

# **The Locus and Appropriateness of Monetary Evaluations: Why Monetary Assessments do not Reflect Predicted Utility**

On Amir, Yale  
Dan Ariely, MIT  
Ziv Carmon, Insead

## **Abstract**

We propose that an important cause of anomalies in consumer judgment and decision-making is that consumers' monetary assessments of purchase options may not reflect the predicted utility associated with those options. We suggest that unlike pleasure evaluations that are intuitive processes and are internally focused, monetary evaluations invoke external cues (e.g., payment or market related reasoning). We argue that this different locus of consideration can explain many anomalies relating to monetary judgments and decisions, depending on the extent of the consistency between internal and external cues.

## Long abstract

We propose that an important cause of anomalies in consumer judgment and decision-making is that consumers' monetary assessments of purchase options may not reflect the predicted utility associated with those options. We argue that unlike pleasure evaluations (assessments of predicted utility) that tend to be natural intuitive judgments for which the locus of consideration is internal, monetary evaluations invoke external market related reasoning. Because of this difference in the locus of consideration individuals' monetary judgments tend to correspond to their predicted utility to the degree that the internal and external cues happen to be aligned.

We find consistent support for our ideas in a series of six experiments. In our studies we show that individuals tend to focus on external cues, attending to attributes compatible with money when they assess experiences in monetary terms, more so than when evaluating them in non-monetary terms. We show, for example, that people's monetary assessments are influenced much more by how expensive a particular concert is to produce than by how much they are likely to enjoy attending the concert. We show this pattern to be true of both WTP and WTA, gains and losses, and assessments made regarding the people themselves as well as others, and across different types of goods and services. We also show that if people consider currencies that have hedonic elements, such as dish washing in exchange for admission to a concert, the anomalies tend to disappear. We further show that people indicate immensely dispersed minimal sums they will accept in exchange for performing unfamiliar tasks. This dispersion is only slightly reduced by elaborate and clear descriptions of what the task entails whereas mentioning prices that have little relevance to the task decreases the price dispersion by two orders of magnitude. That is, consistent with our proposed difference in consideration locus, we find an "anchoring-like" effect for monetary judgments, but not for hedonic or pleasure judgments, when using external cues. And consistent with our idea that monetary reactions have an external locus as they are less natural, we show that monetary evaluations of an experience take substantially longer time to provide than a prediction of one's pleasure from it. Finally, we show that manipulating accessibility of internal hedonic cues affects pleasure judgments but not monetary ones.

We conclude that while individuals are well equipped to evaluate experiences at a hedonic level in some situations and could thus make effective choices between experiences, they are not as proficient in translating the experiences into prices. We discuss why many money-related anomalies are due to people's tendency to rely on external cues when making monetary judgments and decisions (even if those are clearly of questionable relevance) instead of mapping internal cues to monetary judgments.

*“Money has no meaning outside ourselves; it is something we have created that has a powerful shaping effect on the world, because it’s something we all subscribe to.”*

*Douglas Adams*

When it became clear to mankind that not every family could easily manufacture all their needs, and specialization was in order, the early concept of trade developed. Trade, an ingenious idea that allowed the shoemaker to eat, the milkman’s horse to be shoed, and farmer to have a plow made life tolerable. Transaction rates were harder to attain, however, because two halves of a cow do not always equal a whole living one. It is then that the revolutionary idea of a common divisible currency was adopted – money. The invention of money entailed inherent benefits: one could actually store money to produce savings, one could smooth consumption over time, one could divide wealth to be used more efficiently, and so forth.

The intention underlying the use of money was to set the price of products (e.g., an egg), either through market supply-demand processes, or through state-paternalistic processes. Once set, the price would serve the following type of decision paradigm: given the price of an egg and the amount of benefit/pleasure I expect to derive from its consumption, how many, if at all, do I want to buy? The underlying assumption in such a process is that the mapping from pleasure/benefit (utility) into the amount of money one is willing to pay for such benefit is unambiguous. We would like to argue that this is not the case.

Money, in itself, is very different from the consumption experience it enables. One does not usually experience money, and thus does not learn and internalize a direct physical benefit that can be associated with it. Individuals do not always know how much benefit they expect to gain from an experience or even whether an experience is good for them or not (Ariely, Lowenstein, and Prelec, 2003). However, there are clear cases in which one can not only associate a certain level of expected pain or pleasure to an experience, but also correctly rate different hedonic levels of utility. In these cases, one

may hope that individuals would be able to correctly use these insights to establish a corresponding willingness to pay, or reference prices. A major caveat for this thesis, however, comes from the very idea behind the invention money as an abstract. No utility, happiness, or hedonic reaction comes directly from money, and thus, there is no internal psychological scale for money, as there are for pain, temperature, light, etc. For example, using a classic psychological anchoring manipulation, Ariely Lowenstein & Prelec (2003) found that participants who were asked for the last two digits of their social-security numbers and following that evaluated a wireless keyboard in dollar amounts, provided amounts that were highly correlated with their SSN endings ( $r = 0.516$ ,  $p < 0.001$ ). Therefore, if people think about an experience differently when they evaluate an experience in terms of its ability to generate pleasure or happiness and when they need to provide a monetary evaluation, we may expect individuals to come up with ill-fitting monetary numbers, even in cases when the experience itself is easy to assess.

Unlike pleasure evaluations, which are internally scaled, and whose representations mostly do not directly resemble external cues, money is related to many market variables and is thus associated with and influenced by ubiquitous external cues. For example, in an effect termed “Money Illusion”, people preferred a smaller real salary to a bigger one, merely because its nominal value was higher – thus, the market state seemed better whereas the actual one was worse (Shafir, Diamond, & Tversky 1997; Hsee, et al. 1999; See also Soman, Wertenbroch, & Chattopadhyay 2003, for similar currency nummericity effects). Using real data and controlling for wealth and other demographics, Simonsohn & Lowenstein (2003) find that individuals relocating to less expensive cities tend to rent or purchase above the average market price, and those moving from less expensive cities to more expensive ones tend to choose below-average living options. This effect was hypothesized to exist because consumers use the information from their past location instead of their new one, and only adjust after some time to the new environment. Notice that the correct information is available all through the process, but the reference is made with respect to the older information.

We thus propose that the different locus of evaluation of pleasure and monetary evaluations causes inconsistencies between these two response modes. While pleasure evaluations are intuitive internal processes, monetary considerations, absent clear internal cues, are not intuitive and invoke external payment and market related reasoning. This premise leads to the following hypotheses:

1. While pleasure evaluations are intuitive internally based processes, monetary evaluations are based on external cues, and lack an internal intuitive component.
2. To the extent that internal and external cues are misaligned, monetary and pleasure evaluations will be inconsistent.
3. Absent external cues, monetary evaluations will be diffuse.

In the rest of the paper we provide experimental evidence supporting our main thesis and the above hypotheses. We conclude with a discussion of the findings and their implications for marketing, markets, and consumption decisions.

Our strategy in teasing apart the influences of internal vs. external cues begins by locating situations in which there is a mismatch between such cues, and measuring pleasure evaluations vs. monetary ones. Similar to the abovementioned reference prices account, the production cost of a product does not always represent its actual value to an individual in a given situation. We thus predict that if individuals relied on the cost information because of its monetary nature, their monetary assessments may diverge from their happiness evaluations. To test the effect of thinking about the value of an experience in terms of money, as opposed to actual personal value – anticipated pleasure, we conducted the following experiment.

### **Experiment 1 - Concert**

The Experiment was designed to test whether preferences for experiences (a concert) are different when they are expressed in terms of money as opposed to anticipated pleasure. We postulated that when preferences are expressed in terms of money, decision makers would be more sensitive to the external monetary aspects of the

experience (e.g., production cost). Conversely, when preferences are expressed in terms of anticipated pleasure, decision makers would be more sensitive to the internal hedonic aspects of the experience. These expectations are akin to the scale compatibility results – where the scale of the response causes decision makers to put more weight on its dimension when making judgments (Tversky, Satath, & Slovic 1988). However, for the scale compatibility results, there is no right or wrong. In contrast, in our case we assume that expected pleasure is the golden standard for preferences and any deviation from it is a violation not only of normative principals such as indifference to response scale but also a violation of preferences (we later test this assumption explicitly).

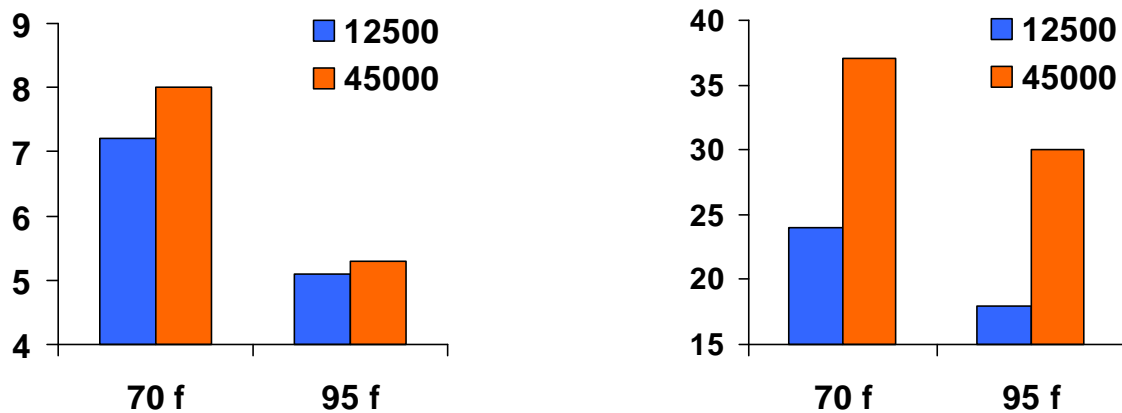
**Method:** Three hundred eighty-eight respondents filled a short survey. The basic design was a cost of production (2) x temperature at the auditorium (2) x response mode (2) between subjects design. The basic scenario described a concert the student association had been planning of a popular singer. The cost of the event to the student association was indicated to be either \$12,500 (low cost) or \$45,000 (high cost), and the expected temperature at the auditorium was either 70 degrees Fahrenheit (low temperature), or 95 degrees Fahrenheit (high temperature). In terms of the response scale, half of the respondents were asked to indicate whether they would pay or not pay an X amount for the ticket. There were 20 amounts in total ranging from \$6 to \$63. The dependent measure we used was the maximum willingness to pay for the ticket. The other half of the respondents were asked to indicate their anticipated pleasure from this performance on a scale from 1 (not very pleasurable) to 10 (very pleasurable).

### **Results & Discussion:**

We first analyzed the monetary responses in a cost of production (2) x temperature at the auditorium (2) ANOVA. The results showed a main effect for production cost [ $F(1,188) = 64.11, p < 0.001$ ], and a main effect for temperature [ $F(1,188) = 19.39, p < 0.001$ ]. As can be seen in Figure 1A (as well as from the F values), production cost had a higher effect on willingness to pay than temperature changes. Next, we analyzed the pleasure responses in a production cost (2) x temperature at the auditorium (2) ANOVA. The results showed a main effect for production cost [ $F(1,192)$

= 4.96,  $p = 0.027$ ], and a main effect for temperature [ $F(1,192) = 119.71$ ,  $p < 0.001$ ]. As can be seen in Figure 1B (as well as from the F values), production cost had a much lower effect on willingness to pay than the temperature changes. It is questionable whether the production cost should have had any effect on our respondents. On one hand, the efficiencies and cost could be irrelevant for the pleasure of the concert, while on the other hand the students could have assumed that a more expensive production could be more pleasurable. Regardless of the magnitude of the production cost, what is clear from the two analyses presented here is that the relative effect of production cost was higher on monetary responses while the auditorium temperatures had more influence on pleasure responses.

•• Figure 1 ••



Since the response scales for the two types of dependent measures were not the same, we next converted all the responses to both measures to Z-scores. An analysis of the Z scores in production cost (2) x temperature at the auditorium (2) x response scale (2) ANOVA, revealed main effects for cost [ $F(1, 380) = 54.66$ ,  $p < 0.001$ ], and for temperature [ $F(1, 380) = 113.92$ ,  $p < 0.001$ ]. More importantly, this analysis revealed an interaction of cost with response scale [ $F(1, 380) = 19.04$ ,  $p < 0.001$ ], and an interaction of temperature with response scale [ $F(1, 380) = 17.71$ ,  $p < 0.001$ ]. The analysis of the Z scores supports the notion that monetary response scales place higher attention on the monetary (perhaps irrelevant) aspects of the transaction, while the pleasure responses

were much more in line with our intuition about behavior and the effects of auditorium temperature on enjoyment.

In sum, the results of Experiment 1 show that monetary response scales cause respondents to focus on aspects that are not the expected pleasure of the experience when making willingness to pay types of decisions. That is, people did not map their internal cues onto to money in the way that maximal WTP responses are usually interpreted. In the case of Experiment 1, participants providing monetary responses seem to have heavily relied on the external information about production cost, but less so on their expected happiness.

Another potential explanation for these results could be scale compatibility, suggesting that when providing monetary responses, dimensions that are closer to money may be overemphasized. That is, information regarding the cost of production will be emphasized when considering monetary reactions. Our current account, however, suggests a deeper explanation. We suggest that people actually consider the problem from a different perspective. We thus need to look for a situation in which the response scales are as compatible with the production cost information, but that one response scale elicits monetary considerations, and one elicits pleasure considerations. Experiment 2 was intended to test this prediction in a setting in which the experience requires investing negative happiness in the form of exerting effort.

### **Experiment 2 – Washing Dishes**

We thus have two goals in Experiment 2. The first was to demonstrate that effort based measures such as willingness to work in exchange for a ticket to the concert would produce the same basic results as the pleasure measure used in Experiment 1 (low effect for production cost and larger effect for temperature). If we were able to show that willingness to exert effort is similar to the pleasure measure, it would be more likely that the monetary scale used is the exception in terms of its ability to reflect the underlying utilities of our respondents. Our second goal was to demonstrate that the dissociation between predicted enjoyment and WTP in Experiment 1 couldn't be explained by a simple scale compatibility account (surface scale compatibility). Scale compatibility can

be conceived in two forms, surface and deep: The surface scale compatibility argument is that the response scale causes decision makers to place greater *weight* on the attributes that fit with this response scale. The deep scale compatibility argument is that the response scale causes decision makers to *think* about the evaluated object from the perspective of the type of attributes that are represented by the response scale – this is the type of scale compatibility we argue for here. In particular, we suggest that the monetary perspective is an external one, while the pleasure perspective is intuitive and internal.

To archive these goals we augmented the basic design of Experiment 1 by changing the dependent measures to capture the amount of effort participants were willing to exert in order to obtain a concert ticket. The effort was measured in two ways: duration of washing dishes or quantity of dishes washed. Because duration of work translated more naturally into hourly wages, very salient to student population, we expected it to be more sensitive to the cost of the performance than the quantity of dishes washed. Conversely, the number of dishes washed can be easily translated to an intuitive internal measure of effort.

**Method:** Three hundred and eighteen respondents filled a short survey. The basic design was the same as in Experiment 1: a cost of production (2) x temperature at the auditorium (2) x response mode (2) between subjects design. The main differences from Experiment 1 were the response scales. Experiment 2 used two response scales: length of time respondents were willing to wash wine glasses in exchange for a ticket (dish-duration), and number of wine glasses they were willing to wash in exchange for a ticket (dish-quantity). Half of the respondents were asked to indicate whether they would be willing to wash wine glasses for X minutes in exchange for a ticket (dish-duration). There were 20 durations in total ranging from 15 minutes to 300 minutes. The dependent measure we used was the maximum duration respondents were willing to wash glasses in exchange for the ticket. The other half of the respondents were asked to indicate whether they would be willing to wash X wine glasses in exchange for a ticket (dish-quantity). There were 20 durations in total ranging from 25 glasses to 500 glasses. The dependent measure we used was the maximum number of glasses respondents were willing to wash in exchange for the ticket. Finally as a manipulation check, after

answering the main questions, respondents rated how pleasurable they expect the concert to be on a scale from 1 (very unpleasurable) to 10 (very pleasurable).

### **Results & Discussion:**

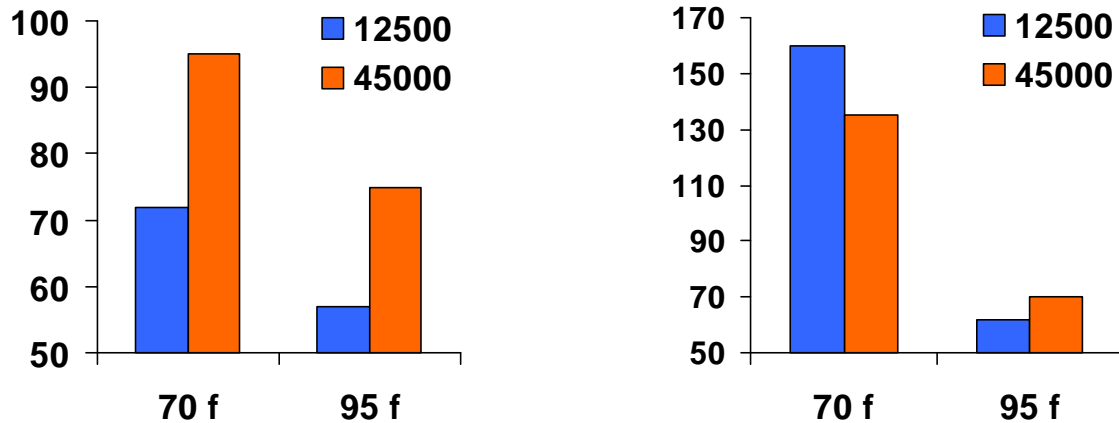
We first wanted to test that the type of dependent measure we used (dish-duration and dish-quantity) did not change the perception of the desirability of the concert. An overall ANOVA revealed a large main effect for temperature [ $F(1,292) = 59.66, p < 0.001$ ], a smaller effect for cost of production [ $F(1,292) = 10.22, p = 0.002$ ], and no significant effects for the type of the depended measure. Thus, showing that the type of dependent measure did not change the way our respondents understood / perceived the concert.

We next analyzed the dish-duration responses in a cost of production (2) x temperature at the auditorium (2) ANOVA. The results showed a main effect for cost of production [ $F(1,132) = 6.89, p = 0.01$ ], and a main effect for temperature [ $F(1,132) = 7.46, p = 0.007$ ]. As can be seen in Figure 2A (as well as from the F values), the magnitude of the two effects (production cost and temperature) was relatively similar. Next, we analyzed the dish-quantity responses in a cost of production (2) x temperature at the auditorium (2) ANOVA. The results showed only a main effect for temperature [ $F(1,178) = 81.19, p < 0.001$ ], and a non significant effect for cost of production [ $F(1,178) = 0.64, p = 0.43$ ]. As can be seen from the comparison of Figure 2A and 2B, while both dependent measures showed sensitivity to temperature, as expected only the dish-duration dependent measure was influenced by production cost.

The results from Experiment 1 suggested that responses on a monetary scale are sensitive to external factors that do not directly influence pleasure (such as production cost). The results of Experiment 2 expand these findings to show that this sensitivity to external factors also exists in cases where the response scale is naturally associated with a monetary sale. In this case we assumed that dish-duration would be naturally converted into money (based on salary per hour which is very salient to students) and does not have a natural intuitive internal representation, but that dish-quantity is less likely to be

converted to money and may be more easily associated with internal associations of effort. As a consequence we expected higher sensitivity for production cost in the dish-duration condition compared with the dish-quantity condition. The results (see Figure 2) confirmed our hypotheses.

••• Figure 2 •••



### Experiment 3 – Fixing your Hard Drive

Experiments 1 and 2 demonstrate that individuals are willing to pay or willing to work, not based on their own internal predictions of enjoyment or effort, but rather based on external, less relevant cues, yet compatible with market payment schemes. The next experiment was meant to test whether individuals will still make biased decisions when the outcome is in the domain of losses, for which there is higher vested interest, and another individual was performing the work. Moreover, by manipulating the order of asking about WTP and about the impact of the experience, we may be able to provide initial evidence for the suggested mechanism: if individuals who first consider the hedonic or personal impact of the experience provide different WTP evaluations than those who directly estimate WTP, we may argue that the latter do not base their estimation on the predicted internal impact (predicted utility). Experiment 3 concerned judgments about payment to a technician for recovering lost data from a defective hard drive. Unlike experiments 1 and 2, in this experiment we introduce an implicit external influence in the form of market fairness. We predict that individuals' maximal WTP

would correspond to this external issue, rather than to their own value of the work, when directly responding to the WTP question, but would correspond to predicted utility when first estimating it and only subsequently providing WTP estimates.

**Method:** Ninety-six respondents filled a short survey. The basic design logic was similar to the previous experiments: 2 (amount of labor: 5 minutes vs. 12 hours) x 2 (size of loss: 1 month of data vs. 5 years of data) x 2 (response mode: WTP vs. importance) x 2 (order: WTP first vs. importance first) between participants design. Experiment 3 used two response scales: amount of money participants were willing to pay for the data recovery (WTP), and rating of the importance level participants assigned to the data recovery (importance), as a 0-100 open response scale.

### Results & Discussion

The main result replicated those of the previous experiments (Table 1): while the amount of labor involved in restoring the data had a main effect on WTP, only the size of the loss had an effect on importance. This resulted in the paradoxical finding that the WTP for a 5 minute restoration of 5 years of data was smaller than the WTP for a 12 hour restoration of 1 month of data! Furthermore, while this anomaly holds when WTP was the first measure, when importance was measured first, the subsequent WTP responses correlate with it. That is, the WTP responses match predicted utility only when the latter was made salient and accessible, otherwise, the WTP responses more closely correspond to the amount of labor (perhaps because of fairness considerations).

•• Table 1 ••

Work Length	Data Lost	WTP First		Importance First	
		WTP	Importance	WTP	Importance
5 minutes	1 month	\$62.31	54.62	\$90.25	69.17
	5 years	\$127.50	71.75	\$114.50	85.50
12 hours	1 month	\$154.58	68.50	\$91.67	65.33
	5 years	\$213.46	74.08	\$184.08	86.25

These results are supported by two ANOVAs, one for each DV, including labor, loss size, and order as independent factors. In the case of importance ratings, the only significant effect was that of the amount of data lost [ $F(1,88) = 5.220$ ,  $p = 0.025$ ]. However, when WTP was estimated, both the amount of data lost and the length of the work time to recover the data had a significant influence [ $F(1,88) = 5.220$ ,  $p = 0.014$ , and  $F(1,88) = 6.716$ ,  $p = 0.011$ , respectively].

The results of Experiment 3 strengthen the results of the previous studies by demonstrating robustness of the phenomena involving monetary evaluations to whether one considers paying or accepting money, and whether one is gaining or preventing losses. More importantly, the results of Experiment 3 provide initial evidence for the proposed theory of locus of cues considered, by demonstrating that unless internal cues related to predicted experience utility are made salient and accessible, monetary evaluation are greatly influenced by external cues and considerations (such as fairness in this particular case). Participants in our experiment actually, perhaps even erroneously, were willing to pay more for a 12 hour restoration work of a small impact loss, than a 5 minute restoration of a very consequential loss!

In sum, the psychology of providing monetary evaluations and the weak mapping between money and effort or money and pleasure may violate the basic principles on which the very idea of a smooth intermediary for trade resides. Great market inefficiencies may be created when individuals do not use internal cues (predicted experience utility) as an input to their decision processes and map them into monetary evaluations. For example, if potential employees do not correctly translate the expected effort of a job into the salary demanded, either unemployment or overpayment and inefficiency may arise. Moreover, any such inefficiency will have a tendency to persist if individuals use existing level of payment as a predominant source of information in determining their own required compensation<sup>1</sup>.

---

<sup>1</sup> The process that is supposed to correct inefficiencies in such cases is the ability of external labor to be hired at more appropriate salaries. However, if such external labor determines appropriateness by observing current compensations, the market will lack a balancing power.

## Experiment 4 – Job Compensation

The experiments thus far showed that there is dissociation between predicted pleasure and willingness to pay. The results showed that when considering money, expected pleasure is not the primary factor that is considered, but other external considerations such as fairness and cost of production have a large influence, even though individuals are able to estimate their own expected pleasure. One component of our argument is the claim that money, unlike pleasure, has virtually no intuitive internal representation. If this is indeed the case, then absent sufficient external cues, monetary evaluation will tend to be extremely diffuse, even in situations where the pain and pleasures are easy to come by (Hypothesis 1). The current experiment was designed to test this aspect of our claim.

The domain in which we will examine this idea is labor, asking for the minimum amount individuals require to be compensated to accomplish a task (WTA). We used two types of tasks: **common-jobs** - tasks that are commonly practiced as paid employment (providing strong external cues); and **uncommon-jobs** - tasks that are only rarely practiced as paid employment and thus have virtually no market price (providing weak external cues). For the common-jobs we expected our participant population to be familiar with both the effort involved and with the expected compensation in the marketplace. For the uncommon-jobs we expected our participant population to be familiar with the effort involved (since the tasks themselves are known and familiar activities), but not with the prevailing compensation in the marketplace. Our hypotheses were that for the common jobs, individuals would state a level of compensation that is similar across individuals and close to the market reference price (i.e., low variance), while for uncommon-jobs, individuals would state a level of compensation that is different across individuals (i.e., high variance). Both of these claims arise from the difficulty of mapping experience to money: when there is an available compatible external cue, it will be used, but when one does not exist, the difficulty will be portrayed by high variance. We also wanted to make sure that the differences across the two types of tasks would not come from participants' uncertainty about what the different tasks entail.

**Method:** In order to examine these two kinds of tasks we first selected a set of tasks that would be equivalent in the level of pleasure / displeasure that they evoke, and also equivalent in the clarity of understanding of what the tasks entail (i.e., uncertainty about the task). A pilot study found eight activities that induced the same level of pleasure / displeasure and were of equivalent certainty level. A separate sample of respondents divided the eight tasks (see Table 2) into common and uncommon jobs. We asked respondents to state the minimum amount that they would be willing to accept in exchange for performing each of the eight tasks (WTA). This resulted in a 2 (common vs. uncommon) condition within participant design.

**Table 2: Standard Deviations of WTA responses**

Task Type	Task	Std Dev
Common	Drive a bus to the SFO airport	46.07
	Walk dog for 30 minutes across campus	20.28
	Deliver newspapers for 30 minutes	15.25
	Shine 3 pairs of shoes	15.56
Uncommon	Kill one mouse	417.99
	Smell 3 shoes	380.15
	Sing songs at a central corner on campus for 30 minutes	362.03
	Paint face blue and walk across campus	272.06

## Results & Discussion

In the main study, 69 respondents stated the lowest level of compensation they would be willing to accept if they were to perform a task. For each of the tasks the standard deviation of the responses was computed and used as the measure of consistency across respondents. We used this variation measure to test the idea that when faced with tasks that have a clear market reference price, individuals will use this price as their own and thus will be homogeneous in their estimate. On the other hand, when faced with tasks that do not have a clear reference price in the market, individuals will be forced to use their own subjective estimation of task difficulty and provide their own price, allowing natural heterogeneity to determine the spread of the results. If participants

would have mapped their predicted pleasure / pain into money, or were they relying on a distinct internal value, we would expect a variance not dissimilar from the common jobs. However, if internal cues for money are weak, and the estimation is thus difficult absent external cues, as we hypothesized, we would expect it to lead to a much higher level of variation across individuals. As can be seen in Table 1, this was the exact pattern of the results. The common jobs had relatively low standard deviations while the uncommon tasks had very high standard deviations [F-values for the differences between variances range from  $F(68,68) = 34.87$  to  $F(68,68) = 751.26$ ,  $p < 0.001$ ]. In fact, even though the overall pleasure of the experiences was pre-tested to be similar (painting one's face vs. walking a dog for example), some people were willing to perform the uncommon jobs for free but no one was willing to perform the common jobs without being compensated.

### **Experiment 5 – Jobs and Information**

If people used external rather than internal information to determine required compensation, they may have still come up with a reasonable number. Specifically, if people were more familiar with the task information they may have generated a better estimate of the amount of effort required to perform the task. On the other hand, we predicted that if individuals do not map internal predicted effort to money, they would use external cues (e.g., reference prices or market values) to come up with their answer. When these cues are salient, the variance will be small as the estimates center around this salient value, but when they are not, the variance will be large. To test whether the increase in variability was driven by a lack of an external cue as we predicted, or whether it was driven by a lack of task relevant information for the uncommon jobs, Experiment 5 was run. Experiment 5 replicated Experiment 4 while directly manipulating the information states participants were in (i.e., amount of task information, and availability of external price cues).

**Method:** One hundred and fourteen individuals participated in an online questionnaire for a chance at winning a CD or a DVD. The experiment was a (2) task information x (2) reference price information x (2) task familiarity design, where the first

two factors were manipulated in a between-participants design, and the latter within participants. We used a different set of tasks in order to increase the generalizability of Experiment 4's findings (see Table 2). The conditions differed on the amount and type of information provided to participants for each task. Half of the participants received a detailed description of the task, while the other half did not, and, independently, half of the participants received information about a remote reference price ("In a related task professionals earn \$xx.xx per hour"), while the other half did not.

After reading instructions, participants were asked for the minimal wage they would be willing to accept (WTA) for performing the task. We emphasized that this amount meant that they would not be willing to perform the task if offered a compensation of \$1 less than their stated amount. After completing this stage, participants were asked to rate the amount of effort each task requires, as well as their familiarity with the task itself. These latter two measures served as manipulation checks.

### **Results & Discussion**

Manipulation checks: Familiarity was measured on a five-point scale. While for the familiar tasks (making coffee and helping elderly people cross the street) the mean familiarity ratings were 2.17 and 3.72 respectively, the means for the unfamiliar tasks (shining shoes and making hamburgers at McDonald's) were -2.04 and -3.75 [smallest  $t(55) = 3.65$ ,  $p < 0.001$ ]. These measures confirm our assumptions about the nature of prior information about the tasks.

Our main hypothesis regarded the influence of different types of information on the variances of responses to the WTA question. The main effect of the role of familiar task as opposed to unfamiliar ones replicated the previous experiment (Table 3): the variances of WTA values in the case of familiar tasks were significantly smaller than those of the unfamiliar tasks [F-values range from  $F(113,113) = 8.69$  to  $F(113,113) = 33.68$ ,  $p < 0.001$ ]. The current experiment adds to the previous findings by revealing that while the effect task specific information on the variances was small, the effect of providing a distant related market value was huge.

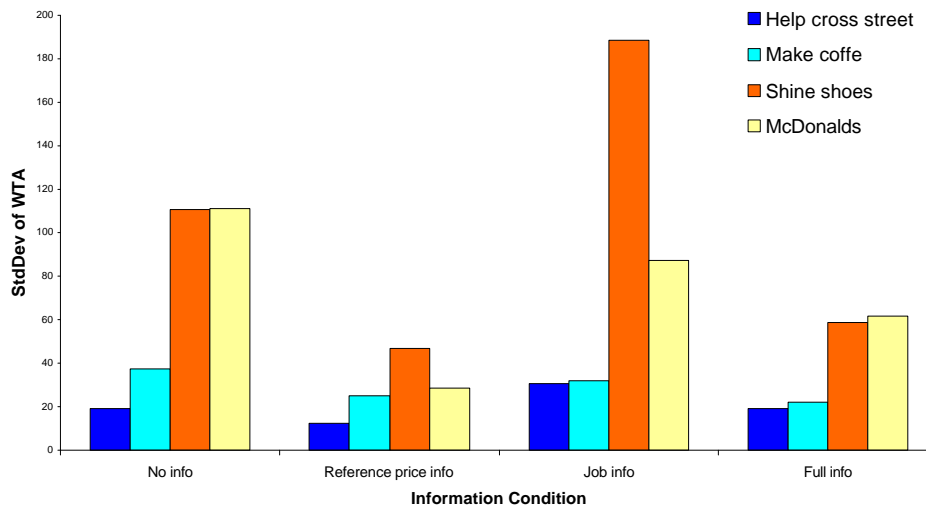
**Table 3 Standard Deviations of WTA for familiar and unfamiliar jobs without added information**

Task Type	Task (all for 3 hours)	Std Dev
Familiar	Make coffee	37.50
	Help the elderly cross the street	19.14
Unfamiliar	Shine shoes next to a business center	110.55
	Make hamburgers at McDonald's	111.08

Analyzing the effects of the two types of information on the variance of WTA responses for familiar and unfamiliar tasks, the following picture emerges. The difference in the variances could be assessed by their ratio, which provides an F-statistic, and because our sample sizes are identical, we can directly compare the size of this statistic as a proxy for the effect size. Compared to the no information condition, adding reference price information had a bigger effect on responses to the unfamiliar tasks than to familiar tasks (10.34 vs. 2.32), and a larger effect than both task information (2.26 vs. 1.97) and full information (3.37 vs. 1.93). Moreover, notice that the relative effect of reference price information on unfamiliar tasks was much larger than any other effect.

These results replicate the results of Experiment 4, showing that familiarity with a task leads to a much narrower distribution of WTA responses. Experiment 5 expands this result by showing that it is the knowledge of some external reference price that is responsible for the major decrease in variance, while the task-specific information which could improve the internal representation of the effort involved does not (Figure 3). The results suggest that individuals do not map their internal representations of experiences onto monetary sums (hence the high variance) and rely on external market cues to respond to such questions. Our results demonstrate that participants follow such cues to a much greater extent than informational cues about the tasks themselves.

**Figure 3 Standard Deviations of familiar and unfamiliar jobs with and without detailed task and reference price information**



### Experiment 6 – The processes of WTP and Pleasure ratings

Thus far, we have presented evidence that individuals are using external, distant, and even irrelevant cues when they are required to provide monetary evaluations (WTP and WTA). We have argued that this stems from the locus of monetary considerations in face of a mapping which unnatural and unlike pleasure considerations – not intuitive. The evidence is consistent with this claim, but thus far, we assumed the later part of the claim – that monetary considerations are far less intuitive than pleasure considerations. In order to demonstrate this, one needs to obtain process measures demonstrating that the process of coming up with a monetary response is more cumbersome than that of coming up with a happiness rating. One such measure may be the amount of time it takes individuals to respond to these questions. Experiment 6 measured reaction times to questions about WTP and expected happiness. The results of Experiment 5 further suggest that the difference between monetary responses and pleasure predictions does not emerge from an information accessibility account. This result is important for it makes the scale compatibility account even less likely. It seems that people “choose” to use the external cues, rather than use them for lack of invoked internal representations of pleasure. To replicate the results and further explore this intuition, Experiment 6 also manipulated the cognitive availability of pleasure related information by eliciting experiences of similar pleasure levels. If the external information is utilized in monetary reactions because of a

lack of accessible pleasure information, making such information available should reduce the diffusion of the WTA responses. However, if the external information is purposefully used, as we hypothesize, then we should find no difference in the WTA response distributions.

**Method:** Three hundred and forty nine individuals responded to a banner ad in a major online retail website and participated in an online questionnaire. We told participants that their favorite performer was coming to town, and that a friend may be able to get them tickets for the show. We asked half of them about their WTP for the ticket, and the other half to rate the level of happiness they would derive from the experience. Before reaching this question, however, a third of the participants were asked to write one experience that was pleasure-equivalent to the concert, a second third were asked to state three such experiences, and the final third proceeded directly to the main DV. This resulted in a (2) DV (WTP vs. happiness) x (3) pleasure-equivalent elicitation (one, three, or none) between participants design. The dependent measures of interest were the WTP and pleasure values, and the time participants took to report these values. For the WTP response, participants wrote the exact amount they would be willing to pay, and for the pleasure rating, participants wrote the number that corresponded to a pleasure scale: 0 (not pleasurable) to 100 (very pleasurable). We measured reaction times from the time the question page was displayed to the time participants clicked the “submit answer” button.

## **Results & Discussion**

The results for the six conditions are summarized in Table 4. The main effect of the dependent measure (i.e., WTP vs. pleasure ratings) replicated the previous experiments. Because of the difference in the possible range of responses (i.e., the pleasure scale was restricted to 100 from above), we compared the rescaled WTP distributions to those of the pleasure ratings<sup>2</sup>. Even after rescaling, the variances of the WTP responses were significantly greater than those of the corresponding pleasure

---

<sup>2</sup> Data was rescaled so that the maximum WTP observation per condition represented 100. In this way the variance is greatly reduced, and we are stacking the cards against us.

ratings ( $F[174,174] = 1.45$ ,  $p = 0.007$  for no PEs;  $F[118,118] = 2.01$ ,  $p < 0.001$  for one PE; and  $F[87,87] = 2.75$ ,  $p < 0.001$  for three PEs). This result is striking even more given that the means themselves contracted.

More important for the current experiment were participants' reaction times when providing monetary values and pleasure ratings. We subjected reaction times in the six conditions to a DV (2) x PEs (3) analysis of variance. The analysis revealed only a main effect for DV, whereby reactions were significantly faster for pleasure as opposed to monetary responses ( $F[1, 344] = 9.13$ ,  $p = 0.0027$ ). Both the effect of eliciting pleasure equivalent experiences, and its interaction with the type of dependent measure were not significant ( $F[2,344] = 0.123$ , ns. and  $F[2,344] = 0.51$ , ns. respectively).

**Table 4 Standard Deviations and Reaction Times across conditions**

DV	# Pleasure equivalents	Mean	StdDev	TTD(s)
WTP	None	32.87*	29.54*	24.75
	One	35.45*	37.86*	25.23
	Three	40.37*	31.22*	24.53
Pleasure ratings	None	65.44	24.56	17.58
	One	66.90	26.70	19.41
	Three	71.42	18.81	18.60

\* Statistics of the rescaled WTP data.

These results replicate the previous findings and add two additional insights. The first is evident from the reaction time results: monetary reactions are significantly slower, consistent with their less intuitive nature. The second arises from the null effect of the pleasure equivalent experiences elicitation, consistent with the different content as opposed to accessibility driving the difference in reaction types. This further suggests that even faced with salient pleasure information of their own making, participants nevertheless used the diffuse external information in the situation to provide their monetary evaluation.

Thus far, we have demonstrated that monetary evaluations do not match predicted pleasure, happiness, or utility when external cues are not aligned with internal cues. We have argued that difference in the locus of the cues used is the reason for many problems with monetary evaluations that stem from the contrast between the intuitive nature of

experienced utility predictions contrasted with the unnatural process of monetary evaluations. The suggested mechanism is evidenced in the experiments thus far by demonstrating that the discrepancy between predicted pleasure and monetary evaluations increases when external cues are not aligned with internal ones and when such cues are made more salient, but decreases somewhat when individual “commit” to expressed internal cues and only then provide monetary evaluations (as when they are elicited first). Another way to explore the mechanism is to manipulate accessibility of internal cues and measure its influence on both experienced utility predictions and monetary evaluations.

### **Experiment 7 – Internal Cues**

Experiment 7 attempted to explore this final idea: if the locus of consideration for monetary evaluations is indeed external but that of pleasure predictions is internal, then manipulating internal cues (or their accessibility) should influence the latter, but not the former. The following experiment tests this prediction.

**Method:** Two hundred and eighty one individuals participated in an online study. Participants were either asked to rate the degree of pleasure they would derive from eating a big pail of ice-cream, on a 0-100 scale, or asked for their maximum WTP for one. Prior to answering this, half of the participants were also engaged in a hedonic priming manipulation that included questions such as “Describe how you would feel after a full body massage” or “Please try to stop for a moment and think of how you are feeling right now”. This priming manipulation was predicted to enhance the accessibility of internal cues. This resulted in a 2 (DV: WTP vs. predicted pleasure) x 2 (internal accessibility: enhanced vs. normal) between participants design. As before, the dependent measures of interest were the WTP and predicted pleasure ratings.

### **Results & Discussion**

The results of the experiment confirmed our predictions: While the hedonic accessibility manipulation influenced the amount of pleasure participants predicted they would derive from the pail of ice-cream [ $t(139) = 2.00, p = 0.047$ ], it did not influence their WTP responses [ $t(142) = 0.72, p = 0.47$ ]. This result is qualified by a significant

interaction between the hedonic accessibility priming manipulation and the predicted pleasure DV [ $t(277) = 2.08, p = 0.039$ ].

Above and beyond replicating a mismatch between WTP and predicted experience utility, these results provide additional evidence for the proposed thesis. While previous experiments manipulated the external cues, influencing WTP evaluations but not pleasure predictions, Experiment 7 manipulated internal cues (and their accessibility) and achieved the converse. When internal cues were made more accessible through a hedonic priming manipulation, pleasure predictions but not WTP estimates were affected. The observed differential influence provides strong evidence for the differential locus of consideration theory.

### **General Discussion**

Money is ubiquitous in daily life. Nevertheless, human experience differs from linear decimal quantities in several important ways that make its mapping to money non-intuitive. Yet, individuals are required to make monetary evaluations when they make purchases, pay for services, or accept a salary for exerting effort, time, and talent. It is in these types of decisions that money, instead of being the “ultimate equalizer” and representing our predicted internal benefits, enabling smooth trade, distorts the way individuals make sense of experiential trade-offs. We argue that the reason for this distortion lies in the mapping between experience and money, while evaluations of the utility of an experience are intuitive and have an internal locus, evaluations of monetary values are more artificial and rely on external cues.

In order to support this argument we needed to show that, in situations where external cues are at odds with internal ones, individuals make monetary decisions that are inconsistent with their predictions of their future happiness. Furthermore, we needed to show that the two types of judgments relied on cues of different location: predicted pleasure relying on internal (and intuitive) cues, but monetary judgments on external (market related) cues. In Experiment 1, we demonstrated that while individuals predict that their happiness levels would be higher when the room temperature is more pleasant, their willingness to pay (WTP) is more influenced by the less relevant cost of production. In this particular experiment, this discrepancy led to an implied preference reversal,

whereby people would pay more for a concert they would enjoy less. Such inconsistencies are not new to the literature. Our contribution, however, lies in suggesting a general underlying reason for their existence – the very different nature and locus of monetary and pleasure considerations. Experiment 2 extended the previous result to individuals' willingness to accept (WTA), and, in addition, demonstrated that this discrepancy could not be explained by a simple scale compatibility account, suggesting a deeper source for the response discrepancy. Experiment 3 further generalized the effect to consequences in the loss domain, and monetary evaluations of another's work, as well as provided initial support for our proposed theory. The results revealed an asymmetric order effect, whereby monetary judgments became much more consistent with predicted utility when measured directly after the latter were elicited, but predicted utility was not influenced by an inconsistent preceding measurement of WTP. We interpret this finding as evidence that when internal cues are made extremely salient (one has just committed to them in writing) they do influence monetary evaluations, but absent this influence – they do not. Together these experiments demonstrate a problem with monetary evaluations, of both WTP and WTA, being discrepant from predicted happiness, effort, and personal gain / loss.

Our proposed different locus theory implies the general use of external cues and not only to the extent that internal signals are weak. A direct corollary is that absent external salient cues, individuals will have a hard time mapping experience to money, resulting in a greater dispersion of the distribution of replies. Experiment 4 investigated and confirmed this hypothesis. For uncommon tasks, absent market pricing information, participants generated greatly dispersed distributions of WTA, whereas the contrary was true for common tasks with a high probability of knowledge of external cues. Experiment 5 replicated this main effect, and in addition, demonstrated that the reduction of variance in the common tasks was mostly driven by knowledge of a reference-price rather than better knowledge of task-specific information. Together these experiments reveal that when forced into the non-intuitive process of mapping experience to money, individuals do not come up with a united coherent evaluation, even though their evaluations of pleasure or effort are.

The high variance in monetary responses observed in our experiments whenever external cues were absent is consistent with our claim that, unlike pleasure evaluations, such responses are not natural or intuitive. In Experiment 6 we tried a more direct measure of the implied process difference – reaction time. Indeed, Experiment 6 demonstrated that responding in monetary terms as opposed to pleasure ratings takes significantly longer. Moreover, the experiment also that this effect is robust to making pleasure equivalent experiences cognitively available in the case of monetary responses rendering alternative accounts of accessibility and differential search or activation less likely.

Finally, the yet untested orthogonal prediction, that manipulation of internal as opposed to external cues should influence utility predictions but not monetary judgments, was tested in Experiment 7. We manipulated the accessibility of internal cues, using a hedonic priming manipulation, and found that, as predicted by the proposed theory, predictions of pleasure, but not monetary evaluations (WTP) were affected.

The set of experiments presented corroborates our thesis, and suggests a reason for many documented inconsistencies between measurements of predicted pleasure, happiness, or experience utility, and corresponding monetary judgments, assessments, or evaluations. The idea that people think about these theoretically equivalent measures in very different ways (i.e., they consider very different input when solving the judgment problem: internal in the case of pleasure predictions, and external in the case of monetary judgments) fits very well with the artificial nature of money as opposed to the intuitive nature of pleasure predictions (perhaps a “system-1 – system-2” discrepancy; Kahneman & Frederick 2003). The current thesis may explain why anchoring effects are profound with monetary evaluations, and once one is made, it becomes an external cue and following judgments align to it (Ariely, Loewenstein, & Prelec 2003), but not for experiential judgments (XXXX Cite to CA2 XXX); it may also shed light on the reason that buyers and sellers think in different ways about the transaction (Carmon & Ariely 2000). ### We can put more cites here... ###

It is perhaps important to clarify that we by no means argue that individuals are wrong to use heuristics such as market values to make sense of their consumption world. Doing so may be adaptive and represent their best reaction to the situation (Payne,

Bettman, & Johnson 1993). For example, in situations where internal cues are subject to extreme influences and one is engaged in transactions that may have long term financial implications (e.g., gambling under the influence), one would be wise to rely on external cues, albeit, the right set of cues. We do, however, argue that this behavior may generate inefficiencies and put people at a disadvantage, especially when the external and internal cues are misaligned. We can divide cases with potential inefficiencies into two types. The first type is when individuals are sufficiently different from the market in a given domain, or the situation one finds herself in is different enough from the standard market situation. The second type is cases where there is no market price (e.g., really-new products or services, extreme changes in lifestyle such as having children), similar to our irregular tasks experiments. In all these events, monetary evaluations are subject to inherent errors. One possible solution is to use non-monetary scales to elicit evaluations, and even non-monetary experiential rewards as compensations in such cases. Another may be to guide people into choosing based on their pleasure evaluations, or make their internal cues more accessible.

In sum, the unnatural advantages of money prevent it from being the smooth equalizer it is perceived to be. The psychology of consumption is thus subject to biases in our everyday lives. We are always able to provide an evaluation when asked to do so, but we must be aware that our evaluation is not necessarily what it theoretically or optimally supposed to have been.

## References

- Ariely Dan, Loewenstein George, & Prelec Drazen (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, Vol.118, No. 1 (February), 73-105.
- Carmon Ziv, & Ariely Dan (2000). Focusing on the forgone: How value can appear so different to buyers and sellers. *Journal of Consumer Research*, Vol. 27, No. 3, 360-370.
- Hsee, Christopher K. (1999). Value Seeking and Prediction – Decision Inconsistency: Why Don't People Take What They Predict They'll Like the Most? *Psychonomic Bulletin & Review*, Vol. 6, No. 4, 555-561.
- Kahneman, Daniel, & Frederick, Shane (2003). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In Gilovich, Thomas, Griffing, Dale, and Kahneman, Daniel (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press.
- Payne, John W., Bettman, James R., and Johnson, Eric J. (1993). *The Adaptive Decision Maker*. New York: Cambridge University Press.
- Shafir, Eldar, Diamond, Peter, & Tversky, Amos (1997). Money Illusion. *The Quarterly Journal of Economics*, Vol. 112, No. 2 (May), 341-374.
- Simonsohn, Uri, & Loewenstein, George (2003). Mistake #37: The Effect of Previously Faced Prices on Current Housing Demand. Carnegie Mellon University Dept. of Social and Decision Sciences Working Paper.
- Soman, Dilip, Klaus Wertenbroch and Amitava Chattopadhyay (2003). Currency Numerosity Effects on the Perceived Value of Transactions. Working paper.
- Tversky, Amos, Sattath, Shmuel, & Slovic, Paul (1988). Contingent Weighting in Judgment and Choice. *Psychological Review*, 95(3), 371-384.