

“COHERENT ARBITRARINESS”: STABLE DEMAND CURVES WITHOUT STABLE PREFERENCES*

DAN ARIELY
GEORGE LOEWENSTEIN
DRAZEN PRELEC

In six experiments we show that initial valuations of familiar products and simple hedonic experiences are strongly influenced by arbitrary “anchors” (sometimes derived from a person’s social security number). Because subsequent valuations are also coherent with respect to salient differences in perceived quality or quantity of these products and experiences, the entire pattern of valuations can easily create an illusion of order, as if it is being generated by stable underlying preferences. The experiments show that this combination of *coherent arbitrariness* (1) cannot be interpreted as a rational response to information, (2) does not decrease as a result of experience with a good, (3) is not necessarily reduced by market forces, and (4) is not unique to cash prices. The results imply that demand curves estimated from market data need not reveal true consumer preferences, in any normatively significant sense of the term.

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 trade-off between the desire for consumption and the displeasure of work. Finally, and most importantly for this paper, consumer microeconomics 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 commodi-

* We thank Colin Camerer, Shane Frederick, John Lynch, James Bettman, and Birger Wernerfelt for helpful comments and suggestions. We are also grateful for financial support to Ariely and Prelec from the Sloan School of Management, to Loewenstein from the Integrated Study of the Human Dimensions of Global Change at Carnegie Mellon University (NSF grant No. SBR-9521914), and to Loewenstein and Prelec from the Center for Advanced Study in the Behavioral Sciences (NSF grant No. SBR-960123 to the Center, 1996–1997).

ties are downward sloping, or whether stock prices respond in the predicted way to share repurchases. However, such “comparative static” relationships are a necessary but not sufficient condition for fundamental valuation (e.g., Summers [1986]). 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 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 paper we show that consumers’ absolute valuation of experience goods is surprisingly arbitrary, even under “full information” conditions. However, we also show that consumers’ *relative* valuations of different amounts of the good appear orderly, as if supported by demand curves derived from fundamental preferences. Valuations therefore display a combination of arbitrariness and coherence that we refer to as “coherent arbitrariness.”

Our findings are consistent with an account of revealed preference which posits that valuations are initially malleable but become “imprinted” (i.e., precisely defined and largely invariant), after the individual is called upon to make an initial decision.¹ Prior to imprinting, valuations have a large arbitrary component, meaning that they are highly responsive to both normative and nonnormative influences. Following imprinting, valuations be-

1. The idea that preferences are not well defined, but become articulated in the process of making a decision is consistent with a large body of research on what decision researchers refer to as “constructed preferences” (e.g., Slovic [1995] and Hoeffler and Ariely [1999]).

come locally coherent, as the consumer attempts to reconcile future decisions of a "similar kind" with the initial one. This creates an illusion of order, because consumers' coherent responses to subsequent *changes* in conditions disguise the arbitrary nature of the initial, foundational, choice.

I. EXPERIMENT 1: COHERENTLY ARBITRARY VALUATION OF ORDINARY PRODUCTS

Our experiments take an old trick from the experimental psychologists' arsenal—the anchoring manipulation—and use it to influence valuation of products and hedonic experiences with normatively irrelevant factors. In a famous early demonstration of anchoring, Tversky and Kahneman [1974] spun a wheel of fortune with numbers that ranged from 0 to 100, asked subjects whether the number 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.² This, and many other anchoring studies seemed to show that people lack preexisting subjective probability distributions over unknown quantities.

The vast majority of anchoring experiments in the psychological literature have focused on how anchoring corrupts subjective judgment, not subjective valuation or preference. Because valuation typically involves judgment, however, it is not surprising that valuation, too, can be moved up or down by the anchoring manipulation. Johnson and Schkade [1989] were the first to demonstrate this experimentally. They showed that asking subjects whether their certainty equivalent for a lottery is above or below an anchor value influences subsequently stated certainty equivalents. Green, Jacowitz, Kahneman, and McFadden [1998], and Kahneman and Knetsch [1993], found the same effect with judgments of willingness-to-pay for public goods; higher values in the initial Yes/No question led to higher subsequent willingness-to-pay.

Our first experiment replicates these results with ordinary

2. For recent studies of anchoring, see, e.g., Chapman and Johnson [1999], Jacowitz and Kahneman [1995], Strack and Mussweiler [1997] and Epley and Gilovitch [2001].

TABLE I
AVERAGE STATED WILLINGNESS-TO-PAY 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	.415	.516	0.328	.328	0.319	.419
	$p = .0015$	$p < .0001$	$p = .014$	$p = .0153$	$p = .0172$	$p = .0013$

The last row indicates the correlations between Social Security numbers and WTP (and their significance levels).

consumer products. The first class meeting of a market research course in the Sloan School MBA program provided the setting for the study. Fifty-five students were shown 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, subjects 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 willingness-to-pay (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-De-groot-Marschak procedure [1963]). 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 percent 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 I. 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 preference 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.

II. COHERENT ARBITRARINESS

The sensitivity of WTP to anchors suggests that consumers do not arrive at a choice or at a pricing task with an inventory of preexisting preferences and probability distributions, which is consistent with a great deal of other psychological evidence [Kahneman and Miller 1986; Payne, Bettman, and Johnson 1993; Drolet, Simonson, and Tversky 2000]. Rather than specific WTP values for products, consumers probably 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, i.e., 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 our actual experiment, suppose that a subject with a social security number ending with 25 has an a priori WTP range of \$5 to \$30 for wine described as "average," and \$10 to \$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 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 between 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.³

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

3. Another research literature, on “evaluability,” is also relevant here. “Evaluability” has been identified as the cause of preference reversals that arise when options are evaluated either jointly (within subject) or separately (between subject). Hsee, Loewenstein, Blount, and Bazerman [1999] explain these reversals by assuming that it is more difficult to evaluate some attributes separately than jointly, depending on whether the attributes 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.

choices in a fashion that is consistent with them, not that these choices are generated from preexisting preferences.

III. VALUATION OF NOVEL PAIN EXPERIENCES

If preferences and valuations at a moment in time are largely inferences that a person draws from the history of his or her own previous decisions, a natural question that arises is whether the inference has a narrow scope (restricted only to very similar previous choices) or whether the scope is more general. For example, if I go on record as being willing to pay \$25 for a wine, will that only influence my subsequent willingness-to-pay for wine, for a broader range of items or experiences, or even for pleasures generally? The broader the scope of inferences, the more will previous choices constrain any future choice. If purchases of specific products and services function as precedents not just for those same items but also for the general valuation of pleasure (including here the avoidance of discomfort), then an adult consumer should have accumulated an inventory of previous choices sufficient to stabilize his or her dollar valuation of hedonic experiences.

In the next five experiments, 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 willingness to accept (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. They therefore entered the pricing phase of the experiment with full information about the experience they were pricing. Second, 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. Third, 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. Fourth, with annoying sounds it is possible to re-create 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].

IV. EXPERIMENT 2: COHERENTLY ARBITRARY VALUATION OF PAIN

The goal of Experiment 2 was to test 1) whether valuation of annoying sounds was susceptible to an anchoring manipulation; 2) whether additional experience with the sounds would erode the influence of the anchor; and 3) 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 the Massachusetts Institute of Technology 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 minutes, and subjects were paid according to their performance as described below.

At the beginning of the experiment, all subjects listened to an annoying, 30-second sound, delivered through headphones. The sound was a high-pitched scream (a triangular wave with frequency of 3,000 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**), and anchor price of 50¢ (**high-anchor**), and no anchor (**no-anchor**). Subjects in the low-anchor [high-anchor] condition first encountered a screen that read:⁴

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

4. In a different study [Ariely, Loewenstein, and Prelec 2002] we tested whether the order in which subjects received the sample and the anchor made a difference. It did not.

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 neither 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 seconds) and were asked to indicate their WTA for the sound. The three durations were presented either in an increasing (10 seconds, 30 seconds, 60 seconds) or decreasing order (60 seconds, 30 seconds, 10 seconds). 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 that price minus 5¢, and whether they would experience it for that price 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.

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 [$M = 43.87$; $F(1,126) = 12.17$, $p < 0.001$].⁵ 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.⁶

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 Ariely and Loewenstein [2000] and Kahneman, Wakker, and Sarin [1997]). The mean price for the 10 second sound [$M = 28.35$] was significantly lower than the mean

5. For the purpose of statistical analysis, responses above 100¢ (7.7 percent) were truncated to 101¢ (the highest random price selected by computer was 100¢, so responses above 101¢ were strategically equivalent). Repeating the analyses using untruncated values did not qualify any of the findings.

6. 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.

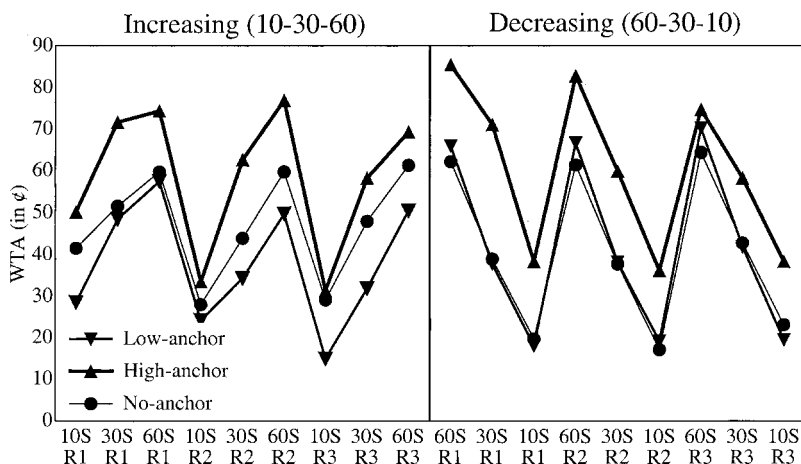


FIGURE I

Mean WTA for the Nine Trials in the Three Anchor Conditions

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

price for the 30 second sound [$M = 48.69$; $F(1,252) = 169.46$, $p < 0.001$], and the mean price for the 30 second sound was lower than the mean price for the 60 second sound [$M = 66.25$; $F(1,252) = 126.06$, $p < 0.001$].

Figure I 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 II provides additional support for the tight connection between WTA and duration. For each subject, we calculated the ratio of WTA in each of the durations to each of the other durations, and plotted these separately for the three conditions. As can be seen in the figure, the ratios of WTAs are stable, and independent of condition (there are no significant differences by condition).

In summary, Experiment 2 demonstrates arbitrary but coherent pricing of painful experiences, even when there is no uncertainty about the nature or duration of the experience. Nei-

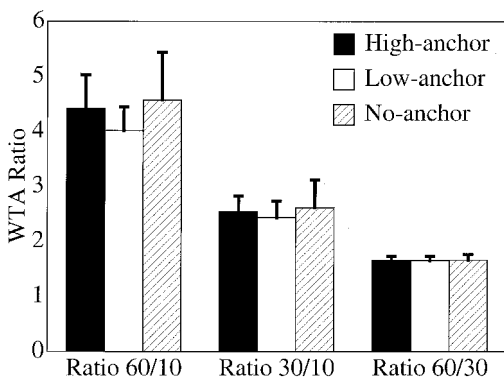


FIGURE II

Mean of Individual WTA Ratios for the Different Durations across the Different Conditions

Error bars are based on standard errors.

ther repeated experience with the event, nor confrontation with the same price distribution, overrode the impact of the initial anchor.

V. EXPERIMENT 3: RAISING THE STAKES

Experiment 3 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 3 used 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 3, 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 trade-offs involving other experiences.

Ninety students from the Massachusetts Institute of Tech-

nology participated in the experiment. The procedure closely followed that of Experiment 2, except that the stimuli were ten times as long: the shortest stimulus lasted 100 seconds; the next lasted 300 seconds, and longest lasted 600 seconds. 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 (for 300 seconds) 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 seconds, 300 seconds, 600 seconds) or a decreasing set (600 seconds, 300 seconds, 100 seconds). 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 2 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 II). At the end of the experiment, subjects were paid according to the payment rule.

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 III 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 [$M_s = \$3.55$, and $\$5.76$; $F(1,88) = 28.45$, $p < 0.001$]. 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

TABLE II

The event	Mean rank
1 Missing your bus by a few seconds	4.3
2 Experiencing 300 seconds 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 minutes in line for your favorite restaurant	6.2
9 Going to a movie theater 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

The different events that subjects were asked to order-rank in terms of their annoyance, at the end of Experiment 3. The items are ordered by their overall mean ranked annoyance from the most annoying (lower numbers) to the least annoying (high numbers).

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

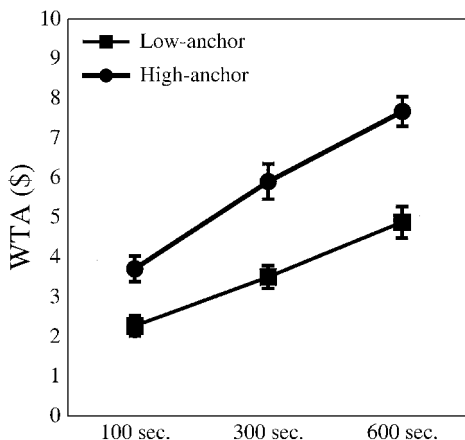


FIGURE III

Mean WTA (in Dollars) for the Three Annoying Sounds

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.

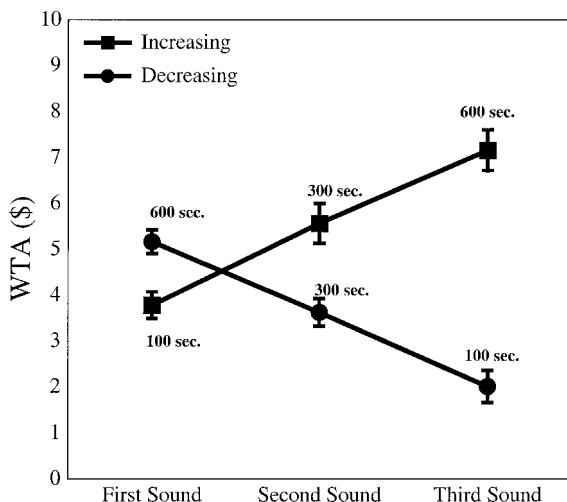


FIGURE IV

Mean WTA (in Dollars) for the Three Annoying Sounds

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

a value, we would expect responses to the initial question to be relatively close to the anchor, regardless of whether the duration was 100 seconds or 600 seconds. However, having committed themselves to a particular value for the initial sound, 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 IV shows that the prediction is supported. Initial valuations of the 600 second tone in the decreasing order condition [$M = \$5.16$] were significantly larger than initial valuations of the 100 second tone in the increasing order condition [$M = \$3.78$; $t(88) = 3.1, p < .01$], but the difference of \$1.38 is not very large. In the second period, both groups evaluated the same 300 second tone, and the valuation in the increasing condition was greater than that of the decreasing condition [$M_s = \$5.56$, and $\$3.65$; $t(88) = 3.5, p < .001$]. By the final period, the two conditions diverged dramatically with WTA being much higher in the increasing condition compared with the de-

creasing condition [$M_s = \$7.15$, and $\$2.01$; $t(88) = 9.4$, $p < .0001$].

We now turn to the rank-ordering of the different events in terms of their annoyance (see Table II). Recall that we wanted to see whether 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 3 demonstrates that coherent arbitrariness persists even with randomly generated anchors and larger stakes. In addition, the last part of Experiment 3 provides some evidence that the effect of the anchor on pricing does not influence the evaluation of the experience relative to other experiences.

VI. EXPERIMENT 4: 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 the second experiment), but then elicited the WTA values through a multiperson auction, rather than using the Becker-DeGroot-Marschak [1963] 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 these markets other

participants 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 from the Massachusetts Institute of Technology 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 minutes.

The design and procedure were very similar to Experiment 2, but we increased the high anchor to \$1.00 (instead of 50¢) and used an auction, rather than individual-level pricing procedure. As in the second experiment, the sound durations were 10, 30, or 60 seconds, subjects were given three opportunities to listen to each of these sounds and the order of the durations was manipulated between subjects. In the increasing condition, durations were presented in the order 10, 30, 60 seconds (repeated three times), and in the decreasing condition the durations were in the order 60, 30, 10 seconds (also repeated three times). All subjects first experienced 30 seconds of the same annoying sound that was used in the second experiment. 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 fourth lowest person.

Subjects 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 for them 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

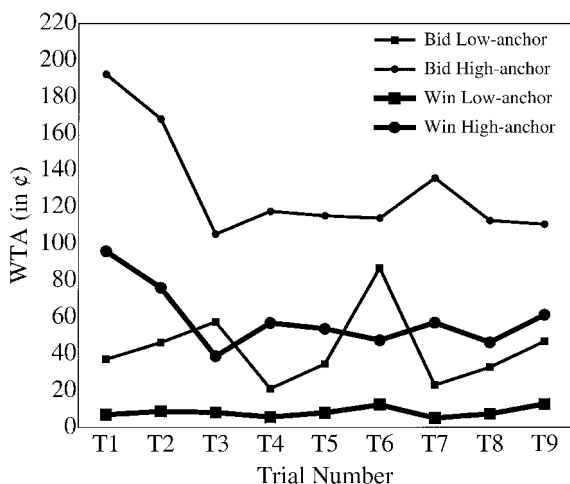


FIGURE V

Mean Bids (WTA) and Mean Payment as a Function of Trial and the Two Anchor Conditions

ended the subjects who “won” the sound received the amount set by the fourth lowest bid.

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 second 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 \$.59, while the mean payment in the low-anchor condition was only \$.08.

The main question that Experiment 4 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 V, there is no evidence of convergence, whether one looks at mean bids or the mean of the prices that emerged from the auction.

Although the bids and auction prices in the different conditions did not converge to a common value, bids *within* each group

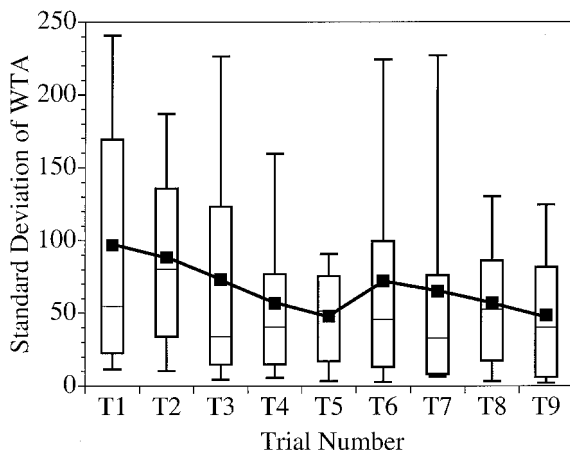


FIGURE VI

The within-Group Standard Deviations of the Bids (WTA), Plotted as a Function of Trial

did converge toward that group’s arbitrary value. Figure VI, 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 4 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.

VII. EXPERIMENT 5: 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 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 nonnormative influences. To test this, Experiment 5 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. On the other hand, if 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 from the Massachusetts Institute of Technology participated in the experiment, which lasted about 25 minutes. 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 seconds. Second, and most important, in each of the three trials subjects were introduced to a new sound with different characteristics: a constant high-pitched sound (the same as in Experiment 2), a fluctuating high-pitched 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 seconds 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 seconds of that sound (WTA). The same Becker-DeGroot-Marschak [1963] procedure used in Experiment 2 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

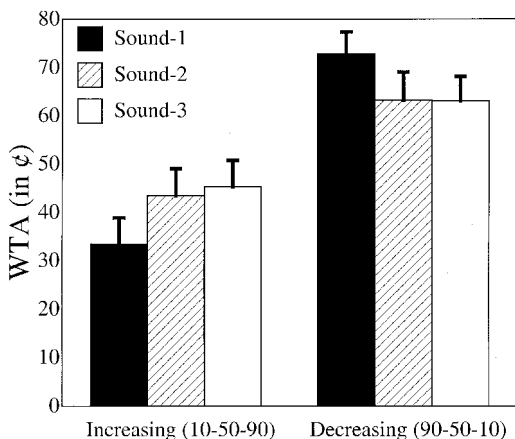


FIGURE VII

Mean WTA (in Cents) for the Three Annoying Sounds

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.

finger in a vise (see Ariely [1998]). The experimenter closed the vise slowly until the subject indicated that he/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 seconds or the vise for 30 seconds.

Results. Figure VII 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 [$M_s = 63.1¢$, and $45.3¢$; $F(1,42) = 5.82$, $p < 0.03$]. Thus, the initial anchor has a stronger effect on 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 5 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¢; $p > .20$ 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¢, $p < .001$ by χ^2 test). Subjects who were in the increasing-anchor condition were much more willing to listen to the sound, compared with 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.

VIII. EXPERIMENT 6: MONEY ONLY?

The previous experiments demonstrated arbitrariness in money valuations. Neither of the follow-up studies (in Experiments 3 and 5), 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 rather that the coherent arbitrariness phenomenon arises only with a relatively abstract response dimension, like money. To address this issue, we conducted an experiment that employed a design similar to that of Experiments 2–4 but which did not involve money. Because Experiments 2–4 had all demonstrated coherence on the dimension of duration, in Experiment 6 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 2–4). They were then shown three containers of different sizes (1 oz., 2 oz., 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 one minute for half the subjects and three minutes 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 oz., 2 oz., 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 of duration equal to 10 seconds, 20 seconds, 30 seconds, etc. up to eight minutes. (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

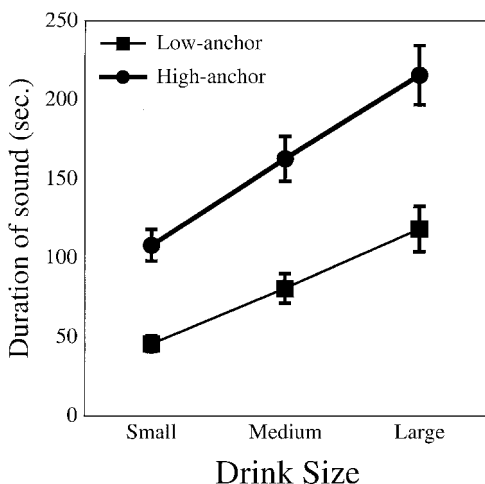


FIGURE VIII

Mean Maximum Duration at Which Subjects Prefer Tone to Drink
Error bars are based on standard errors.

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.

Results. Revealing arbitrariness with respect to tone duration, the anchoring manipulation had a significant impact on trade-offs between the sound’s duration and drink quantity [$F(1,57) = 24.7, p < .0001$]. The mean maximum tone duration at which subjects preferred the tone to the drink (averaging over the three drink sizes) was 82 seconds in the one minute anchor condition, and 162 seconds in the three minute 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 [$F(2,114) = 90.4, p < .0001$] (see Figure VIII).

The experiment demonstrates that arbitrariness is not limited to monetary valuations (and, less importantly, that coherence is not an inherent property of duration).⁷ In combination

7. In a discussion of the arbitrary nature of judgments in “contingent valuation” research, Kahneman, Schkade, and Sunstein [1998] and Kahneman, Ritov, and Schkade [1999] point out a similarity to some classical results in psycho-

with the results of the add-on components of Experiments 3 and 5, it suggests that the web of consistency that people draw from their own choices may be narrow. Thus, for example, a subject in our first 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.

IX. GENERAL DISCUSSION

The main experiments presented here (Experiments 2–4) 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." Experiment 1 demonstrated the pattern with familiar consumer products, and Experiment 6 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.

physical scaling of sensory magnitude. The well-known "ratio scaling" procedure [Stevens 1975] asks subjects to assign positive numbers to physical stimuli (e.g., tones of different loudness) in such a manner that the ratio of numbers matches the ratio of subjectively perceived magnitudes. Sometimes the subjects are told that a reference tone has a certain numerical value (e.g., 100) which is also called "the modulus," while in other procedural variants, subjects have no reference tone and are left to assign numbers as they please. In the latter case, one finds typically that the absolute numbers assigned to a given physical stimulus have little significance (are extremely variable across subjects), but the ratios of numbers are relatively stable (across subjects). Kahneman et al. [1998, 1999] point out that the money scale used in WTP is formally an unbounded ratio scale, like the number scale used in psychophysics, and hence should inherit the same combination of arbitrary absolute but stable relative levels. However, unlike the psychophysical setting in which the response modulus is truly arbitrary (the subjects do not come to the experiment knowing what a 100-point loudness level is), the WTP response scale is not at all arbitrary. Subjects should know what a dollar is worth in terms of other small pleasures and conveniences. If we had asked subjects to evaluate the sounds in terms of uninterpreted "points" rather than dollars, then we would have duplicated the psychophysical procedure of scaling without a modulus, but in that case, of course, the results would be predictable and uninteresting. In any case, the results of Experiment 6 show that exactly the same pattern of coherent arbitrariness can be obtained with well-defined attributes such as duration and drink size.

Our main focus up to this point was to demonstrate the coherent arbitrariness phenomenon, and test whether it is reduced or eliminated by repeated experience, market forces, or higher stakes. Next, we discuss a variety of other phenomena that may be interpreted as manifestations of coherent arbitrariness.

X. A. CONTINGENT VALUATION

The clearest analogy to our research comes from research on contingent valuation, in which people indicate the most they would be willing 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 willingness to pay to clean one polluted lake in Ontario was statistically indistinguishable from willingness to pay to clean all polluted 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, Ritov, and Schkade [1999] and Kahneman, Schkade, and Sunstein [1998]).

This effect 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, applesauce, 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, which explicitly manipulated quantity. This held true even for goods such as toilet paper, for which the meaning of the quantity description (number of rolls) should have been easy to evaluate. Frederick and Fischhoff [p. 116] suggest that this would be a common finding for valuation studies generally—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." In fact, the tendency for within-subject manipulations to produce larger effects than between subject manipulations is a common phenomenon (e.g., Fox and Tversky [1995], Kahneman and Ritov [1994], and Keren and Raaijmakers [1988]).

XI. B. 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" 6,000 today? Or 5,000 or 7,000? or 2,000 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" [p. 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.

XII. C. 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, p. 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 nonunion 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 trade-off 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 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 interindustry wage differentials (see Dickens and Katz [1987], Krueger and Summers [1988], and 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.

XIII. D. 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, Kahneman, Schkade, and Ritov [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 sizable 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 trade-off between the gains from crime and the costs of punishment.

XIV. E. FINAL COMMENTS

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, [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 [page 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, e.g., 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 nonvacuous 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.

SLOAN SCHOOL OF MANAGEMENT, MASSACHUSETTS INSTITUTE OF TECHNOLOGY
 DEPARTMENT OF SOCIAL AND DECISION SCIENCES, CARNEGIE MELLON UNIVERSITY
 SLOAN SCHOOL OF MANAGEMENT, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

- Ariely, Dan, "Combining Experiences over Time: The Effects of Duration, Intensity Changes and On-Papers Line Measurements on Retrospective Pain Evaluations," *Journal of Behavioral Decision Making*, XI (1998), 19–45.
- Ariely, Dan, and George Loewenstein, "The Importance of Duration in Ratings of, and Choices between, Sequences of Outcomes," *Journal of Experimental Psychology: General*, CXXIX (2000), 508–523.
- Ariely, Dan, and G. Zauberger, "On the Making of an Experience: The Effects of Breaking and Combining Experiences on their Overall Evaluation," *Journal of Behavioral Decision Making*, XIII (2000), 219–232.
- Ariely, Dan, George Loewenstein, and Drazen Prelec, "Determinants of Anchoring Effects," Working Paper, 2002.
- Becker, Gary, "Irrational Behavior and Economic Theory," *Journal of Political Economy*, LXX (1962), 1–13.
- Becker, Gary M., Morris H. DeGroot, and Jacob Marschak, "An Experimental Study of Some Stochastic Models for Wagers," *Behavioral Science*, VIII (1963), 199–202.
- Bewley, Truman F., "Why Not Cut Pay?" *European Economic Review*, XLII (1998), 459–490.
- Camerer, Colin, George Loewenstein, and Martin Weber, "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, XCVII (1989), 1232–1254.
- Chapman, Gretchen B., and Eric J. Johnson, "Anchoring, Activation, and the Construction of Values," *Organizational Behavior and Human Decision Processes*, LXXIX (1999), 115–153.
- Cox, James C., and David M. Grether, "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives," *Economic Theory*, VII (1996), 381–405.
- Dickens, William T., and Lawrence F. Katz, "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, 1987).
- Drolet, Aimee L., Itamar Simonson, and Amos Tversky, "Indifference Curves That Travel with the Choice Set," *Marketing Letters*, XI (2000), 199–209.
- Epley, Nicholas, and Thomas Gilovich, "Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors," *Psychological Science*, XII (2001), 391–396.

- Fox, Craig R., and Amos Tversky, "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics*, CX (1995), 585-603.
- Frederick, Shane, and Baruch Fischhoff, "Scope (in)sensitivity in Elicited Valuations," *Risk Decision and Policy*, III (1998), 109-123.
- Gilboa, Itzhak, and David Schmeidler, "Case-Based Decision Theory," *Quarterly Journal of Economics*, CX (1995), 605-639.
- Green, Donald, Karen E. Jacowitz, Daniel Kahneman, and Daniel McFadden, "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," *Resources and Energy Economics*, XX (1998), 85-116.
- Hoeffler, Steve, and Dan Ariely, "Constructing Stable Preferences: A Look into Dimensions of Experience and their Impact on Preference Stability," *Journal of Consumer Psychology*, XI (1999), 113-139.
- Hsee, Christopher K., George Loewenstein, Sally Blount, and Max H. Bazerman, "Preference Reversals between Joint and Separate Evaluations of Options: A Theoretical Analysis," *Psychological Bulletin*, CXXV (1999), 576-590.
- Jacowitz, Karen E., and Daniel Kahneman, "Measures of Anchoring in Estimation Tasks," *Personality and Social Psychology Bulletin*, XXI (1995), 1161-1166.
- Johnson, Eric J., and David A. Schkade, "Bias in Utility Assessments: Further Evidence and Explanations," *Management Science*, XXXV (1989), 406-424.
- Kahneman, Daniel, and Jack Knetsch, "Valuing Public Goods: The Purchase of Moral Satisfaction," *Journal of Environmental Economics and Management*, XXII (1992), 57-70.
- Kahneman, Daniel, and Jack Knetsch, "Anchoring or Shallow Inferences: The Effect of Format," unpublished manuscript, University of California, Berkeley, 1993.
- Kahneman, Daniel, and Dale T. Miller, "Norm Theory: Comparing Reality to its Alternatives," *Psychological Review*, XCIII (1986), 136-153.
- Kahneman, Daniel, and Ilana Ritov, "Determinants of Stated Willingness to Pay for Public Goods. A Study in the Headline Method," *Journal of Risk and Uncertainty*, IX (1994), 5-38.
- Kahneman, Daniel, Ilana Ritov, and David Schkade, "Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues," *Journal of Risk and Uncertainty*, XIX (1999), 220-242.
- Kahneman, Daniel, David A. Schkade, and Cass R. Sunstein, "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," *Journal of Risk and Uncertainty*, XVI (1998), 49-86.
- Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin, "Back to Bentham? Explorations of Experienced Utility," *Quarterly Journal of Economics*, CXII (1997), 375-405.
- Keren, Gideon, and Jeroen B. Raaijmakers, "On Between-Subjects Versus Within-Subjects Comparisons in Testing Utility Theory," *Organizational-Behavior-and-Human-Decision-Processes*, IV (1988), 233-247.
- Krueger, Alan B., and Lawrence H. Summers, "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, LVI (1988), 259-293.
- Mussweiler, Thomas, and Fritz Strack, "Considering the Impossible: Explaining the Effects of Implausible Anchors," *Social Cognition*, XIX (2001), 145-160.
- Payne, John W., James R. Bettman, and Erik J. Johnson, *The Adaptive Decision Maker* (New York: Cambridge University Press, 1993).
- Rae, John, *The Sociological Theory of Capital* (London: Macmillan, (1834), 1905).
- Ross, H. Laurence, "Law, Science, and Accidents: The British Road Safety Act of 1967," *Journal of Legal Studies*, II (1973), 1-78.
- Russo, J. Edward, and France Leclerc, "Characteristics of Successful Product Information Programs," *Journal of Social Issues*, XLVII (1991), 73-92.
- Sen, Amartya Kumar, *Choice, Welfare, and Measurement* (Cambridge, MA: MIT Press, 1982).
- Sherman, Lawrence, "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, 1990).
- Shiller, Robert J., "Human Behavior and the Efficiency of the Financial System," in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, (Amsterdam: North-Holland, Elsevier: 1998).

- Slovic, Paul, "The Construction of Preferences," *American Psychologist*, L (1995), 364–371.
- Stevens, S. S. *Psychophysics. Introduction to Its Perceptual, Neural, and Social Prospects* (New York, NY: Wiley, 1975).
- Strack, Fritz, and Thomas Mussweiler, "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility," *Journal of Personality and Social Psychology*, LXXIII (1997), 437–446.
- Summers, Lawrence H., "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance*, XLI (1986), 591–602.
- Sunstein, Cass R., Daniel Kahneman, David Schkade, and Ilana Ritov, "Predictably Incoherent Judgments," Working paper, University of Chicago Law School, 2002.
- Thaler, Richard H. "Inter-Industry Wage Differentials," *Journal of Economic Perspectives*, III (1989), 181–193.
- Tversky, Amos, and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, CLXXXV (1974), 1124–1131.