaining options, falling within the bow-tie do not have defined preferences with respect to x, such that there is some free play in the tradeoff between the first and the second attributes.

Figure 8.1(b) shows an option (y) that has an undefined preference with respect to (x). The tradeoff between attributes 1 and 2 is not specified with enough precision for a person to know whether option y is better than option x. In such a situation, following Step 2 in the decision process, the person will choose x or y according to some reason or influence that is arbitrary from the standpoint of his or her preferences. The choice could in reality have a large stochastic component, but if the influence of the stochastic component is not apparent, then the person

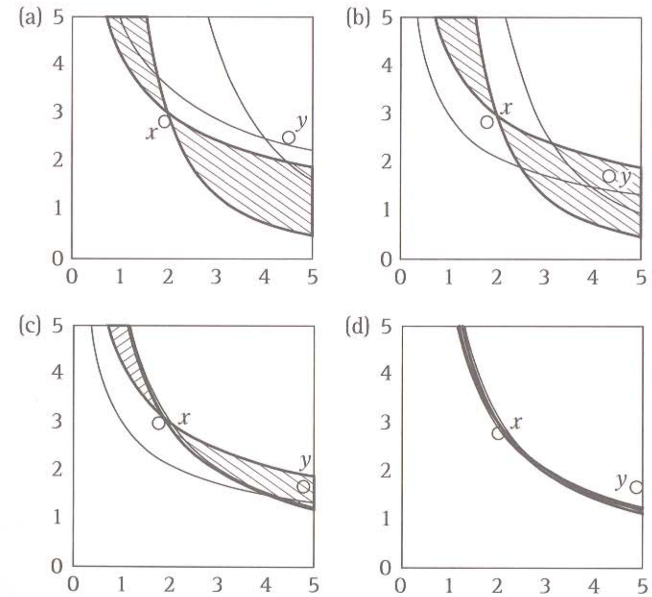

Figure 8.1. *Incomplete preferences, and how they change as a result of an arbitrary initial choice. (a) Option y is preferred to option x; (b) Neither option is preferred; (c) Impact on preferences in (b) of choosing x over y; (d) Impact on preferences in (b) of judging x to indifferent to y*

will subsequently treat the choice as an indicator of preference.[2] This is shown in Figure 8.1(c), where the preference order is "filled in" so as to accommodate the fact that the person has just acted as if x is preferred to y. The region of undefined preferences shrinks, as shown by the reduced size of the shaded region. The tradeoff between attributes is now better specified, and it favors attribute 1 (corresponding to the horizontal axis). Even allowing for some residual free-play in the tradeoff between attributes, option x is now unambiguously preferred. Figure 8.1(d) depicts the individual's preferences after expressing indifference between options x and y. Such a situation might arise if the individual were to make a series of choices sequentially or were asked to adjust one of the attributes of y to establish indifference between the two options. In this situation, as can be seen from Figure 8.1, preferences take the form of conventional (i.e. "thin") indifference curves.

One of the main psychological building blocks of this theory is that consumers do not arrive at a choice or a pricing task with an inventory of preexisting preferences and probability distributions. Instead, consistent with a great deal of other psychological evidence (Kahneman and Miller 1986; Payne, Bettman, and Johnson 1993; Drolet, Simonson, and Tversky 2000), consumers seem to have some range of acceptable values. If a give-or-take price for a product falls outside this range, then the purchase decision is straightforward: "Don't Buy" if the price is above the range, and "Buy" if the price is below the range. But, what if the stated price falls within the WTP range, so that the range does not determine the decision, one way or the other? We do not know much about how a choice in such a case might be made. We do know that if the situation demands a choice, then the person will in fact choose, that is, will either purchase or not purchase. We assume that this "foundational" choice then becomes a part of that person's stock of decisional precedents, ready to be invoked the next time a similar choice situation arises (Gilboa and Schmeidler 1995).

To relate this discussion to the experiment with the common products described above, suppose that a subject with a social security number ending with 25 has an a priori WTP range of \$5–30 for wine described as "average," and \$10–50 for the "rare" wine. Both wines, therefore, might or might not be purchased for the \$25 price. Suppose that the subject indicates, for whatever reason, that she would be willing to purchase the average bottle for \$25. If we were to ask her a moment later whether she would be willing to purchase the "rare" bottle for \$25, the answer would obviously be "yes," because from her perspective this particular "choice problem" has been solved and its solution is known: If an average wine is worth at least \$25, then a rare wine must be worth more than \$25! Moreover, when the subject is subsequently asked to provide WTP values for the wines, then that problem, too, is now substantially

[2] The assumption that people remember their choices but not the preferences which led to those choices, plays a role in several recent economic models of self-knowledge and self-control (Hirschleifer and Welch 1988; Koszegi 1999; Benabou and Tirole 2000; Bodner and Prelec, in press).

constrained: The prices will have to be ordered so that both prices are above \$25 and the rare wine is valued more.

There are many psychological details that we are not specifying. We do not say much about how the choice is made if the price falls within the range, nor do we propose a psychological mechanism for the anchoring effect itself. There are several psychological accounts of anchoring and for our purposes it is not necessary to decide among them (Epley and Gilovich 2001; Mussweiler and Strack 2001). The substantive claims we do make are the following: First, in situations in which valuations are not constrained by prior precedents, choices will be highly sensitive to normatively irrelevant influences and considerations such as anchoring. Second, because decisions at the earlier stages are used as inputs for future decisions, an initial choice will exert a normatively inappropriate influence over subsequent choices and values. Third, if we look at a series of choices by a single individual, they will exhibit an orderly pattern (coherence) with respect to numerical parameters like price, quantity, quality, and so on.

Behaviorally then, consumers in the marketplace may largely obey the axioms of revealed preference; indeed, according to this account, a person who remembered all previous choices and accepted the transitivity axiom would never violate transitivity. However, we cannot infer from this that these choices reveal true preferences. Transitivity may only reflect the fact that consumers *remember* earlier choices and make subsequent choices in a fashion that is consistent with them, not that these choices are generated from preexisting preferences.

3. THE EXPERIMENTS

In the experiments to be described next, we address the question of whether consumers do indeed enter the laboratory with a stable, preexisting valuation of pleasure and pain. In each experiment, subjects stated their WTA pains of different durations (induced by a loud noise played over headphones)—in exchange for payment. Subjects were initially exposed to a sample of the noise, and then asked whether—hypothetically—they would be willing to experience the same noise again in exchange for a payment of magnitude X (with X varied across subjects). Their actual WTAs were then elicited for different noise durations.

We used this artificial hedonic "product" for several reasons. First, we were able to provide subjects with a sample of the experience before they made subsequent decisions about whether to experience it again in exchange for payment. Second, because of the nature of the sensory experience with the sounds, subjects entered the pricing phase of the experiment with full information about the experience they were pricing. Third, we wanted to avoid a situation in which subjects could solve the pricing problem intellectually, without drawing on their own sensory experience. Annoying sounds have no clear market price, so our subjects could not refer to similar decisions made outside the laboratory as a basis for their valuations. Fourth, we wanted to make the money stakes in this decision comparable to the stakes in routine consumer expenditures. The plausible range of values for avoiding the annoying sounds in our experiments ranges from a few cents, to several dollars. Fifth, with annoying sounds it is possible to recreate the same hedonic experience repeatedly, permitting an experiment with repeated trials. Prior research shows that with annoying sounds, unlike many other types of pleasures and pains, there is little or no satiation or sensitization to repeated presentations of annoying sounds (Ariely and Zauberman 2000).

3.1. Experiment 1: Coherently Arbitrary Valuation of Pain

The goal of Experiment 1 was to test (*a*) whether valuation of annoying sounds was susceptible to an anchoring manipulation; (*b*) whether additional experience with the sounds would erode the influence of the anchor; and (*c*) whether valuation would be sensitive to a within-subject manipulation of the duration of the annoying sound, thus demonstrating coherence with respect to this attribute.

One hundred and thirty-two students from MIT participated in the experiment. Approximately half were undergraduates and the rest were MBA students or, in a few cases, recruiters from large investment banks. Subjects were randomly assigned to six experimental conditions. The experiment lasted about 25 min and subjects were paid according to their performance as described below.

At the beginning of the experiment, all subjects listened to an annoying 30-s sound, delivered through headphones. The sound was a high-pitched scream (a triangular wave with frequency of 3000 Hz), similar to the broadcasting warning signal. The main experimental manipulation was the anchor price, which was manipulated between-subject at three levels: An anchor price of 10¢ (low-anchor), an anchor price of 50¢ (high-anchor), and no anchor. Subjects in the low-anchor (high-anchor) condition first encountered a screen that read:

> In a few moments we are going to play you a new unpleasant tone over your headset. We are interested in how annoying you find it to be. Immediately after you hear the tone, we are going to ask you whether you would be willing to repeat the same experience in exchange for a payment of 10¢ [50¢].

Subjects in the no-anchor condition listened to the sound but were not given any external price and were not asked to answer any hypothetical question.

Before the main part of the experiment started, subjects were told that they would be asked to indicate the amount of payment they required to listen to sounds that differed in duration but were identical in quality and intensity to the one they had just heard. Subjects were further told that on each trial the computer would randomly pick a price from a given price distribution. If the computer's price was higher than their price, the subject would hear the sound and also receive a payment corresponding to the price that the computer had randomly drawn. If the computer's price was lower than their price, they would not hear the sound nor receive payment for that trial. Subjects were told that this procedure ensured that the best strategy is to pick the minimum price for which they would be willing to listen to the sound, not a few pennies more and not a few pennies less. The prices picked by the computer were drawn from a triangle distribution ranging from 5¢ to 100¢, with the lower numbers being more frequent than the higher numbers. The distribution was displayed on the screen for subjects to study and, importantly, the distribution was the same for all subjects.

After learning about the procedure, subjects engaged in a sequence of nine trials. On each trial, they were informed of the duration of the sound they were valuing (10, 30, or 60 s) and were asked to indicate their WTA for the sound. The three durations were presented either in an increasing (10, 30, 60 s) or decreasing order (60, 30, 10 s). In both cases, each ordered sequence repeated itself three times, one after the other. After each WTA entry, the computer asked subjects whether they were willing to experience the sound for their WTA minus 5¢, and whether they would experience it for their WTA plus 5¢. If subjects did not answer "no" to the first question and "yes" to the second, the computer drew their attention to the fact that their WTA was not consistent with their responses, and asked to them to reconsider their WTA price.

After finalizing a WTA value, subjects were shown their price along with the random price drawn from the distribution. If the price specified by the subject was higher than the computer's price, the subject did not receive any payment for that trial and continued directly to the next trial. If the price set by the subject was lower than the computer's price, the subject heard the sound over the headphones, was reminded that the payment for the trial would be given to them at the end of the experiment, and then continued to the next trial. At the end of the nine trials, all subjects were paid according to the payment rule.

3.1.1. *Results*

A set of simple effect comparisons revealed that average WTA in the high-anchor condition ($M = 59.60$) was significantly higher than average WTA in either the low-anchor condition ($M = 39.82$; $F(1, 126) = 19.25$, $p < 0.001$) or the no-anchor condition ($M = 43.87$; $F(1, 126) = 12.17$, $p < 0.001$). WTA in the low-anchor condition was not significantly different from WTA in the no-anchor condition ($p = 0.37$). Because subjects in the high-anchor condition specified higher WTAs, they naturally listened to fewer sounds ($M = 2.8$) than subjects in the low-anchor and no-anchor conditions (M's = 4.5 and 4.1; $F(1, 126) = 14.26$, $p < 0.001$). High-anchor subjects also earned significantly less money on average ($M = \$1.53$) than those in the no-anchor condition and the low-anchor condition, (M's = \$2.06, and \$2.16; $F(1, 126) = 7.99$, $p < 0.005$).

Although there was a significant drop in WTA values from the first to the second replication ($F(1, 252) = 17.54$, $p < 0.001$), there was no evidence of convergence of WTA among the different anchor conditions. Such convergence would have produced a significant interaction between the repetition factor and the anchoring manipulation, but this interaction was not significant. A variety of different tests of convergence produced similar results. First we carried out an ANOVA analysis in which we took only the first and last trial as the repeated measure dependent variable. Again, the interaction between trial (first versus last) and the anchoring manipulation was nonsignificant. We also estimated the linear trend of WTA over time for each subject. The estimated trends were decreasing, but the rate of decline did not differ significantly between the two anchoring conditions.

WTA values were highly sensitive to duration in the expected direction ($F(2,252) = 294.46$, $p < 0.001$) (for more discussion of sensitivity to duration see Kahneman, Wakker, and Sarin 1997; Ariely and Loewenstein, 2000). The mean price for the 10-s sound ($M = 28.35$) was significantly lower than the mean price for the 30-s sound ($M = 48.69$; $F(1,252) = 169.46$, $p < 0.001$), and the mean price for the 30-s sound was lower than the mean price for the 60-s sound ($M = 66.25$; $F(1,252) = 126.06$, $p < 0.001$).

Figure 8.2 provides a graphical illustration of the results thus far. First, the vertical displacement between the lines shows the powerful effect of the anchoring manipulation. Second, despite the arbitrariness revealed by the effect of the anchoring manipulation, there is a strong and almost linear relationship between WTA and duration. Finally, there is no evidence of convergence between the different conditions across the nine trials.

Figure 8.3 provides additional support for the tight connection between WTA and duration. For each subject, we calculated the ratio of WTA in each of

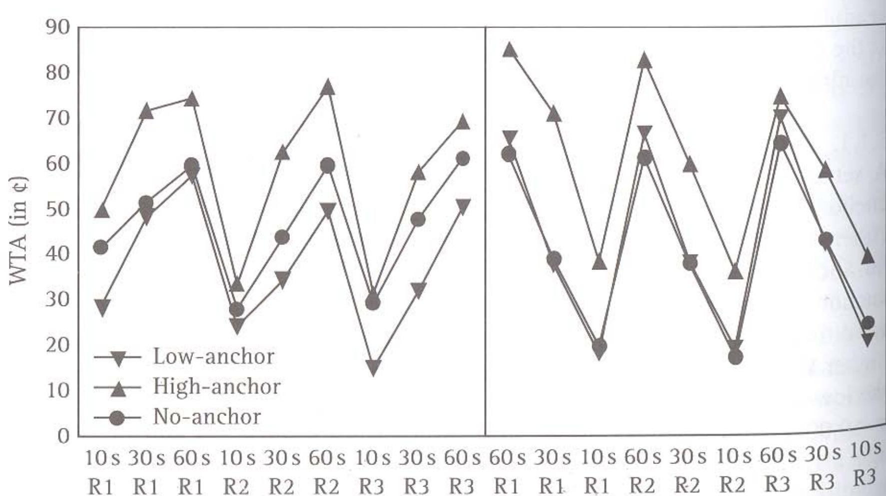

Figure 8.2. *Mean WTA for the nine trials in the three anchor conditions*

Notes: The panel on the left shows the increasing condition (duration order of 10, 30, and 60 s). The panel on the right shows the decreasing condition (duration order of 60, 30, and 10 s).

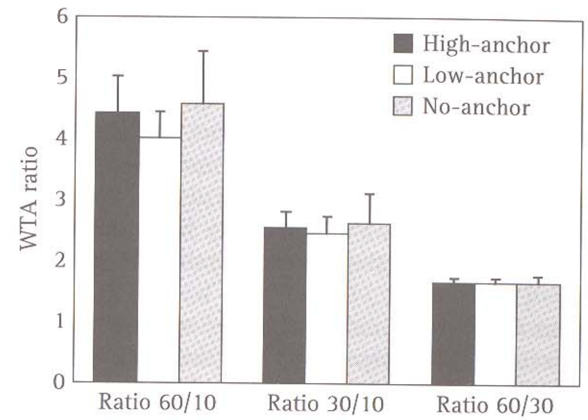

Figure 8.3. *Mean of individual WTA ratios for the different durations across the different conditions*

Note: Error bars are based on standard errors.

the durations to each of the other durations, and plotted these separately for the three conditions. As can be seen in Figure 8.3, the ratios of WTAs are stable and independent of condition (there are no significant differences by condition).

In summary, Experiment 1 demonstrates arbitrary but coherent pricing of painful experiences, even when there is no uncertainty about the nature or duration of the experience. Neither repeated experience with the event, nor confrontation with the same price distribution, overrode the impact of the initial anchor.

3.2. Experiment 2: Raising the Stakes and Using Random Anchors

Experiment 2 was designed to address two possible objections to the previous procedure. First, it could be argued that subjects might have somehow believed that the anchor was informative, even though they had experienced the sound for themselves. For example, they might have thought that the sound posed some small risk to their hearing, and might have believed that the anchor roughly corresponded to the monetary value of this risk. To eliminate this possibility, Experiment 2 used the subjects' own social security numbers as anchors. Second, one might be concerned that the small stakes in the previous experiment provided minimal incentives for accurate responding, which may have increased the arbitrariness of subjects' responses and their sensitivity to the anchor. Experiment 2, therefore, raised the stakes by a factor of ten. In addition, at the end of the experiment, we added a question designed to test whether the anchor-induced changes in valuation carry over to tradeoffs involving other experiences.

Ninety students at MIT participated in the experiment. The procedure followed closely that of Experiment 1, except that the stimuli were ten times as long: The shortest stimulus lasted 100 s, the next lasted 300 s, and longest lasted 600 s. The manipulation of the anchor in this experiment was also different. At the onset of the experiment, subjects were asked to provide the first three digits of their social security number and were instructed to turn these digits into a money amount (e.g. 678 translates into $6.78). Subjects were then asked whether, hypothetically, they would listen again to the sound they just experienced (300 s) if they were paid the money amount they had generated from their social security number.

In the main part of the experiment, subjects had three opportunities to listen to sounds in exchange for payment. The three different durations were again ordered in either an increasing set (100, 300, 600 s) or a decreasing set (600, 300, 100 s). In each trial, after they indicated their WTA, subjects were shown both their own price and the random price drawn from the distribution (which was the distribution used in Experiment 1 but multiplied by 10). If the price set by the subject was higher than the computer's price, subjects continued directly to the next trial. If the price set by the subjects was lower than the computer's price, subjects received the sound and the money associated with it (the amount set by the randomly drawn number), and then continued to the next trial. This process repeated itself three times, once for each of the three durations.

After completing the three trials, subjects were asked to rank order a list of events in terms of how annoying they found them (for a list of the different tasks see Table 8.2). At the end of the experiment, subjects were paid according to the payment rule.

Table 8.2. *The different events that subjects were asked to order-rank in terms of their annoyance, at the end of Experiment 3*

	The event	Mean rank
1	Missing your bus by a few seconds	4.3
2	Experiencing 300 s of the same sound you experienced	5.1
3	Discovering you purchased a spoiled carton of milk	5.2
4	Forgetting to return a video and having to pay a fine	5.4
5	Experiencing a blackout for an hour	5.8
6	Having a blood test	6.0
7	Having your ice cream fall on the floor	6.0
8	Having to wait 30 min in line for your favorite restaurant	6.2
9	Going to a movie theatre and having to watch it from the second row	6.7
10	Losing your phone bill and having to call to get another copy	7.3
11	Running out of toothpaste at night	8.1

Note: The items are ordered by their overall mean ranked annoyance from the most annoying (lower numbers) to the least annoying (higher numbers).

3.2.1. *Results*

The three digits entered ranged from 041 (translated to $0.41) to 997 (translated to $9.97), with a mean of 523 and a median of 505. Figure 8.4 compares the prices demanded by subjects with social security numbers above and below the median. It is evident that subjects with lower social security numbers required substantially less payment than subjects with higher numbers ($Ms = \$3.55$ and $\$5.76$; $F(1, 88) = 28.45$, $p < 0.001$). Yet, both groups were coherent with respect to duration, demanding more payment for longer sounds ($F(2, 176) = 92.53$, $p < 0.001$). As in the previous experiment, there was also a small but significant interaction between anchor and duration ($F(2, 176) = 4.17$, $p = 0.017$).

If subjects have little idea about how to price the sounds initially, and hence rely on the random anchor in coming up with a value, we would expect WTA in the first trial to be relatively close to the anchor, regardless of whether the duration was 100 or 600 s. However, having committed themselves to a particular value on the first trial, we would expect the increasing duration group to then adjust their values upward while the decreasing group should adjust their anchor downward. This would create a much larger discrepancy between the two groups' valuations of the final sound than existed for the initial sound. Figure 8.5 shows that the prediction is supported. Initial valuations of the 600-s tone in the decreasing order condition ($M = \$5.16$) were significantly larger than initial valuations of the 100-s tone in the increasing order condition ($M = \$3.78$; $t(88) = 3.1$, $p < 0.01$), but the difference of $1.38 is not very large. In the second

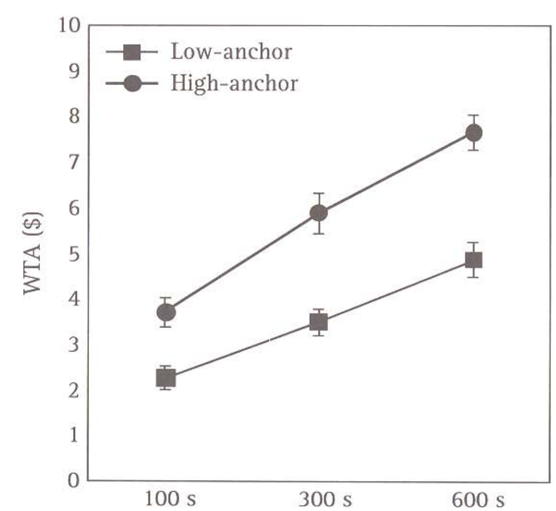

Figure 8.4. *Mean WTA (in $) for the three annoying sounds*

Notes: The data are plotted separately for subjects whose three-digit anchor was below the median (low-anchor) and above the median (high-anchor). Error bars are based on standard errors.

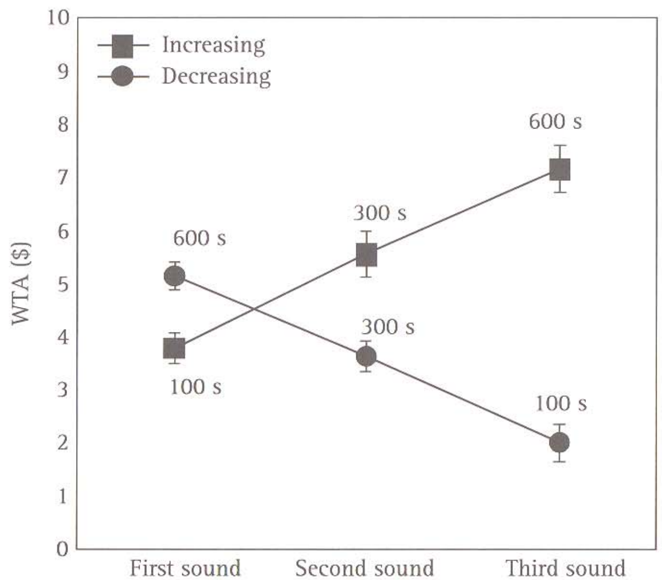

Figure 8.5. *Mean WTA (in \$) for the three annoying sounds*

Notes: The data are plotted separately for the increasing (100, 300, 600 s) and the decreasing (600, 300, 100 s) conditions. Error bars are based on standard errors.

period, both groups evaluated the same 300-s tone, and the valuation in the increasing condition was greater than that of the decreasing condition ($Ms = \$5.56$ and $\$3.65$; $t(88) = 3.5$, $p < 0.001$). By the final period, the two conditions diverged dramatically with WTA being much higher in the increasing condition compared with the decreasing condition ($Ms = \$7.15$ and $\$2.01$; $t(88) = 9.4$, $p < 0.0001$).

We now turn to the rank-ordering of the different events in terms of their annoyance (see Table 8.2). Recall that we wanted to see if the same anchor that influenced subjects' pricing would also influence the way they evaluated the sounds independently of the pricing task. The results showed that the rank-ordering of the annoyance of the sound was not influenced by either the anchor ($F(1, 86) = 1.33$, $p = 0.25$) or the order ($F(1, 86) = 0.221$, $p = 0.64$). In fact, when we examined the correlation between the rank ordering of the annoyance of the sound and the initial anchor, the correlation was slightly negative (−0.096), although this finding was not significant ($p = 0.37$).

In summary, Experiment 2 demonstrates that coherent arbitrariness persists even with randomly generated anchors and larger stakes. In addition, the last part of Experiment 2 provides some evidence that the effect of the anchor on pricing does not influence the evaluation of the experience relative to other experiences.

3.3. Experiment 3: Coherently Arbitrary Valuations in the Market

We now consider the possibility that the presence of market forces could reduce the degree of initial arbitrariness or facilitate learning over time. Earlier research that compared judgments made by individuals who were isolated or who interacted in a market found that market forces did reduce the magnitude of a cognitive bias called the "curse of knowledge" by approximately 50 percent (Camerer, Loewenstein, and Weber 1989).

To test whether market forces would reduce the magnitude of the bias, we exposed subjects to an arbitrary anchor (as in Experiment 1), but then elicited the WTA values through a multi-person auction, rather than using the Becker–Degroot–Marschak procedure. Our conjecture was that the market would not reduce the bias, but would lead to a convergence of prices within specific markets. Earlier research found that subjects who had bid on gambles in an auction similar to ours, adjusted their own bids in response to the market price, which carried information about the bids of other market participants (Cox and Grether 1996). Relying on others' values can be informative in some purchase settings, but in our markets all participants in the same group had been exposed to the same arbitrary anchor. Moreover, having experienced a sample of the noise, subjects had full information about the consumption experience, which makes the valuations of others prescriptively irrelevant.

Fifty-three students at MIT participated in the experiment, in exchange for a payment of \$5 and earnings from the experiment. Subjects were told that they would participate in a marketplace for annoying sounds, and that they would bid for the opportunity to earn money by listening to annoying sounds. They participated in the experiment in groups, varying in size from six to eight subjects. The experiment lasted approximately 25 min.

The design and procedure were very similar to Experiment 1, but with anchors of 10¢ and \$1.00 and an auction, rather than individual-level pricing procedure. Sound durations were 10, 30, or 60 s, and subjects were given three opportunities to listen to each of these sounds in an order that was manipulated between subjects. In the increasing condition, durations were presented in the order 10, 30, 60 s (repeated three times), and in the decreasing condition the durations were in the order 60, 30, 10 s (also repeated three times). All subjects first experienced 30 s of the same annoying sound that was used in the previous experiments. Next, the bidding procedure was explained to the subjects as follows:

> On each trial, the experimenter will announce the duration of the sound to be auctioned. At this stage every one of you will be asked to write down and submit your bid. Once all the bids are submitted, they will be written on the board by the experimenter, and the three people with the lowest bids will get the sound they bid for and get paid the amount set by the bid of the forth lowest person.

Each subject were then asked to write down whether, in a hypothetical choice, a sum of X (10¢ or 100¢ depending on their condition) would be sufficient compensation to listen to the sound again. At this point the main part of the experiment started. On each of the nine trials, the experimenter announced the duration of the sound that was being auctioned; each of the subjects wrote a bid on a piece of paper and passed it to the experimenter, who wrote the bids on a large board. At that point, the three lowest bidders were announced, and they were asked to put on their headphones and listen to the sound. After the sound ended the subjects who "won" the sound received the amount set by the fourth lowest bid.

3.3.1. *Results*

The general findings paralleled those from the previous experiments. In the low-anchor condition, the average bids were 24¢, 38¢, and 67¢ for the 10, 30, and 60 s sounds respectively (all differences between sound durations are significant within a condition), and in the high-anchor condition, the corresponding average bids were 47¢, \$1.32, and \$2.11. Overall, mean WTA in the low-anchor condition was significantly lower than WTA in the high-anchor condition ($F(1,49) = 20.38$, $p < 0.001$). The difference in the amount of money earned by subjects in the two conditions was quite stunning: The mean payment per sound in the high-anchor condition was \$0.59 while the mean payment the low-anchor condition was only \$0.08 ($F(1,49) = 26.24$, $p < 0.001$).

The main question that Experiment 3 was designed to address is whether the WTA prices for the low- and high-anchor conditions would converge over time. As can be seen from Figure 8.6 there is no evidence of convergence, whether one looks at mean bids or the mean of the prices that emerged from the auction.

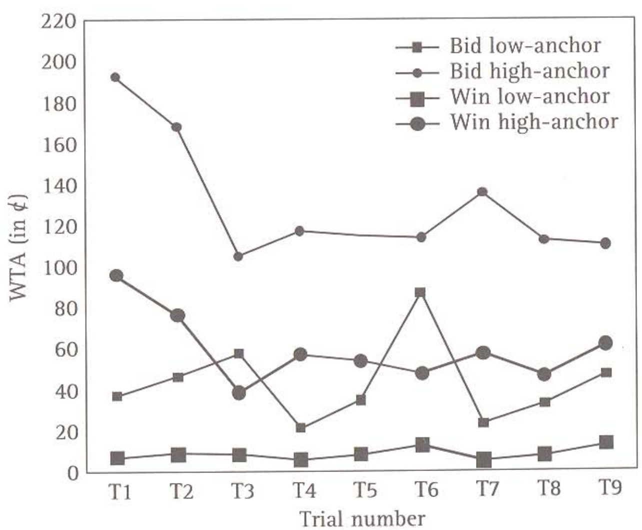

Figure 8.6. *Mean bids (WTA) and mean payment as a function of trial and the two anchor conditions*

Although the bids and auction prices in the different conditions did not converge to a common value, bids *within* each group did converge toward that group's arbitrary value. Figure 8.7, which plots the mean standard deviation of bids in the eight different markets for each of the nine trials, provides visual support for such convergence. To test whether convergence was significant, we first estimated the linear trend in standard deviations across the nine rounds separately for each group. Only one of the eight within-group trends was positive (0.25) and the rest were negative (ranging from -0.76 to -14.89). A two-tailed t-test of these eight estimates showed that they were significantly negative ($t(7) = 2.44$, $p < 0.05$).

In summary, Experiment 3 demonstrates that coherent arbitrariness is robust to market forces. Indeed, by exposing people to others who were exposed to the same arbitrary influences, markets can strengthen the impact of arbitrary stimuli, such as anchors, on valuation. The real-world analog of Experiment 3 would be an experiment where a unique new product (such as Viagra) was introduced at two different price points in two segregated and economically non-communicating markets. Individuals in one market would all be "anchored" to a single value. Such real life experiments might be difficult today, because of information flows. Nevertheless, our experimental results predict that the two markets would reach entirely different price equilibria, even if the economic fundamentals (production functions, consumers' tastes) were *ex ante* identical.

3.4. Experiment 4: The Impact of Multiple Anchors

According to our account of preference formation, the very first valuation in a given domain has an arbitrary component that makes it vulnerable to anchoring

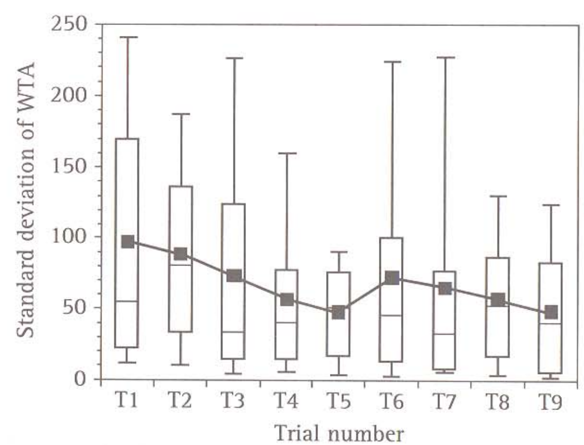

Figure 8.7. *The within-group standard deviations of the bids (WTA), plotted as a function of trial*

and similar manipulations. However, once individuals express these somewhat arbitrary values, they later behave in a fashion that is consistent with them, which constrains the range of subsequent choices and renders them less subject to non-normative influences. To test this, Experiment 4 exposed subjects to three different anchors instead of only one. If the imprinting account is correct, then the first anchor should have a much greater impact on valuations compared with later ones. If, on the other hand, subjects are paying attention to anchors because they believe they carry information, then all anchors would have the same impact as the initial one (similarly, Bayesian updating predicts that the order in which information arrives is irrelevant).

At the end of the pricing part of the experiment, we gave subjects a direct choice between an annoying sound and a completely different unpleasant stimulus. We did this to see whether the influence of the anchor extends beyond prices to qualitative judgments of relative aversiveness.

Forty-four students at MIT participated in the experiment, which lasted about 25 min. The experiment followed a procedure similar to the one used in Experiment 2, with the following adjustments. First, there were only three trials, each lasting 30 s. Second, and most important, in each of the three trials subjects were introduced to a new sound with different characteristics: A constant high pitch sound (the same as in the first three experiments), a fluctuating high pitch sound (which oscillated around the volume of the high pitched sound), or white noise (a broad spectrum sound). The important aspect of these sounds is that they are qualitatively different from each other, but similarly aversive.

After hearing each sound, subjects were asked if, hypothetically, they would listen to it again for 30 s in exchange for 10¢, 50¢, or 90¢ (depending on the condition and the trial number). Subjects in the increasing conditions answered the hypothetical questions in increasing order (10¢, 50¢, 90¢), and subjects in the decreasing conditions answered the hypothetical questions in decreasing order (90¢, 50¢, 10¢). Each of these hypothetical questions was coupled with a different sound. After answering each hypothetical question, subjects went on to specify the smallest amount of compensation they would require to listen to 30 s of that sound (WTA). The same Becker–Degroot–Marschak procedure used in Experiment 1 determined whether subjects heard each sound again and how much they were paid for listening to it.

After the three trials, subjects were asked to place their finger in a vise (see Ariely 1998). The experimenter closed the vise slowly until the subject indicated that he or she just began to experience the pressure as painful—a point called the "pain threshold." After the pain threshold was established, the experimenter tightened the vise an additional 1 mm (a quarter-turn in the handle) and instructed the subject to remember the level of pain. Subjects then experienced the same sound, and answered the same anchoring question that they had been asked, in the first trial. They were then asked if they would prefer to experience the same sound for 30 s or the vise for 30 s.

3.4.1. *Results*

Figure 8.8 displays mean WTAs for the three annoying sounds, and the two anchoring orders. With respect to the first bid, the low-anchor generated significantly lower bids ($M = 33.5$¢) than the high-anchor ($M = 72.8$¢; $F(1, 42) = 30.96$, $p < 0.001$). More interesting is the way subjects reacted to the second bid, which had the same anchor (50¢) for both conditions. In this case, we can see that there was a carryover effect from the first bid, so that the mean WTA price for the sound in the increasing condition ($M = 43.5$¢) was lower than the sound in the decreasing condition ($M = 63.2$¢; $F(1, 42) = 6.03$, $p < 0.02$). The most interesting comparison, however, is the WTA associated with the third sound. For this sound, subjects in both conditions had been exposed to the same three anchors, but the effects of the initial anchor and the most recent anchor (preceding the final stimulus) were in opposition to each other. In the increasing condition, the initial anchor was 10¢ and the most recent anchor was 90¢. In the decreasing condition, the initial anchor was 90¢ and the most recent anchor was 10¢. If the most recent anchor is stronger than the initial anchor, then WTA in the increasing condition should be higher than the one in the decreasing condition. If the initial anchor is stronger than the most recent anchor, as predicted by the imprinting account, then WTA in the decreasing condition should be higher than WTA in the increasing condition. In fact, WTA was higher in the decreasing condition compared with the increasing condition ($Ms = 63.1$¢ and 45.3¢; $F(1, 42) = 5.82$, $p < 0.03$). Thus, the initial anchor has a stronger effect on

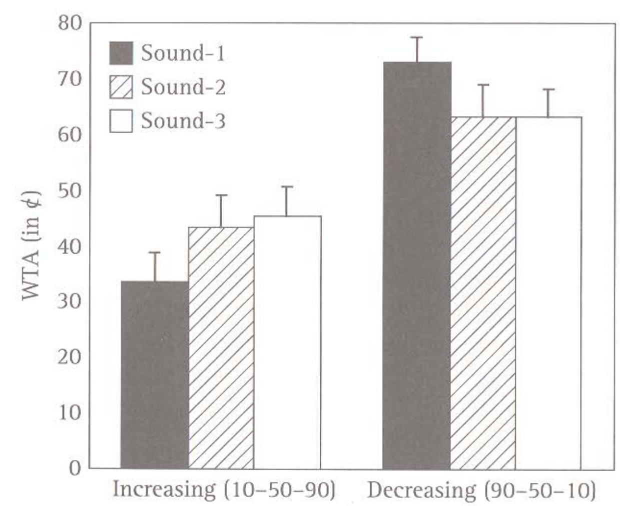

Figure 8.8. *Mean WTA (in cents) for the three annoying sounds*

Notes: In the increasing condition the order of the hypothetical questions was 10¢, 50¢, and 90¢, respectively. In the decreasing condition the order of the hypothetical questions was 90¢, 50¢, and 10¢, respectively. Error bars are based on standard errors.

WTA than the anchor that immediately preceded the WTA judgment, even though the initial anchor had been associated with a qualitatively different sound.

Another way to examine the results of Experiment 4 is to look at the binary responses to the hypothetical questions (the anchoring manipulation). In the first trial, the proportion of subjects who stated that they would be willing to listen to the sound they had just heard for X¢ was different, but not significantly so, across the two anchor values (55 percent for 10¢ and 73 percent for 90¢; $p > 0.20$ by χ^2 test). The small differences in responses to these two radically different values supports the idea that subjects did not have firm internal values for the sounds before they encountered the first hypothetical question. On the third trial, however, the difference was highly significant (41 percent for 10¢ and 82 percent for 90¢, $p < 0.001$ by χ^2 test). Subjects who were in the increasing anchor condition were much more willing to listen to the sound, compared to subjects in the decreasing anchor condition, indicating that they were sensitive to the change in money amounts across the three hypothetical questions. Consistent with the imprinting account proposed earlier, subjects acquired a stable internal reservation price for the sounds.

The response to the choice between the sound and vise pain revealed that subjects in the increasing anchor condition had a higher tendency to pick the sound (72 percent), compared with the decreasing anchor condition (64 percent), but this difference was not statistically significant ($p = 0.52$). These results again fail to support the idea that the anchor affects subjects' evaluations of the sound relative to other stimuli.

3.5. Experiment 5: Money Only?

The previous experiments demonstrated arbitrariness in money valuations. Neither of the follow-up studies (in Experiments 2 and 4), however, found that the anchoring manipulation affected subsequent choices between the unpleasant sounds and other experiences. This raises the question of whether these null results reflect the fact that the effects of the anchor are narrow, or that the coherent arbitrariness phenomenon arises only with a relatively abstract response dimensions, like money. To address this issue, we conducted an experiment that employed a design similar to that of Experiments 1–3 but which did not involve money. Because Experiments 1–3 had all demonstrated coherence on the dimension of duration, in Experiment 5, we attempted to demonstrate arbitrariness with respect to duration.

Fifty-nine subjects were recruited on the campus of the University of California at Berkeley with the promise of receiving $5.00 in exchange for a few minutes of their time and for experiencing some mildly noxious stimuli. After consenting to participate, they were first exposed to the two unpleasant stimuli used in the experiment: A small sample of an unpleasant tasting liquid composed of equal parts Gatorade and vinegar, and an aversive sound (the same as used in Experiments 1–3). They were then shown three containers of different

sizes (1, 2, and 4 oz), each filled with the liquid they had just tasted and were asked to "please answer the following hypothetical question: Would you prefer the middle size drink or X minutes of the sound" where X was 1 min for half the subjects and 3 min for the other half (the anchor manipulation). After the initial anchoring question subjects were shown three transparent bottles with different drink quantities in each (1, 2, and 4 oz). For each of the three drink quantities, subjects indicated whether they would prefer to drink that quantity of liquid or endure a sound lasting 10, 20, 30 s, etc. up to 8 min. (The specific instructions were "On each line, please indicate if you prefer that duration of the sound to the amount of the drink. Once you have answered all the questions the experimenter will pick one of the lines at random, and you will be asked to experience the sound described on that line or the drink depending on your preference in that line.") To simplify the task, the choices were arranged in separate blocks for each drink size, and were arranged in order of duration.

3.5.1. *Results*

Revealing arbitrariness with respect to tone duration, the anchoring manipulation had a significant impact on tradeoffs between the sound's duration and drink quantity ($F(1, 57) = 24.7$, $p < 0.0001$). The mean maximum tone duration at which subjects preferred the tone to the drink (averaging over the three drink sizes) was 82 s in the 1-min anchor condition and 162 s in the 3-min anchor condition. Revealing consistency with respect to tone duration, however, subjects were willing to tolerate longer sound durations when the other option involved larger drink size ($F(2, 114) = 90.4$, $p < 0.0001$) (see Figure 8.9).

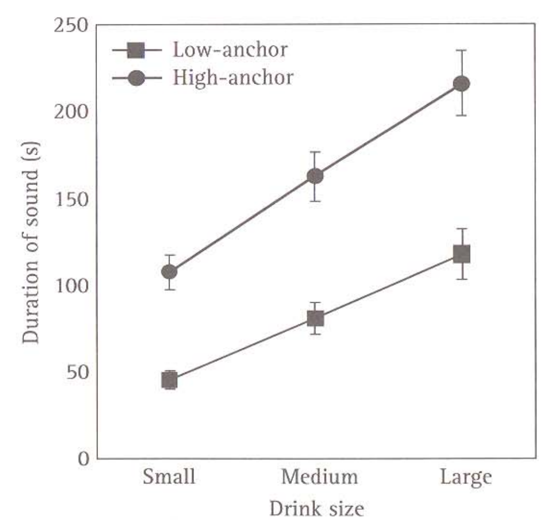

Figure 8.9. *Mean maximum duration at which subjects prefers tone to drink*

Note: Error bars are based on standard errors.

Experiment 5 demonstrates that arbitrariness is not limited to monetary valuations (and, less importantly, that coherence is not an inherent property of duration). In combination with the results of the add-on components of Experiments 2 and 4, it suggests that the web of consistency that people draw from their own choices may be narrow. Thus, for example, a subject in our common products experiment with a high social security number who priced the average wine at $25, would almost surely price the higher quality wine above $25. However, the same individual's subsequent choice of whether to trade the higher quality wine for a different type of good might be relatively unaffected by her pricing of the wine, and hence by the social security number anchoring manipulation.

4. GENERAL DISCUSSION

The first four experiments presented here show that when people assess their own willingness to listen to an unpleasant noise in exchange for payment, the money amounts they specify display the pattern that we call "coherent arbitrariness." The common products experiment demonstrated the pattern with familiar consumer products, and Experiment 5 showed that the pattern is not restricted to judgments about money. Coherent arbitrariness has two aspects: Coherence, whereby people respond in a robust and sensible fashion to noticeable changes or differences in relevant variables, and arbitrariness, whereby these responses occur around a base-level that is normatively arbitrary.

Next, we discuss a variety of other phenomena that may be interpreted as manifestations of coherent arbitrariness.

4.1. Contingent Valuation

The clearest analogy to our research comes from research on contingent valuation, in which people indicate the most they would have willingness to pay (WTP) for a public benefit (e.g. environmental improvement). Of particular relevance to coherent arbitrariness is the finding that people's willingness to pay for environmental amenities is remarkably unresponsive to the scope or scale of the amenity being provided (Kahneman and Knetsch 1992). For example, one study found that WTP to restore the fish population in one lake in Ontario was statistically indistinguishable from willingness to restore it in all lakes in Ontario (Kahneman and Knetsch 1992).

Importantly, insensitivity to scale is most dramatic in studies that employ between-subjects designs. When scope or scale is varied within-subject, so that a single person is making judgments for different values, the valuations are far more responsive to scale (see Kahneman, Schkade, and Sunstein 1998; Kahneman, Ritov, and Schkade 1999). In fact, the tendency for within-subject manipulations to produce larger effects than between subject manipulations is a common phenomenon (e.g. Keren and Raaijmakers 1988; Kahneman and Ritov 1994; Fox and Tversky 1995).

Increased sensitivity in within-subjects designs has even been observed in a study that examined intuitive pricing of common household items. Frederick and Fischhoff (1998) elicited WTPs for two different quantities of common market goods (e.g. toilet paper, apple sauce, and tuna fish) using both a between-subjects design (in which respondents valued either the small or large quantity of each good) and a within-subjects design (in which respondents valued both the small and large quantity of each good). The difference in WTP was in the right direction in both designs, but it was much greater (2.5 times as large) in the within-subjects condition, despite the fact that for goods such as toilet paper the meaning of the quantity description (number of rolls) should have been easy to evaluate. Frederick and Fischhoff (1998: 116) suggest that "valuations of any particular quantity (of good) would be sensitive to its relative position within the range selected for valuation, but insensitive to which range is chosen, resulting in insensitive (or incoherent) values across studies using different quantity ranges."

4.2. Financial Markets

Like the price one should ask to listen to an aversive tone, the value of a particular stock is inherently ambiguous. As Shiller (1998) comments, "Who would know what the value of the Dow Jones Industrial Average should be? Is it really 'worth' 6000 today? Or 5000 or 7000? Or 2000 or 10,000? There is no agreed-upon economic theory that would answer these questions." In the absence of better information, past prices (asking prices, prices of similar objects, or other simple comparisons) are likely to be important determinants of prices today." In a similar vein, Summers (1986) notes that it is remarkably difficult to demonstrate that asset markets reflect fundamental valuation. It is possible to show that one or more prediction of the strong markets theory are supported, but "the verification of one of the theory's predictions cannot be taken to prove or establish a theory" (1986: 594). Thus, studies showing that the market follows a random walk are consistent with fundamental valuation, but are insufficient to demonstrate it; indeed, Summers presents a simple model in which asset prices have a large arbitrary component, but are nevertheless serially uncorrelated, as predicted by fundamental valuation.

While the overall value of the market or of any particular company is inherently unknowable, the impact of particular pieces of news is often quite straightforward. If Apple was expected to earn $\$x$ in a particular year but instead earned $\$2x$, this would almost unquestionably be very good news. If IBM buys back a certain percentage of its own outstanding shares, this has straightforward implications for the value of the remaining shares. As Summers (1986) points out, the market may respond in a coherent, sensible fashion to such developments even when the absolute level of individual stocks, and of the overall market, is arbitrary.

4.3. Labor Markets

In the standard account of labor supply, workers intertemporally substitute labor and leisure with the goal of maximizing utility from lifetime labor, leisure, and consumption. To do so optimally, they must have some notion of how much they value these three activities, or at least of how much they value them relative to one another. Although it is difficult to ascertain whether labor supply decisions have an element of arbitrariness, due to the absence of any agreed-upon benchmark, there is some evidence of abnormalities in labor markets that could be attributed to arbitrariness. Summarizing results from a large-scale survey of pay-setting practices by employees, Bewley (1998: 485) observes that "Non-union companies seemed to be isolated islands, with most workers having little systematic knowledge of pay rates at other firms. Pay rates in different non-union companies were loosely linked by the forces of supply and demand, but these allowed a good deal of latitude in setting pay." Wage earners, we suspect, do not have a good idea of what their time is worth when it comes to a tradeoff between consumption and leisure, and do not even have a very accurate idea of what they could earn at other firms. Like players in the stock market, the most concrete datum that workers have with which to judge the correctness of their current wage rate is the rate they were paid in the past. Consistent with this reasoning, Bewley continues, "though concern about worker reaction and morale curbed pay cutting, the reaction was to a reduction in pay relative to its former level. The fall relative to levels at other firms was believed to have little impact on morale, though it might increase turnover." In other words, workers care about changes in salary but are relatively insensitive to absolute levels or levels relative to what comparable workers make in other firms. This insensitivity may help to explain the maintenance of substantial inter-industry wage differentials (see Dickens and Katz 1987; Krueger and Summers 1988; Thaler 1989). Similarly, coherent arbitrariness is supported by the quip that a wealthy man is one who earns $100 more than his wife's sister's husband.

4.4. Criminal Deterrence

Imagine an individual who is contemplating committing a crime, whether something as minor as speeding on a freeway, or something as major as a robbery. To what extent will such an individual be deterred by the prospect of apprehension? Research on criminal deterrence has produced mixed answers to this question, with some studies finding significant negative effects of probability or severity of punishment on crime, and others reaching more equivocal conclusions. These studies have employed different methodologies, with some examining cross-sectional differences in crime and punishment across states, and others examining changes over time. Coherent arbitrariness has important implications for these studies. Like many other types of cost–benefit calculations, assessing the probabilities and likely consequences of apprehension is difficult,

as is factoring such calculations into one's decision-making calculus. Thus, this is a domain characterized by value uncertainty where one might expect to observe the coherent arbitrariness pattern. Coherent arbitrariness, in this case, would mean that people would respond sensibly to well-publicized *changes* in deterrence levels but much less to absolute levels of deterrence (for a discussion of similar results in civil judgments, see Sunstein et al. 2002). We would predict, therefore, that one should find short-term deterrence effects in narrowly focused studies that examine the impact of policy changes, but little or no deterrence effects in cross-sectional studies. This is, indeed, the observed pattern. Interrupted time series studies have measured sizeable reactions in criminal behavior to sudden, well-publicized increases in deterrence (Ross 1973; Sherman 1990), but these effects tend to diminish over time. The implication that we draw is that the prevailing level of criminal activity does not reflect any underlying fundamental tradeoff between the gains from crime and the costs of punishment.

5. FINAL COMMENTS

Recently, there has been a significant outflow of research on the accuracy of people's predictions of their own future feelings (Loewenstein and Angner, forthcoming). The almost uniform conclusion from this research is that people are not very good at predicting how different experiences and outcomes will make them feel (Gilbert, Driver-Linn, and Wilson, forthcoming), and are especially bad at predicting how long their feelings will last (Gilbert et al. 1998). The sound experiments presented here–in which subjects valued experiences that they had just sampled, points to an even more dire conclusion: People may not even be able to assess their own feelings at the moment when they are making the judgment. Moreover, research presented in a prequel to this volume (Schooler, Ariely, and Loewenstein, forthcoming) shows that either trying to enjoy an experience or even just monitoring one's enjoyment of it actually decreased reported enjoyment. Taken as a body, this research suggests that it may not be possible to maximize pleasure, and indeed might not be worth doing even if it were possible.

More prosaically, our experiments highlight the general hazards of inferring fundamental valuation by examining individuals' responses to change. If all one observed from our experiment was the relationship between valuation and duration, one might easily conclude that people were basing their WTA values on their fundamental valuation for the different stimuli. However, the effect of the arbitrary anchor shows that, while people are adjusting their valuations in a coherent, seemingly sensible fashion to account for duration, they are doing so around an arbitrary base value. Moreover, this effect does not diminish as subjects gain more experience with the stimulus or when they provide valuations in a market context.

A key economic implication of coherent arbitrariness is that some economic variables will have a much greater impact than others. When people recognize that a particular economic variable, such as a price, has changed, they will respond robustly but when the change is not drawn to their attention, they will respond more weakly, if at all. This point was recognized early-on by the economist John Rae who, in 1834, noted that

> When any article rises suddenly and greatly in price, when in their power, they are prone to adopt some substitute and relinquish the use of it. Hence, were a duty at once imposed on any particular wine, or any particular sort of cotton fabric, it might have the effect of diminishing the consumption very greatly, or stopping it entirely. Whereas, were the tax at first slight, and then slowly augmented, the reasoning powers not being startled, vanity, instead of flying off to some other objects, would be apt to apply itself to them as affording a convenient means of gratification. (1834: 374)

The speed at which an economic variable changes is only one of many factors that will determine whether it is visible to individuals—whether it "startles" the reasoning powers, as Rae expressed it. Other factors that can make a difference are how the information is presented—for example, whether prices of alternative products are listed in a comparative fashion or are encountered sequentially (see Russo and Leclerc 1991), and whether prices are known privately or discussed. Thus, for example, large salary differentials may be easier to sustain in a work environment in which salary information is not discussed. In sum, changes or differences in prices or other economic conditions will have a much greater impact on behavior when people are made aware of the change or difference than when they are only aware of the prevailing levels at a particular point in time.

These results challenge the central premise of welfare economics that choices reveal true preferences—that the choice of A over B indicates that the individual will in fact be better off with A rather than with B. It is hard to make sense of our results without drawing a distinction between "revealed" and "true" preferences. How, for example, can a pricing decision that is strongly correlated with an individual's social security number reveal a true preference in any meaningful sense of the term? If consumers' choices do not necessarily reflect true preferences, but are to a large extent arbitrary, then the claims of revealed preferences as a guide to public policy and the organization of economic exchange are weakened. Market institutions that maximize consumer sovereignty need not maximize consumer welfare.

As many economists have pointed out (e.g. Sen 1982), the sole psychological assumption underlying ordinal utility is that people will behave consistently. Our work suggests that ordinal utility may, in fact, be a valid representation of choices under specific, albeit narrow, circumstances, without revealing underlying preferences, in any non-vacuous sense of "preference." When people are aware of changes in conditions, such as the change in price in the example just given, they will respond in a coherent fashion that mimics the behavior of individuals with fixed, well-defined preferences. However, they will often not respond reasonably to new opportunities or to hidden changes in old variables, such as price or quality. The equilibrium states of the economy may, therefore, contain a large arbitrary component, created by historical accident or deliberate manipulation.

REFERENCES

Ariely, Dan (1998), "Combining experiences over time: The effects of duration, intensity changes & on-papers line measurements on retrospective pain evaluations." *Journal of Behavioral Decision Making*, 11: 19–45.

—— and Loewenstein, George. (2000), "The importance of duration in ratings of, and choices between, sequences of outcomes." *Journal of Experimental Psychology: General*, 129: 508–23.

—— and Zauberman, G. (2000), "On the making of an experience: The effects of breaking and combining experiences on their overall evaluation." *Journal of Behavioral Decision Making*, 13: 219–32.

—— Lowenstein, George F., and Prelec, Drazen (2003), "Coherent Arbitrariness: Stable Demand Curves without Stable Preferences," *Quarterly Journal of Economics*, 118: 73–105.

Becker, Gary M. (1962). "Irrational behavior and economic theory." *Journal of Political Economy* 70: 1–13.

Becker, Gordon, De Groot, M., and Marschak, Jacob. (1964), "Measuring utility by a single-response sequential method." *Behavioral Science*, 9: 226–32.

—— DeGroot, Morris H., and Marschak, Jacob. (1963), "An experimental study of some stochastic models for wagers." *Behavioral Science*, 8: 199–202.

Bewley, Truman F. (1986), Knightian Decision Theory: Part I. Yale University, Cowles Foundation Discussion Paper, No. 807.

—— (1998), "Why not cut pay?" *European Economic Review*, 42: 459–90.

Bodner, Ronit, and Prelec, Drazen. (2003), "Self-signaling in a neo-Calvinist model of everyday decision making", in *Essays in Psychology and Economics*, Vol I. I. Brocas and J. Carillo (eds.), (Oxford: Oxford University Press).

Camerer, Colin, Loewenstein, George, and Weber, Martin (1989), "The curse of knowledge in economic settings: An experimental analysis." *Journal of Political Economy*, 97: 1232–54.

Chapman, Gretchen B. and Johnson, Eric J. (1999), "Anchoring, activation, and the construction of values." *Organizational Behavior and Human Decision Processes*, 79: 115–53.

Cox, James C. and Grether, David M. (1996), "The preference reversal phenomenon: Response mode, markets and incentives." *Economic Theory*, 7: 381–405.

Dickens, William T. and Katz, Lawrence F. (1987), "Inter-industry wage differences and industry characteristics." In K. Lang and J. S. Leonard (eds.), *Unemployment and the Structure of Labor Markets* (Oxford: Basil Blackwell).

Drolet, Aimee L., Simonson, Itamar, and Tversky, Amos (2000), "Indifference curves that travel with the choice set." *Marketing Letters*, 11(3): 199–209.

Epley, Nicholas and Thomas, Gilovich. (2001), "Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors." *Psychological Science*, 12: 391–96.

Fox, Craig R. and Tversky, Amos (1995), "Ambiguity aversion and comparative ignorance." *Quarterly Journal of Economics*, 110(3): 585–603.

Frederick, Shane and Fischhoff, Baruch (1998), "Scope (in)sensitivity in elicited valuations." *Risk Decision and Policy*, 3: 109–23.

Gilbert, Daniel T., Driver-Linn, E., and Wilson, Timothy D. (2002), "The trouble with Vronsky: Impact bias in the forecasting of future affective states." In L. Feldman-Barrett and P. Salovey (eds.), *The Wisdom of Feeling*, 114–43.

—— Pinel, Elizabeth C., Wilson, Timothy D., Blumberg, Stephen J., and Wheatley, Thalia P. (1998), "Immune neglect: A source of durability bias in affective forecasting." *Journal of Personality and Social Psychology*, 75(3): 617–38.

Gilboa, Itzhak and Schmeidler, David (1995), "Case-based decision theory." *Quarterly Journal of Economics*, 110: 605–39.

Green, Donald, Jacowitz, Karen E., Kahneman, Daniel, and McFadden, Daniel. (1998), "Referendum contingent valuation, anchoring, and willingness to pay for public goods." *Resources and Energy Economics*, 20: 85–116.

Hirschleifer, David and Welch, Iuo (1988), "A rational economic approach to the psychology of change: Amnesia, inertia, and impulsiveness." Working paper, November.

Hoeffler, Steve and Ariely, Dan (1999), "Constructing stable preferences: A look into dimensions of experience and their impact on preference stability." *Journal of Consumer Psychology*, 11: 113–39.

Hsee, Chris. K., Lowenstein George F., Blount, Sally, and Bazerman, Max H. (1999), "Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125: 576–90.

Jacowitz, Karen E. and Kahneman, Daniel (1995), "Measures of anchoring in estimation tasks." *Personality and Social Psychology Bulletin*, 21: 1161–6.

Johnson, Eric J. and Schkade, David A (1989), "Bias in utility assessments: Further evidence and explanations." *Management Science*, 35: 406–24.

Kahneman, Daniel and Knetsch, Jack (1992), "Valuing public goods: The purchase of moral satisfaction." *Journal of Environmental Economics and Management*, 22: 57–70.

—— and —— (1993), "Anchoring or shallow inferences: The effect of format." Unpublished manuscript, University of California, Berkeley.

—— and Miller, Dale. T. (1986), "Norm theory: Comparing reality to its alternatives." *Psychological Review*, 93: 136–53.

Kahneman, Daniel and Illana, Ritov (1994), "Determinants of stated willingness to pay for public goods. A study in the headline method." *Journal of Risk and Uncertainty*, 9: 5–38.

—— —— and Schkade, David (1999), "Economic preferences or attitude expressions? An analysis of dollar responses to public issues." *Journal of Risk and Uncertainty*, 19: 220–42.

—— Schkade, David A., and Sunstein, Cass R (1998), "Shared outrage and erratic awards: The psychology of punitive damages." *Journal of Risk and Uncertainty*, 16: 49–86.

—— Wakker, Peter P., and Sarin, Rakesh (1997), "Back to Bentham?: Explorations of experienced utility." *Quarterly Journal of Economics*, 112: 375–405.

Keren, Gideon and Raaijmakers, Jeroen B. (1988), "On between-subjects versus within-subjects comparisons in testing utility theory." *Organizational Behavior and Human Decision Processes*, 4: 233–47.

Krueger, Alan B. and Summers, Lawrence H. (1988), "Efficiency wages and the inter-industry wage structure." *Econometrica*, 56: 259–93.

Loewenstein, G. and Angner, E. (2003), "Predicting and indulging changing preferences." In G. Loewenstein, D. Read, and R. Baumeister (eds.), *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice* (New York: Russell Sage Foundation Press).

Mussweiler, Thomas and Strack, Fritz (2001), "Considering the impossible: Explaining the effects of implausible anchors." *Social Cognition*, 19: 145–60.

Payne, John. W., Bettman, James. R., and Johnson, Erik. J. (1993), *The Adaptive Decision Maker* (New York: Cambridge University Press).

Rae, John (1834/1905), *The Sociological Theory of Capital* (London: Macmillan).

Ross, H. Laurence (1973), "Law, science, and accidents: The British road safety act of 1967." *Journal of Legal Studies*, 2: 1–78.

Russo, J. Edward and Leclerc, France (1991), "Characteristics of successful product information programs." *Journal of Social Issues*, 47(1): 73–92.

Schooler, J., Ariely, D., and Loewenstein, G. (2003), "The pursuit of happiness can be self-defeating." In J. Carrillo and I. Brocas (eds.), *Psychology and Economics* (Oxford: Oxford University Press), forthcoming.

Sen, Amartya Kumar (1982), *Choice, Welfare, and Measurement* (Cambridge, MA: MIT Press).

Sherman, Lawrence (1990), "Police crackdowns: Initial and residual deterrence." In Michael Tonry and Norval Morris (eds.), *Crime and Justice: A Review of Research* (Chicago: University of Chicago Press).

Shiller, Robert. J. (1998), "Human behavior and the efficiency of the financial system." In John B. Taylor and Michael Woodford (eds.), *Handbook of Macroeconomics* (North Holland: Elsevier).

Simonson, Itamar and Tversky, Amos (1992), "Choice in context: Tradeoff contrast and extremeness aversion." *Journal of Marketing Research*, 29: 281–95.

Strack, Fritz and Mussweiler, Thomas (1997), "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility." *Journal of Personality and Social Psychology*, 73: 437–46.

Summers, Lawrence H. (1986), "Does the stock market rationally reflect fundamental values?" *Journal of Finance*, 41(3): 591–601.

Sunstein, Cass R., Kahneman, Daniel, Schkade, David, and Ritov, Ilana (2002), "Predictably incoherent judgments." Working paper, University of Chicago Law school.

Thaler, Richard H. (1989), "Inter-industry wage differentials." *Journal of Economic Perspectives*, 3: 181–93.

Tversky, Amos and Kahneman, Daniel (1974), "Judgment under uncertainty: Heuristics and biases." *Science*, 185: 1124–31.