Coherent Arbitrariness:
Duration-sensitive pricing of hedonic stimuli around an arbitrary anchor

Dan Ariely[1], George Loewenstein[2], & Drazen Prelec[1]

1. Sloan School of Management
Massachusetts Institute of Technolo0gy

2. Department of Social and Decision Sciences
Carnegie Mellon University

We thank Shane Frederick, John Lynch, Jim 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, and to Loewenstein from the Integrated Study of the Human Dimensions of Global Change at Carnegie Mellon University (NSF grant # SBR-9521914). The paper was begun while Loewenstein and Prelec were on sabbatical at the Center for Advanced Study in the Behavioral Sciences, which was supported by NSF grant # SBR-960123 to the Center.

Correspondences: Dan Ariely, MIT, 38 Memorial Dr, E56-329 Cambridge, MA. 02142; ariely@mit.edu web.mit.edu/ariely/www.

Coherent Arbitrariness:

Duration-sensitive pricing of hedonic stimuli around an arbitrary anchor

Abstract:

In three experiments, subjects stated their willingness to accept pain – from listening to annoying sounds – in exchange for payment (WTA). Subjects were presented with annoying sounds of different durations, indicated their WTA, and received the sounds and payment that resulted from their prices. At the onset of each experiment subject were asked to listen to the sound. Since the sound was very simple, from that point subjects had full information about the hedonic experience. After the initial exposure subjects were asked to state whether, hypothetically, they would be willing to listen to the noise for 30 seconds for either a large or small payment. Subsequently, their actual WTA was elicited to listen to the noise for different intervals (10, 30 and 60 seconds in the first experiment). WTA values exhibited a pattern that we label "coherent arbitrariness." Suggestive of coherence, prices were systematically related to noise duration. But, suggestive of arbitrariness, prices were powerfully influenced by the arbitrary high/low anchor accompanying the hypothetical question. The first study documented the effect at the individual level, the second in experimental markets, and the third examined more deeply the effect of the initial anchor.

Coherent Arbitrariness:

Duration-sensitive pricing of hedonic stimuli around an arbitrary anchor

Economic theories of valuation generally assume that prices of commodities and assets derive 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 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 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 and other benefits 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 valuation are consistent with theoretical prediction. For example, whether labor supply changes appropriately as a function of wages, whether (compensated) demand curves for commodities are downward sloping, or whether stock prices respond to unexpected changes in earnings. It has often been noted, however, that such "comparative statics" relationships are a necessary but not a sufficient condition for fundamental valuation. 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 the 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 sorting out the pleasures and pains that might be 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 simple pains is surprisingly arbitrary, even under "full information" conditions. However, we also show that consumers' *relative* valuation of simple pains appears orderly, 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."

In three studies we find that when people specify the minimum compensation they require to be exposed again to an unpleasant hedonic stimulus — one they experienced moments earlier. Specifically, subjects listened to 30 seconds of an aversive noise and were then asked whether hypothetically they would be willing to listen to the noise for 30 seconds in exchange for either a large or small payment (50¢ versus 10¢ in the first study). They then proceeded to a series of actual pricing rounds in which they stated their minimum price for listening to such noises of various durations, and listened to the sound if their price was lower than a randomly drawn value. The minimum price was elicited either through an individual incentive-compatible mechanism (Study 1) or via an experimental auction (Study 2). In either case, the stated prices were strongly influenced by the arbitrary initial cue (50¢ versus 10¢) and the impact of the cue was not erased by repeated exposure to the noise (Studies 1, 2, 3) or "market forces" (Study 2). At the same time, however, the prices for each individual were smoothly related to noise duration, such that the coherent demand curves for accepting noise could potentially conceal the absence of fundamental valuation.

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 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.

In what follows, Section 1 reviews literature that is relevant to the issue of fundamental valuation. Section 2 proposes a simple formulation of preferences that exhibit the combination of coherence and arbitrariness. Section 3 presents results from three experiments. Section 4 concludes with a 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, 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 (for recent studies of anchoring, see, e.g., Jacowitz and Kahneman (1995), Strack and Mussweiler (1997), Chapman and Johnson (1999)).

The large literature on anchoring effects contains several studies that demonstrate the impact of the anchor value on preferences. Johnson and Schkade (1989) have shown that certainty equivalents for gambles are affected by prior anchoring questions (i.e., whether the subject's certainty equivalent is above or below the anchor). Higher anchor values lead subsequently to higher stated certainty equivalents. Kahneman and Knetsch (1993), and Green, Jacowitz, Kahneman, and McFadden (1998) found the same effect with judgments of willingness-to-pay for public goods; higher values in the initial Yes/No question led to higher the subsequent willingness-to-pay.

A second relevant line of research deals with preference reversals between joint evaluation (in which several options are presented simultaneously and evaluated comparatively) and separate evaluation (in which options are presented in isolation and evaluated separately). Kahneman, Ritov and Schkade (1999) provide an example of this in the context of punitive damages. When two cases were considered evaluated separately by different groups, respondents assessed greater damages for a business fraud case than for a product-related injury to child; however, when the same respondents viewed both cases, most of them assessed greater damages to the personal injury case. Hsee et al. (1999) review a large number of documented preference reversals of this type and show that they can be explained on the basis of a general principle that they call “evaluability.” The evaluability principle posits 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 a well-established standards. Evaluability is related to coherent arbitrariness on three grounds. First, it shows that absolute judgments are difficult – thus supporting the idea that judgments are not based on fundamental values. Second, it shows that relative judgments are simpler and more natural than absolute judgments, thus supporting the idea that relative judgments can appear coherent even if they do not reflect fundamental values. Finally, Evaluability also points to the difficulty of trying to examine the existence of fundamental values with relative methods.

Also related are a series of phenomena that are collectively referred to as “context effects.” Context effects occur when the attractiveness of options are unreasonably influenced by what other options are made available. One of the most important examples is the “compromise effect” in which people choose the intermediate or “compromise” alternatives across a series of three-alternative choices (Simonson & Tversky, 1992); their preferred choice, in absolute terms, moves along with the choice set. Notably, this effect persists even with options that can be directly experienced. For example, consumers will exhibit a strong bias toward the middle quality paper towel out of a set of three, even if they are given the opportunity to handle the towels. Many explanations have been offered for these effects, but, regardless of their underlying cause, they reinforce the conclusion that people have an imperfect understanding of their own values. The general idea that people are uncertain of their values and use available information to construct them “on the fly” goes under the label of “constructive preferences” (Payne, Bettman & Johnson, 1992).

2. Coherent arbitrariness

The work we have just described suggests that people might have very stable relative preferences but unstable or altogether nonexistent absolute (or “fundamental”) preferences. In this section we will indicate how standard utility theory, which assumes fundamental valuation, could be modified to accommodate this. We use the formal framework of an incomplete preference order, which means that there exist pairs of 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, $x = (x_1,..,x_n)$ and $y = (y_1, .., y_n)$. 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 $x \succ y$, if and only if the utility of x is strictly greater than the utility of y, $u(x) > u(y)$. 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, $u(x_1,..,x_n,\alpha_1,..,\alpha_n)$). 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]

$$x \succ_A y \text{ iff } \underset{a \in A}{Min} \sum_i \left(a_i u_i(x_i) - a_i u_i(y_i) \right) > 0.$$

The preference relation, $\succ_A$, 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 will 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 then 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 very simple case, suppose that an identical choice presents itself. We predict that the foundational choice will simply be repeated, because from the consumer's perspective this particular "choice problem" has been solved. Even if the new choice differs in some respects, however, the consumer will use whatever available precedents exist (Gilboa & 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 which are consistent with preferences "revealed" by that initial choice.

Specifically, we assume that the person follows the following choice process:

[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. The α-weights in our model correspond to subjective probabilities in Bewley's representation.

Step 1 If $x \succ_A y$, choose x (and if $y \succ_A x$, choose y).

Step 2 If neither $x \succ_A y$ nor $y \succ_A x$, 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 $x \succ_B y$ (if Step 2 yields a choice of y, then restrict A to largest subset C such that $y \succ_C x$).

Figure 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 Panel A, we see that instead of one, there are now two 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. There is free play in the trade-off between the first and the second attributes.

Panel B shows an option (y) that has 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 arbitrary reason or influence (arbitrary at least from the standpoint of his or her preferences). If option y is chosen over x then, following Step 3, the preference order will be “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 in Panel C in which permissible preferences are designated by the shaded region. The trade-off between attributes is now better specified, and it favors attribute 1 (corresponding to the vertical axis). Even allowing for some residual free-play in the trade-off between attributes, option y is now unambiguously preferred. Panel (D) depicts the individual's preferences after expressing indifference between options x and y. Such a situation might arise if the individual 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 the figure, preferences take the form of a conventional – i.e., “thin” indifference curves.

••• Figure 1 •••

In summary, there are two main implications of this model. The first implication is that there are situations — when preferences are undefined — in which choices will be extremely sensitive to normatively irrelevant influences and considerations (like the anchoring, context, and other effects mentioned in the previous section). As a result, an initial decision will have normatively inappropriate influence over subsequent choices. The second implication is that, if we look at a series of choices in a given domain, they will exhibit an orderly pattern ("coherence") with respect to numerical parameters like price, quantity, quality and so on; the impact of the arbitrary foundational choice is normally concealed.

3. Studies

The three studies we now describe test this model in the context of an experimental marketplace for novel hedonic products – annoying sounds. The sounds have three desirable properties: (1) they permit delivery of many real stimuli to a single subject, (2) the annoyance that they produce shows little or no adaptation over time (Ariely & Zauberman, forthcoming), and, most importantly, (3) the experience has no obvious cognitive component: thinking more about the sound does cannot change its annoyance. The value of avoiding our annoying sounds ranges from a few cents, to at most a few dollars, which places this "product" alongside many other small purchases and decisions that the subjects are likely to be familiar with (e.g., bus fares, vending machine snacks, shirt cleaning, etc.).

In each experiment, subjects set their willingness to accept pain in exchange for payment (WTA) for annoying sounds of different durations. All subjects were initially exposed to the sound, and therefore had full information about the hedonic experience. After this initial exposure, subjects were presented with an anchoring question, specifically, whether — hypothetically — they would be willing to listen to the noise for 30 seconds for either a large or small payment. Their actual WTA was then elicited for different noise durations. The first study documents the effect at the individual level, the second in experimental markets, and the third examines more deeply the effect of the initial anchor.

Experiment 1

Experiment 1 was designed to examine three issues. First, it tested whether seemingly uninformative questions could influence people's WTA for annoying sounds, thus revealing an absence of fundamental valuation. Second, it tested whether responses would be sensitive to a within-subject manipulation of duration, thus demonstrating coherence with respect to this attribute. Finally, it tested whether the arbitrary influence of the anchor diminishes over time with the accumulation of experience.

Method

Subjects: 132 students from Massachusetts Institute of Technology participated in the experiment. Subjects were recruited by advertisements around campus and in the computer lab where the experiment took place. About half the subjects were undergraduate students and the other half were MBA students or, in a few cases, recruiters from investment banks. Subjects were randomly assigned to the different experimental conditions. The study lasted about 25 min and subjects were paid according to their performance (more on this later).

The stimuli: The stimuli were annoying sounds of different durations. The stimuli were generated using a tone generating application (SoundEdit) to create a triangular wave in a frequency of 3,000 Hz. All stimuli sounded like a high pitched scream, similar to the broadcasting warning signal. Stimuli were presented in three durations: 10 seconds, 30 seconds, and 60 seconds.

Design: 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**), no anchor (**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¢)."

After listening to an annoying sound for 30 seconds, these subjects indicated whether hypothetically they would be willing to listen to that sound again for 10¢ (50¢). Subjects in the No-Anchor listened to the sound but were not given any external price and they were not asked to answer any hypothetical question.

In the main part of the experiment, subjects had nine opportunities to listen to sounds in exchange for payment. The three different durations were ordered in two different ways: an increasing set (10 sec, 30 sec, 60 sec.), and a decreasing set (60 sec, 30 sec, 10 sec.). Subjects in the **Increasing** condition were exposed to three of the increasing sets, one after the other. Subjects in the **Decreasing** condition were exposed to three of the Decreasing sets, one after the other. In both Increasing and Decreasing conditions, subjects were exposed to the sequence of three duration sounds three times.

The experiment can thus be thought of as consisting of two between-subject factors: anchor (3 levels) and order (2 levels), and two within-subject factors: duration (3 levels) and replication (3 levels).

Procedure: At the beginning of the experiment, subjects listened to the annoying sound for 30 seconds. Next, the bidding procedure was explained. Subjects were told that they would face real decisions for real money in which they would be asked to indicate the amount they required to listen to the 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 the subject's price, the subject would hear the sound and also receive the payment stated by the computer (the price the computer randomly drew). If the computer's price were lower than the subject's price, they would not hear the sound or receive incremental payment. 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 price distribution was 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.

After learning about the procedure, subjects engaged in a sequence of 9 trials. On each trial they were told that the sound would be the same as the introductory sound they heard but with a duration of X seconds (with X being either 10, 30, or 60). On each trial the computer asked the subjects what was their WTA price for a sound of X duration. Once they entered their WTA, the computer asked them two questions. The first was whether they would accept WTA-5¢ (the price they had set minus 5¢) for listening to the sound and the second was whether they would accept WTA+5¢ (the price they had set plus 5¢) for listening to the sound. 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 again. This was done to ensure that subjects understood the meaning of WTA and that the number truly reflected the price for which they were willing to listen to the sound.

After finalizing a WTA value, subjects were shown their price along with the random price drawn from the distribution. 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 and then continued to the next trial. This process repeated itself 9 times. At the end of the experiment, subjects were paid according to the payment rule, debriefed, and thanked

Results:

Some of the responses (7.7%) were above 100¢, which does not provide a meaningful measure (since the highest random price selected by the computer was 100¢). Therefore for purpose of statistical analysis responses above 100¢ were truncated to be 101¢. We repeated all analyses for both the truncated and untruncated versions and the truncation does not qualify any of the findings. For simplicity, we present only the results that are based on the truncated data.

An ANOVA was carried on the complete design of the experiment. This was a 3 (Anchor) by 2 (Order) by 3 (Durations) by 3 (Replications), mixed (between-subject/within-subject) design. This analysis revealed a main effect for Anchor [$F(2,126) = 10.74$, $p < 0.001$], a main effect for Duration [$F(2,252) = 294.46$, $p < 0.001$], and a main effect for Replication [$F(2,252) = 8.79$, $p < 0.001$]. In addition, there was a Duration by Order interaction [$F(2,252) = 11.2$, $p < 0.001$], and a Replication by Order interaction [$F(2,252) = 5.79$, $p < 0.005$].

To better understand these results, we examined the main effect of the Anchor manipulation on the overall consumption of sounds and their corresponding payments (See

Figure 2). A set of simple effect comparison revealed that average WTA in the Low-Anchor condition ($M = 39.82$) was lower than in the High-Anchor condition ($M = 59.60$), a statistically significant difference [$\Delta M = 19.78$, $F(1,126) = 19.25$, $p < 0.001$]. Average WTA in the No-Anchor condition ($M = 43.87$) was lower than in the High-Anchor condition ($M = 59.60$), again a statistically significant difference [$\Delta M = 15.73$, $F(1,126) = 12.17$, $p < 0.001$]. But WTA in the Low-Anchor condition ($M = 39.82$) was not significantly different from WTA in the No-Anchor condition ($M = 43.87$), ($\Delta M = 4.05$, $p = 0.37$). Because subjects in the High-Anchor condition specified higher WTA, they naturally listened to fewer sounds than subjects in the other two groups [$F(1,126) = 14.26$, $p < 0.001$]. Subjects in the High-Anchor condition listened to an average of 2.8 sounds whereas subjects in the Low-Anchor condition listened to 4.5 and those in the No-Anchor condition listened to 4.1, on average. Subjects in the High-Anchor condition, not surprisingly, also earned significantly less money ($1.53, on average) than those in the No-Anchor condition ($2.06, on average) and the Low-Anchor condition ($2.16, on average) [$F(1,126) = 7.99$, $p < 0.005$].

••• Figure 2 •••

In addition to the main effect of Anchor, we also tested whether WTA changed over time. Averaging over all three Anchor conditions and both duration orders, the average WTA was higher in the first replication ($M = 51.13$) than the second ($M = 46.20$), or third ($M = 45.96$) replications. The difference between the first replication and the other two replications was statistically significant [$F(1,252) = 17.54$, $p < 0.001$], but the difference between replication 2 and replication 3 was not statistically significant [$F(1,252) = 0.29$, $p = 0.86$]. Although WTA decreased after the first replication, there was no evidence of convergence of WTA among the different Anchor conditions. Such convergence would have produced a significant interaction between Repetition and Anchor, but this interaction was not significant. Figure 3 presents the trajectory of WTA over time in the three Anchor conditions.

••• Figure 3 •••

Next, we examine the effects of Duration. 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, 1999; and Kahneman, Wakker, and Sarin, 1997). The mean price for the 10 second sound ($M = 28.35$) was significantly lower than the mean 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 ($M = 66.25$) [$F(1,252) = 126.06$, $p < 0.001$]. Visual support for the consistent effect of Duration can be seen in Figure 4, by observing the level of mean WTAs as a function of sound duration.

Figure 4 provides a graphical illustration of results thus far. First, by observing the vertical displacement between the lines it is again evident that subjects in the different Anchor conditions required different levels of compensation for hearing the annoying sounds. Second, despite this arbitrariness in the overall magnitude of the WTA prices, there is a strong and almost linear relationship between WTA and duration. Finally, as was evident in Figure 3, Figure 4 also shows that no learning took place across the nine trials.

••• Figure 4 •••

Figure 5 provides additional support for the tight connection between WTA and duration. The left-hand panel of the figure presents the ratios of WTA in the 60 second to WTA in the 10 second conditions, separately for each of the three anchor conditions; the middle panel presents these ratios for the 30 second condition relative to the 10 second condition, and the right-hand panel does the same for the 60 second condition relative to the 30 second condition. For each of these ratios, it can be seen, there is a tight relationship between WTA and duration. Moreover, the similarity in the ratios across the different conditions (there are no significant differences across conditions in each of the ratios) suggest that all subjects used a similar consistent response strategy.

••• Figure 5 •••

In summary, experiment 1 demonstrates the phenomenon we label “coherent arbitrariness.” WTA values are highly coherent in the sense of displaying an exquisite sensitivity to the duration of the stimuli. However, the sensitivity of WTA to the anchor shows that these values are to some extent arbitrary. The lack of convergence over the nine trials among the anchor conditions shows that the arbitrariness was not eliminated by experience with the stimuli.

Experiment 2

Although the effect of the initial anchor did not diminish over the nine trials, it is possible that the presence of market forces or information could reduce the degree of initial arbitrariness or facilitate learning over time. Earlier research that compared judgments made by individuals who were isolated or 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 & Weber, 1989).

To test whether market forces would reduce the magnitude of the bias, we exposed subjects to an arbitrary anchor (as in the first experiment), but then elicited WTA values in a multi-person auction rather than with the Becker-Degroot-Marschak procedure. Our conjecture was that the market would not reduce the bias, but would lead to a convergence of within-group WTA values. Consistent with a large literature demonstrating that people adjust their behavior and values to conform to the people around them (see, e.g., Becker, 1991; Bartels, 1985; Akerlof, Yellen & Katz, 1996; Bertrand, Luttmer & Mullainathan, 1998), we expected subjects to use others' values as information in constructing their own values, and hence we expected to observe convergence of bids within specific markets.

Method

Subjects: Fifty three students from Massachusetts Institute of Technology participated in the experiment. Subjects were recruited by advertisements around campus and participated in exchange for a payment of $5 and earnings from the experiment. Subjects participated in the experiment in groups and the groups were randomly assigned to the different experimental conditions. Group size varied from 6 subjects to 8 subjects. The study lasted about 25 min.

Design: Experiment 2 was very similar to Experiment 1, but with groups rather than individuals, different levels of the initial Anchor, and a different bidding procedure. The anchor was manipulated at two levels: 10¢ (Low-Anchor) and $1.00 (High-Anchor). As in the first study, the sound durations were 10, 30, or 60 seconds, subjects were given three opportunities to listen to each of these sounds (Replication), and the order of the durations (Order) was manipulated between subjects. In the Increasing condition durations were presented in the order 10, 30, 60 sec (repeated three times), and in the Decreasing condition the durations were in the order 60, 30, 10 sec (also repeated three times). In sum, the design of the experiment was a 2 (Anchor) by 3 (Durations) by 3 (Replications) by 2 (Order), mixed design. Anchor and Order were manipulated between-groups and Duration and Replication were manipulated within-groups.

Procedure: Subjects participated in groups of six to eight. They 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.

All subjects first experienced, for 30 seconds, the same annoying sound that was used in the first experiment. Next, the bidding procedure was explained to the subjects:

> "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 all be written on the board by the experimenter and the three people with the lowest bids will get the sound they bid for and the money set by the bid of the forth lowest person."

Next the subjects were asked to write on a piece of paper whether a sum of X (10¢ or 100¢ depending on their condition) would (hypothetically) 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 9 trials, the experimenter announced the duration of the sound that was being auctioned, each of the subjects wrote their bid on a piece of paper and passed it to the experimenter. All the bids were than written on a large board and the three lowest bidders were announced. The three lowest bidders were asked to put on their headphones, listen to the sound, and got paid the amount set by the lowest fourth bidder. At the end of the experiment, subjects were paid according to the payment rule.

Results

An ANOVA was carried out on the complete experimental design in which individual subjects' bids were the dependent variable. This was a 2 (Anchor) by 2 (Order) by 3 (Durations) by 3 (Replications) design in which Anchor and Order were between-subject factors and Duration and Replication were within-subject factors. This analysis revealed a main effect for Anchor [$F(1,49) = 20.38$, $p < 0.001$], a main effect for Duration [$F(2,98) = 34.54$, $p < 0.001$], and a Duration by Anchor interaction [$F(2,98) = 12.1$, $p < 0.001$]. Follow up tests showed that the mean WTA in the Low-Anchor condition ($M = 42.91$) was lower than the WTA in the High-Anchor condition ($M = 130.14$), and that this difference was statistically significant [$F(1,49) = 20.38$, $p < 0.001$]. The amounts of money that were earned followed the same pattern. The mean payment per sound in the High-Anchor condition (M = \$.59) was higher than the mean payment in the Low-Anchor condition (M = \$.08), a statistically significant difference [$F(1,49) = 26.24$, $p < 0.001$].

Analysis of the effect of duration showed that WTA for the 10 second sounds ($M = 35.3$) was lower than the WTA for the 30 second sounds ($M = 84$) [$F(1,98) = 15.76$, $p <0.001$], which was lower than the WTA for the 60 second sounds ($M = 137.8$) [$F(1,98) = 18.83$, $p <0.001$]. Finally, the Duration by Anchor interaction showed that the prices for the different durations had a much higher dispersion in the High-Anchor condition than in the Low-Anchor condition. In the Low-Anchor condition the average bids were 23.6, 37.9, and 67.2 for the 10, 30, and 60 seconds respectively. In the High-Anchor condition the average bids were 47.4, 132, and 211 for the 10, 30, and 60 seconds respectively. Note that the ANOVA assumes that effects are additive rather than multiplicative. A straightforward interpretation of the interaction, then, is that subjects adjusted their WTA values according to the ratio of durations.

The next important issue is whether the WTA prices in Experiment 2 showed a pattern of convergence between the low and high anchor conditions. As can be seen from Figure 6 as well as from the lack of Replication by Anchor interaction, there is no evidence of convergence even when the prices in each round are determined by the market forces.

••• Figure 6 •••

At the end of the experiment we asked the subjects to write down the minimum amount for which they would be willing to be exposed to 10, 30, and 60 second sounds. We told subjects that their response on these last measures would not be subjected to the group bidding procedure but instead, would reflect the minimum amount they would be willing to accept in order to be exposed to these sounds. The results for these hypothetical responses also showed a large difference between the two Anchor conditions, with the average responses in the High-Anchor condition ($M = 179.67$) being higher than the average responses in the Low-Anchor condition ($M = 58.21$), and the overall difference between them being statistically significant [$F(2, 64) = 40.85$, $p < 0.001$].

Despite the fact that the bids (and amounts won) in the different conditions (see Figure 6) do not converge to a common value, it is possible that the bids <u>within</u> each group did converge to that group's arbitrary value. To test this hypothesis, we calculated the standard deviation of the responses within each group for each of the 9 trials and examined how these standard deviations changed over time (see Figure 7). Visual inspection of Figure 7 suggest that such convergence did occur. To test whether this convergence was statistically significant, we created a convergence measure for each group. This measure was created by estimating the linear slope within each group of the nine standard deviations. Only one of the eight slopes was positive (0.25) and the rest were negative (ranging from –0.76 to –14.89). A one sample t-test of these eight estimates showed that they were indeed negative [$t(7) = 2.44$, $P < 0.05$].

••• Figure 7 •••

Experiment 3

Experiments 1 and 2 demonstrated that seemingly irrelevant manipulations can have a strong and long-lasting impact on people's valuations of hedonic experiences. With Experiment 3 we was designed to test whether an information-based interpretation could explain the results of the previous experiments. Such an explanation would posit that subjects interpret the anchor as a suggestion or indication of an optimal price.

Experiment 3 exposed subjects to three, instead of one, arbitrary anchor. If the imprinting account that we provided in Section 2 is correct, then the first anchor should have a much greater impact than the later ones. If the information account is correct, and subjects conform to Bayesian updating, then the first anchor should not have any special effect relative to the other two. Bayesian updating predicts that the order in which information arrives will be irrelevant.

Method

Subjects: 44 students from Massachusetts Institute of Technology participated in the experiment. Subjects were recruited by advertisements around campus. Subjects were randomly assigned to the different experimental conditions. The study lasted about 25 min and subjects were paid according to their performance.

The stimuli: The stimuli used were annoying sounds with a duration of 30 seconds, and 3 different characteristics. The characteristics were either constant high pitch (same as in Experiment 1), fluctuating high pitch (which oscillated around the volume of the high pitched sound), and white noise (which is a broad spectrum sound). The important aspect of these sounds is that they were qualitatively different from each other, although they were intended to be roughly similar in terms of aversiveness.

Design: The experiment had two factors. The first factor was the anchor value, which was manipulated at three levels: 10¢ (Low-Anchor), 50¢ (Mid-Anchor), and 90¢ (High-Anchor). This factor was manipulated within subject such that all subjects responded to all three anchor values. Each of the anchor values was coupled with one of the three stimuli (type of sound), and the coupling of the stimuli and external prices was counterbalanced between subjects. The second factor was the Order of the anchors that were presented to subjects. Subjects in the Increasing condition experienced first the Low-Anchor, followed by the Mid-Anchor and the High-Anchor. Subjects in the Decreasing condition experienced first the High-Anchor, followed by the Mid-Anchor and the Low-Anchor.

In sum, the design of the experiment was a 3 (sound-type) by 3 (Anchor) by 2 (Order), mixed design, with Anchor manipulated within-subjects and Order and sound-type manipulated between-subjects.

Procedure: The experiment followed a procedure similar to the one used in Experiment 1. On each of the three trials subjects were introduced to a new sound. After hearing that sound they were asked if hypothetically they would listen to that sound 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 one of the three highly distinct sounds. After answering each hypothetical question, subjects went on to specify the amount they require to listen to 30 seconds of that sound (WTA). Whether subjects heard each sound and how much they were paid for listening, was determined by the same Becker-Degroot-Marschak procedure used in Experiment 1. This procedure repeated itself three times with different sound and a different external price.

At the end of this phase of the experiment, subjects were confronted with a qualitatively different type of choice — whether they would listen to one of the previously experienced annoying sounds or whether they would endure a certain level of physical pain, produced by squeezing their finger with a vise (see Ariely, 1998). The results of this second-stage choice were inconclusive, and we do not report them here.

Results

First, we conducted an overall ANOVA using the level of the Order of Anchor as a between factor (Increasing or Decreasing) and the three repeated measurers of the different Anchor levels as a within factor. As can be seen in Figure 8, the results showed a significant effect for Order, with the subjects who started with the Decreasing Anchor condition (90-50-10) demanding a higher price ($M = 66.4$) compared with the Increasing Anchor condition (10-50-90), who demanded a lower price ($M = 40.8$). Interestingly, the effect of Anchor itself was not significant [$F(2, 84) = 0.51$, $p > .9$], but the Order by Anchor interaction was significant [$F(2, 84) = 6.0$, $p < .005$]. Examining the lambda of both significant effects (a measure of effect size) revealed that, although both effects were clearly significant, the Order had a larger effect (lambda = 16.3) compared with the Order by Anchor interaction (lambda = 11.9).

••• Figure 8 •••

We can also examine WTA separately within each of the order conditions. With respect to the first bid, it is not surprising that the Anchor manipulation had an effect [$F(1,42) = 30.96$, p

< 0.001], causing the first sound in the Increasing condition (with an anchor of 10¢) to be evaluated lower than the first sound in the Decreasing condition (See Figure 8). More interesting is the way subjects reacted to the second bid, which was the same (50¢) for both conditions. In this case, we can see that there was a carry-over effect from the first bid, such that the mean WTA price for the sound in the Increasing condition was lower than the sound in the Decreasing condition [$F(1,42) = 6.03$, $p < 0.02$]. The most interesting comparison, however, is the WTA associated with the third sound. For this sound, the effects of the initial anchor and the specific anchor (the one preceding the specific stimuli) were in opposition to each other. In the Increasing condition, the initial anchor was 10¢ and the specific anchor was 90¢. In the Decreasing condition, the initial anchor was 90¢ and the specific anchor was 10¢. If the specific anchor is stronger than the initial anchor then WTA in the Increasing condition will be higher than the one in the Decreasing condition. If the initial anchor is stronger than the specific anchor, WTA in the Decreasing condition will be bigger than the one in the Increasing condition. In fact, WTA was higher in the Decreasing condition compared with the Increasing condition [$F(1,42) = 5.82$, $p < 0.03$]. Thus, the initial anchor has a stronger effect on WTA than the anchor that immediately preceded the WTA judgment, even though the initial anchor was associated with a qualitatively different sound. At the same time, the third anchor does seem to have an impact, as the WTA advantage of the Decreasing over the Increasing series is greater in the first WTA elicitation than in the third WTA elicitation (testing the interaction: Order x Trial (1 and 3 only), [$F(1,84)=4.46$, $p<.04$].

Looking at the binary responses to the hypothetical questions (the anchoring manipulation), the proportion of subjects who stated that they would be willing to listen to the sound they had just heard for X¢ in the first trial was greater, but not significantly across the two anchor values (55% for 10¢, and 73% for 90¢; $p>.20$ by $\chi 2$ test). This suggests that subjects did not have firm internal values for the sounds before encountering the hypothetical question (the alternative interpretation would be that only 73-55=18% of the subjects had prior WTA values in between 10¢ and 90¢, which is not very plausible). On the third trial, however, the difference was highly significant (41% for 10¢, and 82% for 90¢, $p<.001$ by $\chi 2$ 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. Consistent with the imprinting account proposed in the second section of the paper, over time subjects acquired a stable internal reservation price for the sounds.

4. Discussion

The studies just presented 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 refer to as coherent arbitrariness. Coherent arbitrariness has two central features: coherence, whereby people respond in a sensible fashion to changes in economic variables or to highly salient differences between alternatives, and arbitrariness, whereby these systematic patterns occur around a base-level that is to some extent arbitrary.

As noted in the introduction, our study is by no means the first to demonstrate arbitrariness in valuation. Indeed, Kahneman, Schkade and Sunstein (1998) and Kahneman, Ritov and Schkade (1999) have discussed this phenomenon in the context of contingent valuation (CV) judgments, where subject indicate the most they would be willing to pay (WTP) for a public benefit (e.g., environmental improvement). It is well known that WTP judgments are insensitive to quantitative dimensions of the benefit (e.g., whether a small or a large number of lakes is cleaned up), in between-subjects designs, that is, when different respondents consider the different quantitative scenarios. However, when attention is drawn to the different quantitative possibilities, then individual respondents do show some sensitivity to the quantitative dimension. This is exactly the pattern that we label coherent arbitrariness. The specific empirical contribution of our studies is to show that this type of result applies to actual WTP decisions (rather than hypothetical judgments) for a purely hedonic good. Our stimuli were basic, perceptual, immediate and self-contained. After initial mandatory exposure to the annoying sound, subjects had full information about the experience, and it is hard to see what additional information that override or modify the direct evidence of their senses. If people do have fundamental values at all, then one would expect to detect them in these ideal conditions.

In their discussion of the arbitrary nature of CV judgments, Kahneman et al. (1998, 1999) point out a similarity to some classical results in psychophysical scaling of sensory magnitude. The relevant "ratio scaling" procedure (Stevens, 1975), asks subjects to assign positive numbers to physical stimuli (e.g., tones of different loudness) such 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 the subjects have no reference tone and are left to assigns numbers as they please. In the latter case, one finds typically that the absolute numbers assigned to a given physical stimulus have no great 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. Indeed, in the context of CV judgments, it is plausible that subjects lack a suitable modulus for scaling different public benefits. However, we are hesitant to assimilate our experimental findings to the psychophysical phenomenon of scaling “without a modulus”, because we feel that the task of formulating a WTP for a personal and isolated consumption opportunity, like listening to an annoying tone, is one where the subjects should be able to draw on any number of relevant moduli — specifically, the price-pleasure combinations provided by habitual small purchases (e.g., snacks, transportation, newspapers). It is their failure to draw on these utterly familiar moduli that constitutes the most surprising aspect of our results.

Beyond this, our experiments highlight the general hazards of inferring fundamental valuation from data about responses to change. If all one observed from our experiment was the relationship between valuation and duration, one might plausibly surmise that people were basing their WTA values on the fundamental aversiveness of listening to the sound. 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, the results show that this effect does not diminish as subjects gain more experience with the stimulus or when they provide valuations in a market context.

Because our stimuli provided such a strong test, we believe that our results would generalize to any commodity or experience that is characterized by value uncertainty. Moreover, we would expect to observe similar patterns in many other domains of economic behavior. Indeed, patterns of behavior that are suggestive of coherent arbitrariness can be detected in several lines of existing research. We next elaborate on a few examples that are consistent with coherent arbitrariness.

Financial markets: Like the price one should ask to listen to an aversive tone, the value of a particular stock is inherently ambiguous in that it depends on so many variables that are unknowable. 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 (or asking prices or 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 predictions 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 that show that the market follows a random walk are consistent with the notion of 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.

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 straight-forward. If Microsoft was expected to earn $x in a particular year but instead earned $2x, this is almost unquestionably very good news. If IBM buys back a certain percentage of its own outstanding shares, this has straightforward implications for the values 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.

<u>Labor markets</u>: 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 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 argue, 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 to judge the correctness of their current wage rate is the rate they were paid in the past. As 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.” Workers, in other words, 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, Thaler, 1989; Dickens & Katz, 1987; Krueger & Summers, 1988).

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, understanding the probabilities and likely consequences of apprehension is difficult, as is factoring such calculations into one’s decision making calculus. Thus, this would seem to be a domain that is characterized by value uncertainty where one might expect to observe the coherent arbitrariness pattern. Coherence, 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. We predict that one should find short-term deterrence effects in narrowly focussed studies that examine the impact of policy changes, but little or no deterrence effects in cross-sectional studies. This is, indeed, the observed pattern. Consistent with coherent arbitrariness, 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 trade-off between the gains from crime and the costs of punishment.

<u>Contingent valuation</u>: Contingent valuation is a method for eliciting the value that people place on non-market goods such as environmental amenities. Research on contingent valuation has revealed an "embedding effect" (Kahneman & Knetsch, 1992) such that people's willingness to pay for different environmental amenities is remarkably unresponsive to the scope or scale of the amenity being provided. For example, one study found that willingness to pay to clean one lake in Ontario was statistically indistinguishable from willingness to pay to clean *all* polluted lakes in Ontario. Importantly, these effects have only been observed most consistently in studies that employed a between-subjects design. Few studies have explicitly compared the magnitude sensitivity of both between subject and within subject designs. However, Frederick & Fischhoff (1998) elicited WTPs for two different quantities of common market goods (e.g., toilet paper, applesauce, & 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 (a factor of 2.5) greater 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 be relatively easy to evaluate in an absolute sense. Frederick and Fischhoff (1998, 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 <u>which</u> 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 & Tversky, 1995; Kahneman & Ritov, 1994; Keren, Raaijmakers & Jeroen, 1988).

While we think that coherent arbitrariness is a widespread characteristic of preference we do not claim that it applies universally to all commodities. Some commodities absorb such a large fraction of consumers' expenditures that demand will inevitably respond to absolute price levels even though there is arbitrariness in precise valuation (Becker, 1963). For example, higher gasoline prices *do* produce enduring changes in consumption (smaller, more energy efficient cars) perhaps because fuel costs are a large share of many families' budgets. We suspect, however, that car-buyers would respond more extremely to a recent run-up in gasoline prices than they would to one that took place in the more distant past, even controlling for inflation. Moreover, we would expect a much larger behavioral response to an abrupt increase in gasoline prices than to a gradual increase of the same ultimate magnitude.

Final comments

Our main focus in this paper has been to document the coherent arbitrariness phenomenon and to test whether it is reduced by market interactions. Our studies show that it is a robust phenomenon that is not eliminated by stationary replication in a market. Although we speculated briefly in the introduction about the underpinnings of coherent arbitrariness, our studies were not designed to prove or disprove any particular psychological mechanism. Future studies which do examine underlying mechanisms, will need to account for two components of the phenomenon, namely, the normatively incorrect impact of the initial anchor, and the normatively excessive stability of prices — i.e., the “failure to learn” — over repeated trials.

References

Akerlof, George A, Yellen, Janet L., and Katz, Michael L. "An analysis of out-of-wedlock childbearing in the United States," *Quarterly Journal of Economics*, 1996, *111*, pp. 277-317.

Ariely, Daniel. "Combining Experiences Over Time: The Effects of Duration, Intensity Changes & On- Papers Line Measurements on Retrospective Pain Evaluations." *Journal of Behavioral Decision Making*, 1998, *11*, pp. 19-45.

Ariely, Daniel and Loewenstein, George. "Encoding Sequences and Conveying Information in Ratings and Choice: The Case of Sequences' Characteristics and Duration." Working paper MIT, 1999.

Ariely, Daniel and Zauberman, G. "On the Making of an Experience: The Effects of Breaking and Combining Experiences on their Overall Evaluation." *Journal of Behavioral Decision Making*, 1999 (forthcoming).

Bartels, Larry M. "Expectations and Preferences in Presidential Nominating Campaigns." *American Political Science review*, 1985, pp. 804-815.

Becker, Gary "Irrational behavior and economic theory." *Journal of Political Economy*, 1962, *70*, pp. 1-13.

Becker, Gary "A Note on Restaurant Pricing and Other Examples of Social Influences on Prices," *Journal of Political Economy*, 1991, *99*, pp. 1109-1116.

Becker, Gary M., DeGroot, Morris H. and Marschak, Jacob. "An Experimental Study of Some Stochastic Models for Wagers." *Behavioral Science*, 1963, *8*, 199-202.

Bertrand, Marianne, Luttmer, Erzo, F.P. and Mullainathan, Sendhil. "Network Effects and Welfare Cultures." Unpublished Working Paper, 1998.

Bewley, Truman F. "Why Not Cut Pay?" *European Economic Review*, 1998, *42*, pp. 459-490.

Bewley, Truman F. "Knightian Decision Theory: Part I." Working Paper, Cowles Foundation for Research in Economics at Yale University, New Haven, Connecticut, 1986.

Camerer, Colin, Loewenstein, George and Weber, Martin. "The Curse of Knowledge in Economic Settings: An Experimental Analysis. *Journal of Political Economy*, 1989, *97*, pp. 1232-1254.

Chapman, Gretchen B., and Johnson, Eric J. “Anchoring, Activation, and the Construction of Values.” *Organizational Behavior and Human Decision Processes*, 1999, *79*, pp. 115-153.

Dickens, William T. and Katz Lawrence F. "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.

Fox, Craig R. and Tversky, Amos (1995). "Ambiguity Aversion and Comparative Ignorance." *The Quarterly Journal of Economics*, 1995, pp. 585-603.

Frederick, Shane and Fischhoff, Baruch. "Scope (in)sensitivity in Elicited Valuations." *Risk Decision and Policy*, 1998, *3*(2), pp. 109-123.

Gilboa, Itazhak and Schmeidler, David. "Case-based Decision Theory." *Quarterly Journal of Economics*, 1995, *110*(3), pp. 605-39.

Green, Donald, Jacowitz, Karen E., Kahneman, Daniel, and McFadden, Daniel. “Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods.” *Resources and Energy Economics*, 1998, *20*, pp. 85-116.

Hsee, Christopher K., Loewenstein, George, Blount, Sally, and Bazerman, Max. "Preference Reversals Between Joint and Separate Evaluations of Options: A Theoretical Analysis." *Psychological Bulletin*, 1999, *125*(5), pp. 576-590.

Jacowitz, Karen E. and Kahneman, Daniel. "Measures of Anchoring in Estimation Tasks." *Personality and Social Psychology Bulletin*, 1995, *21*(11), pp. 1161-?.

Johnson, Eric J. and Schkade, David A. “Bias in Utility Assessments: Further Evidence and Explanations.” *Management Science*, 1989, *35*, pp. 406-424.

Kahneman, Daniel and Knetsch, Jack (1992). Valuing Public Goods: The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management*, 1992, *22*, pp. 57-70.

Kahneman, Daniel and Knetsch, Jack. “Anchoring or Shallow Inferences: The Effect of Format.” Unpublished manuscript. University of California, Berkeley, 1993.

Kahneman, Daniel and Ritov, Ilana. "Determinants of Stated Willingness to Pay for Public Goods. A Study in the Headline Method." *Journal of Risk And Uncertainty*, 1994, 9, pp. 5-38.

Kahneman, Daniel, Ritov, Ilana and Schkade, David. "Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues. *Journal of Risk and Uncertainty*, 1999, *19*, pp. 220-242.

Kahneman, Daniel, Schkade, David A. and Sunstein, Cass R. "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages. *Journal of Risk and Uncertainty,* 1998, *16*, pp. 49-86.

Kahneman, Daniel, Wakker, Peter P. and Sarin, Rakesh. "Back to Bentham?: Explorations of experienced utility." *Quarterly Journal of Economics*, 1997, *112*, pp. 375-405.

Keren, Gideon, Raaijmakers, B. and Jeroen, G. "On Between-subjects Versus Within-subjects Comparisons in Testing Utility Theory. *Organizational-Behavior-and-Human-Decision-Processes*. April 1988, *4*1(2), pp. 233-247.

Krueger, Alan B. and Summers Lawrence H. "Efficiency Wages and the Inter-industry Wage Structure." *Econometrica*, 1988, *56*, pp. 259-293.

Payne, John W., Bettman, James R. and Johnson, Eric J. "Behavioral Decision Research: A Constructive Processing Perspective." *Annual Review of Psychology*, 1992, *43*, pp. 87-131.

Ross, H. Laurence (1973). “Law, Science, and Accidents: The British Road Safety Act of 1967.” *Journal of Legal Studies*, 1973, *2*, pp. 1-78.

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, 1990

Shiller, Robert.J. “Human Behavior and the Efficiency of the Financial System,” in John B Taylor and Michael Woodford, eds. *Handbook of Macroeconomics*, 1998.

Simonson, Itamar and Tversky, Amos. "Choice in Context: Tradeoff Contrast and Extremeness Aversion. *Journal of Marketing Research*, 1992, *29*, pp. 281-295.

Stevens, S. S. *Psychophysics. Introduction to Its Perceptual, Neural, and Social Prospects.* Wiley: New York, 1975.

Strack, Fritz and Mussweiler, Thomas. "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility." Journal name? 1997, *73*(3), pp. 437-??

Summers, Lawrence H. “Does the Stock Market Rationally Reflect Fundamental Values?” *Journal of Finance*, 1986, *41*, pp. 5901-602.

Thaler, Richard H. "Inter-industry Wage Differentials." *Journal of Economic Perspectives*, 1989, *3*(2), pp. 181-193.

Tversky, Amos and Kahneman, Daniel. "Judgment Under Uncertainty: Heuristics and Biases." *Science*, 1974, *185*, pp. 1124-1131.

Figure 1: Incomplete preferences, and how they change as a result of an arbitrary initial choice**.**

Panel A: Option Y is preferred to option X.

Panel B: neither option is preferred.

Panel C: Impact on preferences in Panel B of choosing X over Y.

Panel D: Impact on preferences in Panel B of judging X to indifferent to Y

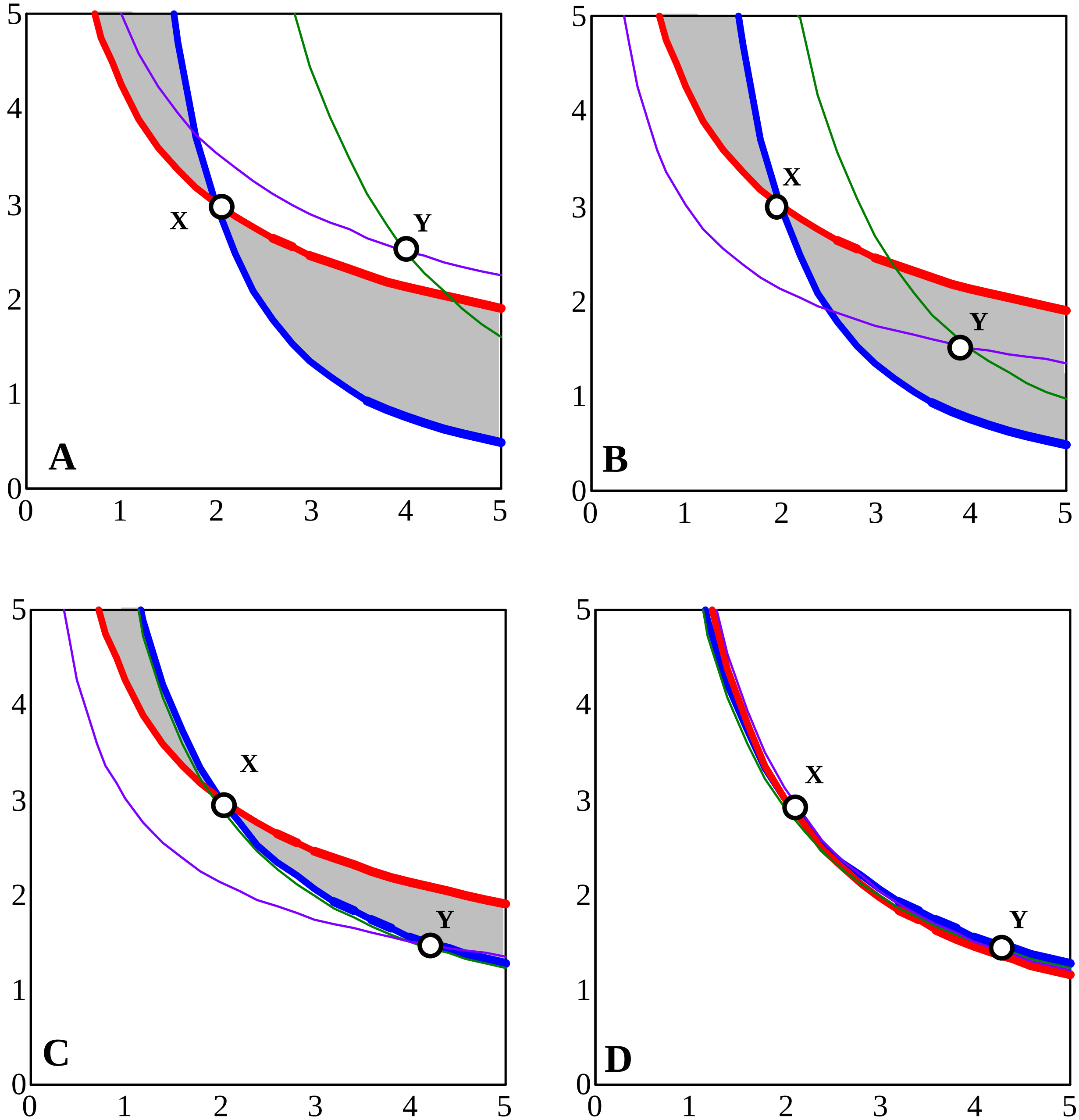

Figure 2: Panel A shows average WTA prices across in the different conditions. Error bars are based on standard errors.

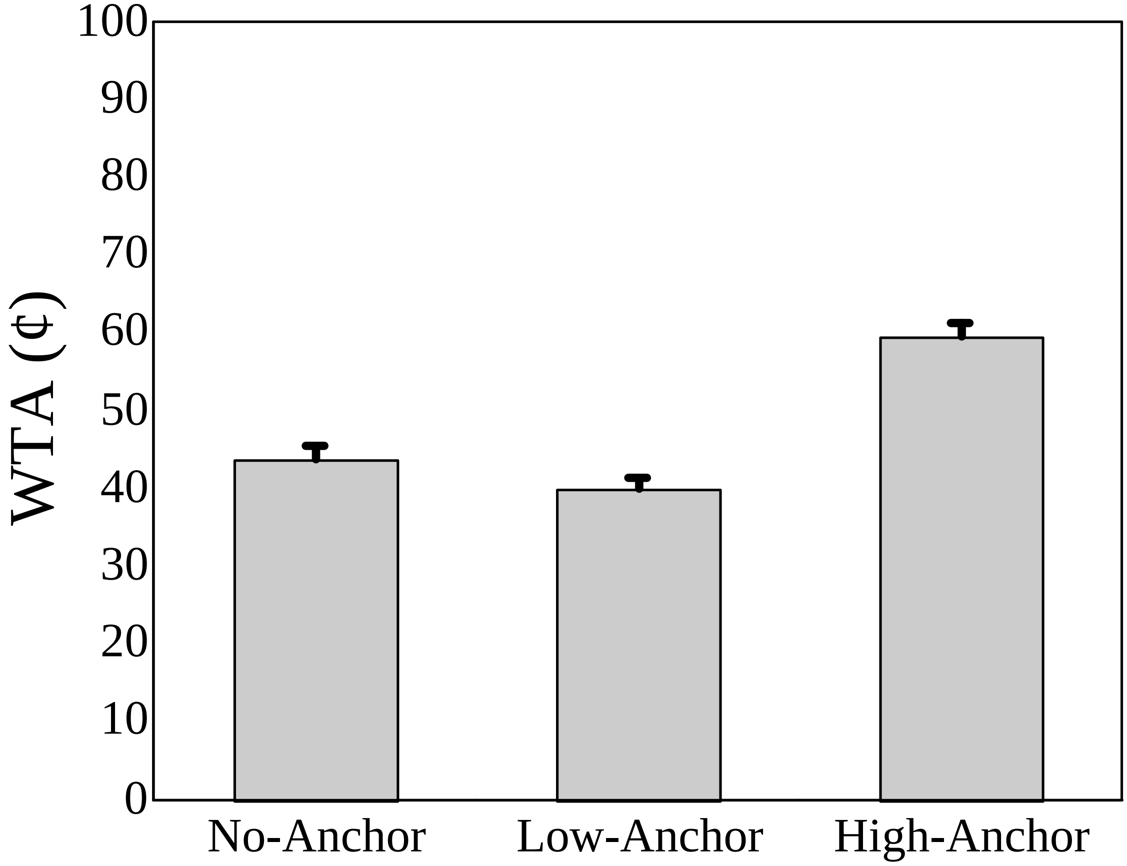

Figure 3: Mean WTA for the nine trials in the three Anchor conditions

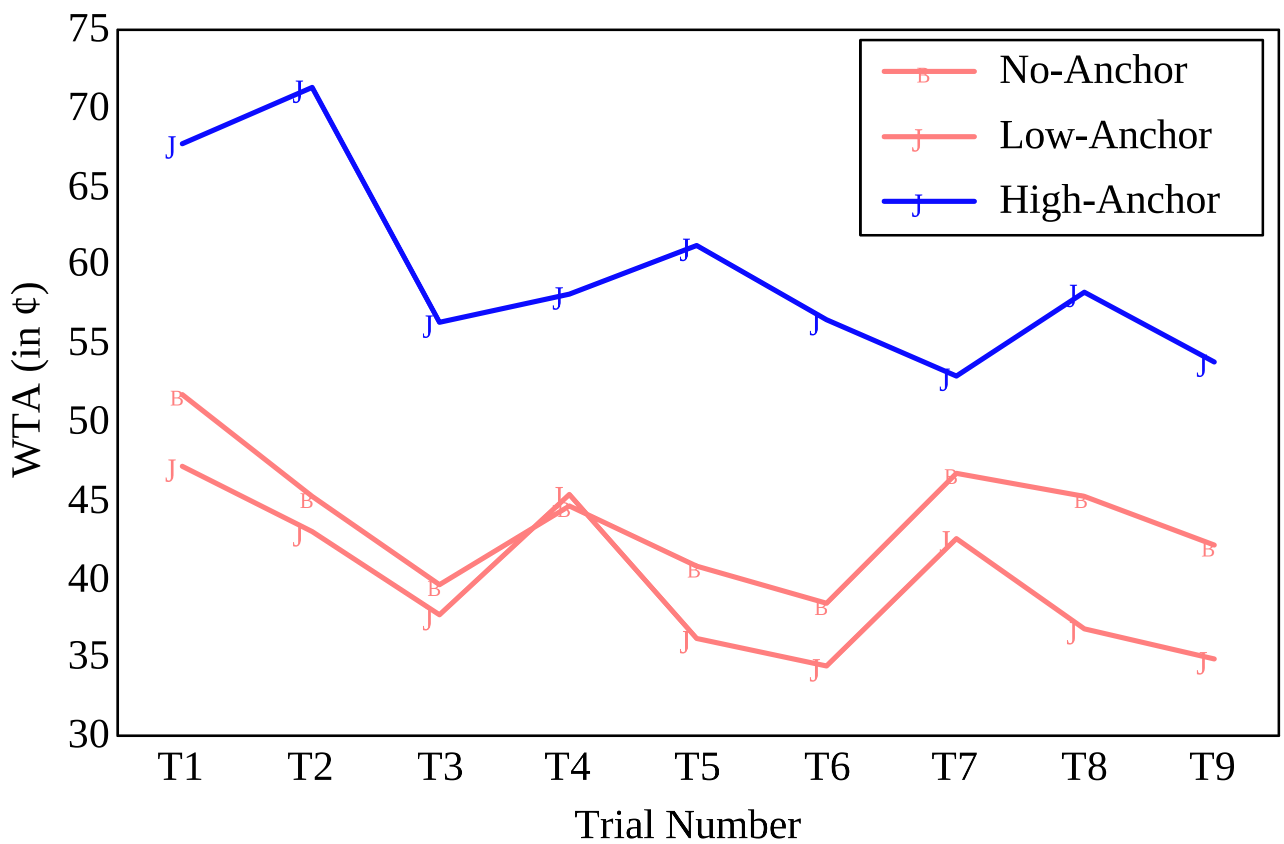

Figure 4: Mean WTA for the nine trials in the three Anchor conditions. The panel on the left shows the trials for the Increasing condition (duration order of 10, 30, and 60 sec.). The panel on the right shows the trials for the Decreasing condition (duration order of 60, 30, and 10 sec.).

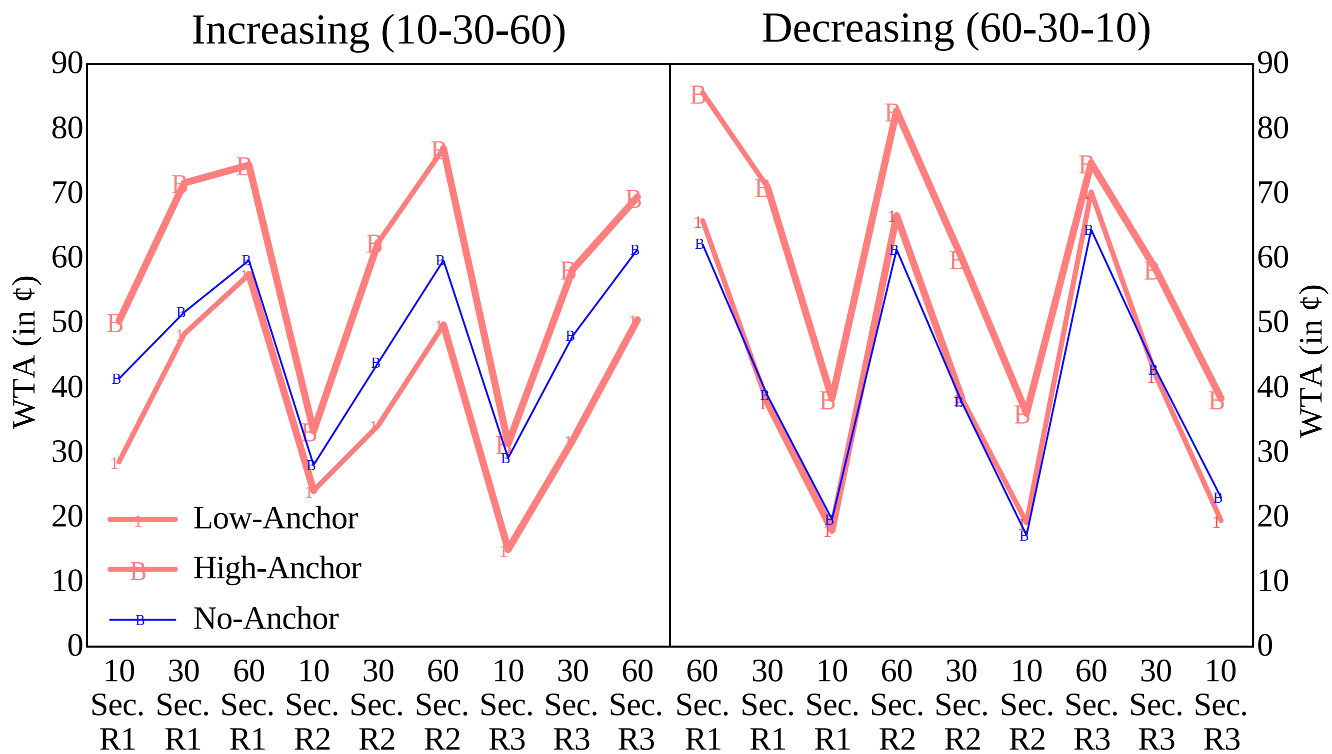

Figure 5: WTA ratios across the different conditions. Error bars are based on standard errors.

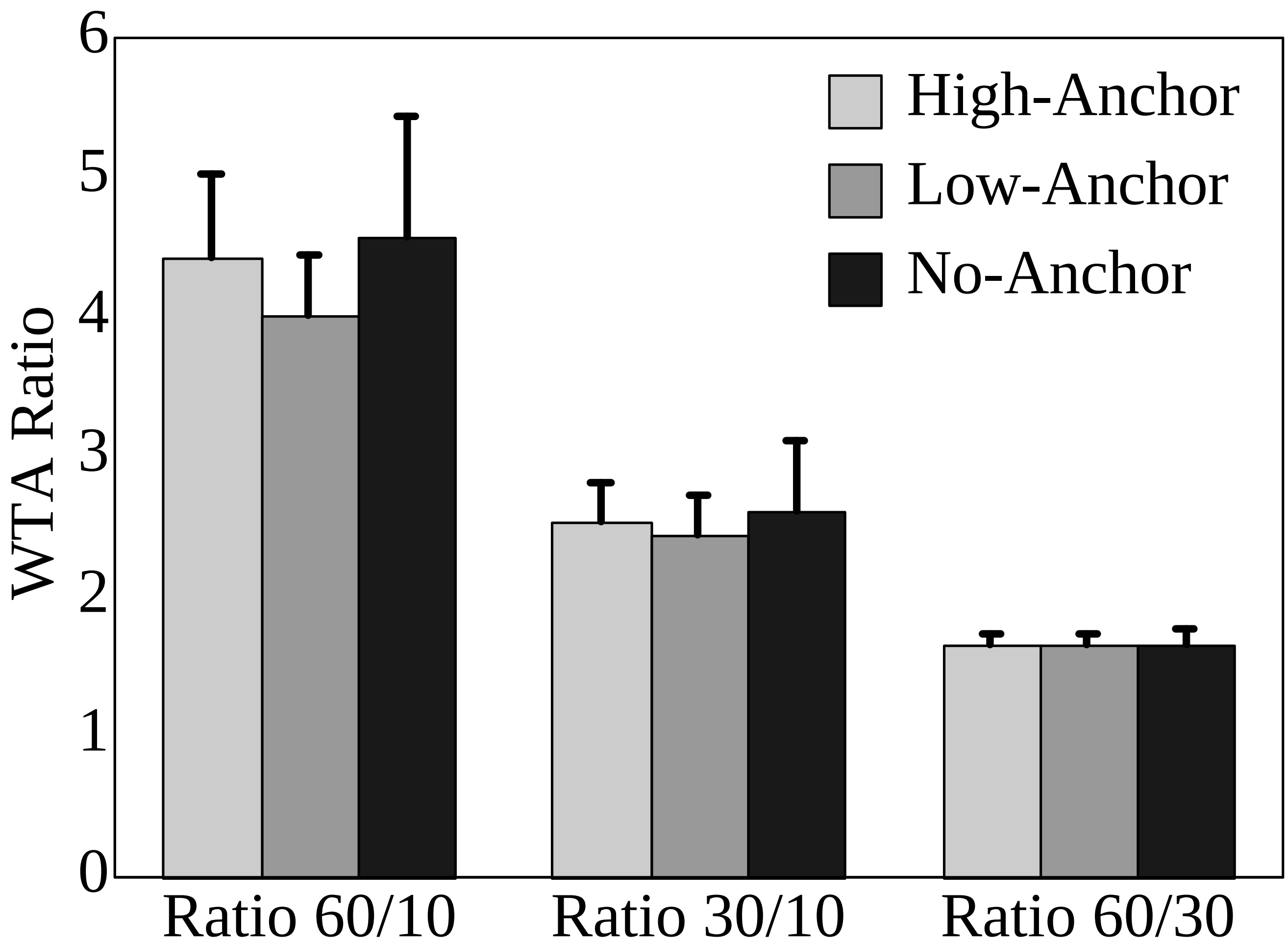

Figure 6: Bids (WTA) and payment as a function of trial and the two Anchor conditions.

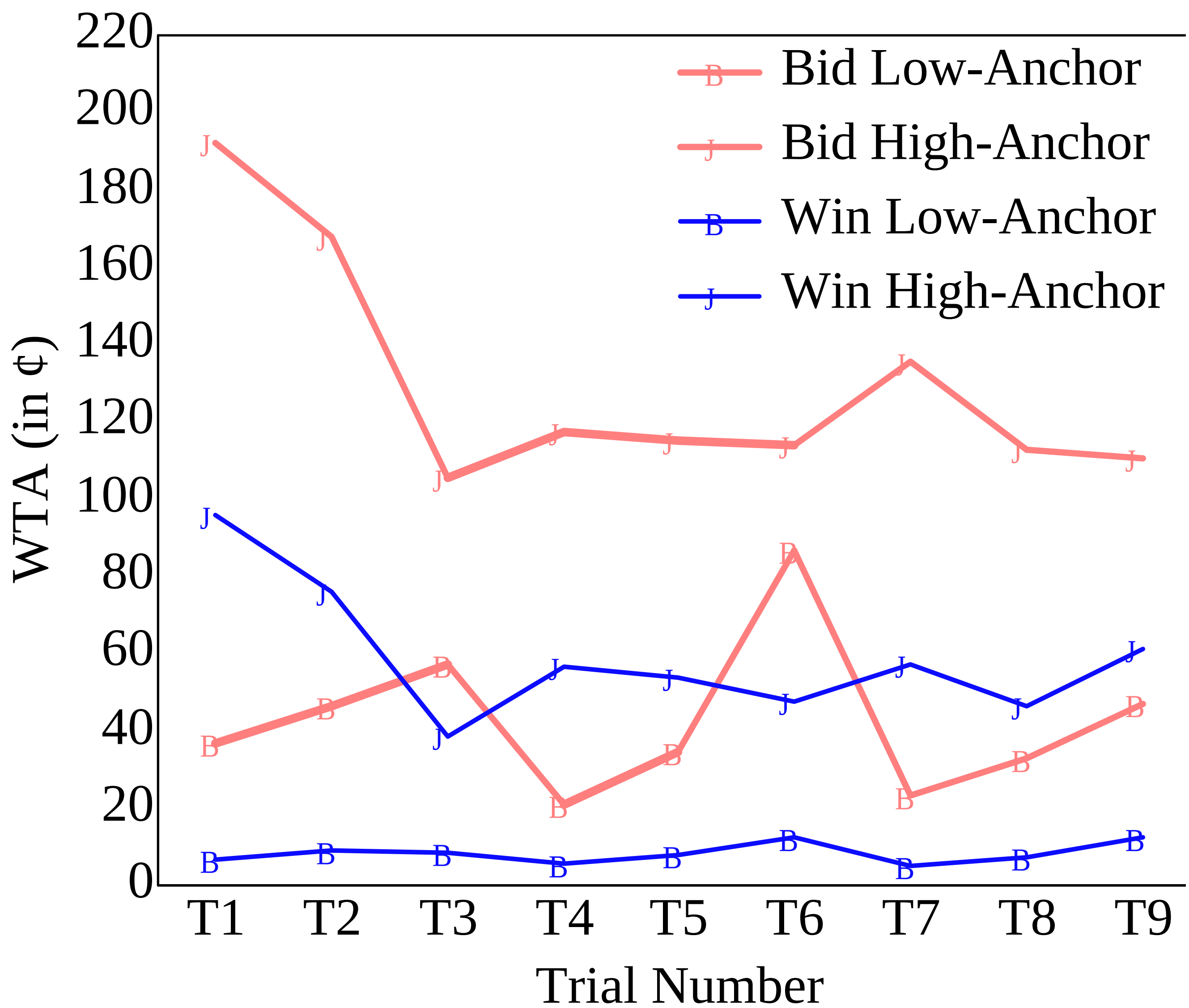

Figure 7: The within-group standard deviations of the bids (WTA), plotted as a function of trial.

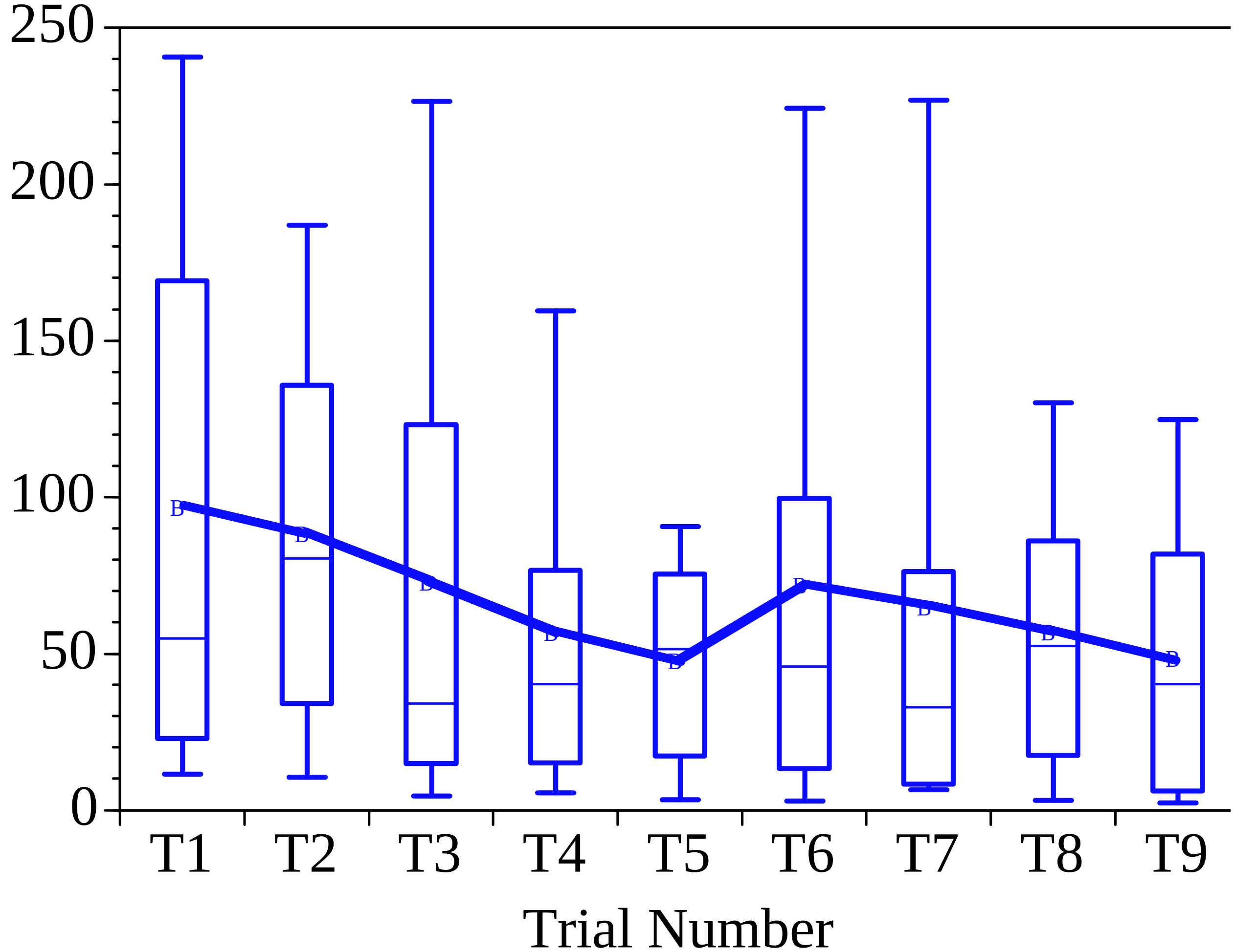

Figure 8: Mean willingness to hear (in cents) for the three annoying sounds. In the Increasing condition the order of the hypothetical questions was 10¢, 50¢, and 90¢ respectively. In the Decreasing condition the order of the hypothetical questions was 90¢, 50¢, and 10¢ respectively. Error bars are based on standard errors.