A Model of Focusing in Economic Choice*

Botond Kőszegi
University of California, Berkeley

Ádám Szeidl
University of California, Berkeley

March 6, 2011

Abstract

We present a theory of individual choice in which the decisionmaker focuses more on, and hence weights more heavily, attributes on which the options in her consideration set are more different. Consistent with evidence on salience in monetary choices, our model predicts that the decisionmaker is biased toward options whose advantages are concentrated in fewer attributes. In intertemporal choice, because a single period's choice can lead to a different concentration of consequences than a lifetime perspective that integrates many choices, the model often predicts time inconsistency in behavior. The decisionmaker exhibits present bias in "lifestyle" decisions whose consequences are distributed over many future dates, but also overcommits to an increasing number of future goals with a single large benefit each. In response to the bias toward concentration, profit-maximizing firms design products with one core attribute, and split prices into as many pieces as they can. A strong firm designs products which are strong on its competitor's weak attribute, while a weak firm copies the strong firm's strength. We also propose a theory of consideration-set determination in which the agent considers the set of options that maximizes a combination of utility and differences between attributes.

*We thank David Ahn, Stefano DellaVigna, Michael Grubb, Joan de Martí Beltran, and Matthew Rabin for insightful discussions. Kőszegi thanks the Center for Equitable Growth and the National Science Foundation (Award #0648659) and Szeidl thanks the Alfred P. Sloan Foundation for financial support.

1 Introduction

How a person chooses from a given set of options often depends not only on the intrinsic attributes of her options, but also on which attributes she focuses on at the moment. While this basic observation has long been recognized in psychology, it has had few systematic applications to economic questions. In this paper, we develop a model based on the idea that a person focuses more on attributes in which her options differ more. We apply our theory to choice over time and product design, and show that it naturally reconciles conflicting evidence and intuitions that have previously been unexplained or explained using different models. Because our model makes a prediction in any choice problem in which a classical model would, it also provides a simple way to incorporate this type of focusing into other economic applications.

We present the basic framework in Section 2. Each option c in the decisionmaker's finite consideration set C provides utility on K different "attributes," with the *consumption utility* from a choice being additively separable: $U(c) = \sum_{k=1}^{K} u_k(c)$.[1] While consumption utility corresponds to the decisionmaker's true experienced utility, she instead acts to maximize *focus-weighted utility* $\sum_{k=1}^{K} g_k u_k(c)$, where g_k is a weight that captures the extent to which she focuses on attribute k. Of the many determinants of focus, in this paper we model one that both seems powerful and is theoretically implementable. Building on evidence from Dunn, Wilson and Gilbert (2003) and Schkade and Kahneman (1998) and adapting insights from Bordalo, Gennaioli and Shleifer (2010), we assume that the decisionmaker focuses disproportionately on attributes in which her options generate a greater range of consumption utility: $g_k = g(\Delta_k(C))$, where $\Delta_k(C) = \max_{c \in C} u_k(c) - \min_{c \in C} u_k(c)$ and $g(\cdot)$ is an increasing function. When looking at her options, the decisionmaker rightly thinks about the biggest differences, but wrongly *overweights* those big differences in making her choice – a mistake consistent with the view of Schkade and Kahneman (1998) that "nothing in life is quite as important as you think it is while you are thinking about it."

In Section 3, we state and discuss the central prediction of our model, that the decisionmaker exhibits a "bias towards concentration." Because she focuses too much on a few big benefits relative

[1] In this paper, we consider only riskless choices. In Kőszegi and Szeidl (2011, hopefully), we study an extension that allows for uncertainty by thinking of ordered lottery outcomes as attributes.

to many small losses, she is too prone to choose options whose advantages are concentrated in fewer attributes. This prediction—equivalent to our main assumption that $g(\cdot)$ is increasing—serves as a key empirical motivation for our model because it helps organize under one framework several findings about salience in monetary choices. Specifically, if payments received on different dates are separate attributes, our model predicts that a person feels more strongly about one big payment than about a sequence of smaller payments. This can explain why most taxpayers choose one lump-sum tax refund over lower monthly withholdings (Jones 2010), and provides one reason for the common practice of financing consumer products. Similarly, if the posted price and the extra costs of a product are separate attributes, consumers will perceive a product to be less expensive if its price is split into multiple components. This can explain the finding of Chetty, Looney and Kroft (2009) that demand for a taxable item in the supermarket is lower if its tax-inclusive price is posted on the shelf, and creates a motive for the retailer practice of splitting prices into multiple components. For example, airlines separate out a ticket's price from taxes and fees, and mortgage brokers split their compensation into a multitude of fees over which the borrower has no choice.

The bias towards concentration does no arise in balanced decisions, where the costs and benefits of two alternatives are identically concentrated, and hence a person maximizes consumption utility. This observation allows us to elicit consumption utility and other model ingredients from choice behavior, and also provides guidance as to the kinds of situations in which choices reflect welfare. In addition, because for small differences between options the focus on each attribute is close to the baseline level, if $g(0) > 0$ our model also predicts that a person maximizes consumption utility for small stakes. In predicting that focus matters for large stakes but may not matter for small stakes, our model contrasts with the implications of boundedly rational models of attention and information processing, such as Reis (2006), Gabaix, Laibson, Moloche and Weinberg (2006), Sims (2010), and Salant (forthcoming), in which a person processes information more fully, and hence makes a better choice, in more important decisions.

In Section 4, we explore the implications of our model for intertemporal choice. Assuming that utility outcomes on different dates correspond to different attributes, the central principle emerging from our model is that it is the *concentration*, rather than the *temporal placement*, of an option's

costs and benefits that determines the agent's assessment of the option. Furthermore, because the concentration of costs and benefits can be different from the ex-post perspective of a single choice and the ex-ante perspective that aggregates multiple decisions, the decisionmaker often exhibits a kind of time inconsistency. Our model predicts both how the choice problem shapes the strength and direction of time inconsistency, and whether the ex-ante or ex-post preference more accurately reflects welfare.

As an example of one prototypical type of decision situation, consider the daily decision of whether to exercise, where doing so results in pain today and tiny benefits for many periods in the future. Because on each day a person focuses too little on the many small gains relative to the one big cost, she exhibits overly present-oriented behavior. But when considering the entire sequence of decisions ex ante, the benefits accumulate into larger and hence more salient gains, so that the person would commit to more future-oriented behavior. This scenario fits many types of decisions, including lifestyle choices such as diet and consumption-savings decisions, that have been invoked as prime examples of present bias in the literature. In these settings, our model both explains present bias and predicts that the ex-ante decision—it being a more balanced choice—more accurately reflects welfare.

As an example of a different prototypical type of decision situation, consider choices, such as writing a book, where a sequence of sacrifices leads to a single large goal. Ex ante, the decisionmaker focuses too much on the large goal and hence is too prone to commit to writing the book. On any individual day ex post, however, the person weighs effort that day against a marginally better book, and working becomes less attractive. This scenario fits career and other "achievement"-type decisions, for which some researchers believe that individuals commit themselves to overly future-oriented paths (Scitovsky 1976, Kahneman, Krueger, Schkade, Schwarz and Stone 2006, Hamermesh and Slemrod 2008). In this case, our theory says that the individual's ex-post decision to do less for the achievement—being the balanced choice—more accurately reflects her welfare. Nevertheless, since the diffuse costs associated with each *additional* achievement look small relative to the single benefit, we also predict that the decisionmaker often chooses to add ever more challenges on top of her current commitments, reducing her consumption utility. As these examples show, our model

simultaneously explains common examples of present-oriented and future-oriented behavior, and also provides a simple welfare guide based on the relative balancedness of the ex-ante and the ex-post decision.

In Section 5, we explore how firms design products to take advantage of consumers' choice-set-dependent focus, deriving some specific predictions and demonstrating that questions that have so far largely been studied in marketing can be analyzed using economic methods. Our model predicts that a monopolist evenly divides the price among available payment components, while it concentrates product value on a single attribute. The former implication is consistent with the retailer practice of splitting prices discussed above, and the latter seems central in marketing analyses of product positioning. For example, Aaker (1991) argues that "a positioning strategy which involves too many product attributes can result in a fuzzy, sometimes contradictory, confused image. In part, the problem is that the motivation and ability of the audience to process a message involving multiple attributes is limited."

We also analyze firms' incentives in choosing the attribute to concentrate on. As examplified by the association of Volvo with safety and BMW with performance, competing firms often concentrate on a single attribute in which they are perceived to be market leaders. But as illustrated by the flood of touch-screen phones following the introduction of the iPhone, another common positioning strategy is to copy the key attribute of a competitor. Our model predicts that concentrating on the other firm's weakness is optimal for strong competitors who can design high-utility attributes that attract focus. In contrast, copying the defining—most salient—attribute of the incumbent is optimal for weaker firms who are not able to shift consumer focus. In the context of cell phones, firms who cannot attract focus with an innovative new attribute should add a touch screen to ensure that consumers also focus on their product.

An important property of our model is that unchosen options often affect focus, so that assumptions about which "irrelevant" options make it into the decisionmaker's consideration set affect predictions. While our theory of choice from a consideration set C can be combined with any theory of how C is formed, in Section 6 we propose a model of endogenous consideration-set determination that is consistent with the spirit of our model that differences attract attention. We

propose that from her choice set, the decisionmaker only considers the set of options that maximizes a linear combination of average consumption utility and average differences between options. We illustrate the use of our theory through an example. If a consumer's choice set is whether or not to buy a particular item, unless the item is very expensive or very cheap her consideration set will be equal to her choice set. But if her choice set includes two possible attractive but different purchases, not buying might not make it into her consideration set. With her decision effectively shifting from *whether* to buy to *what* to buy, differences in money outlays become smaller, and hence the weight the consumer puts on the money dimension diminishes. This can lead her to more easily buy a more expensive product.

With the endogenous theory of consideration sets, our model has the useful methodological feature that applying it to a new (deterministic) economic setting requires—besides assumptions a classical model would also make—only the specification of relevant attributes for that setting. Doing this is straightforward in intertemporal decisions where our definition of dates as attributes leaves zero degrees of freedom. In other situations, the definition of attributes is less clear, and our theory provides no guidance for this modeling choice. In our applications, we make what we believe are reasonable attribute choices, but a reader who disagrees can also straightforwardly apply our model to her preferred set of attributes. In addition, results on how predictions depend on attributes may be useful for empirical work in identifying attributes from choice behavior. We discuss further issues with our model, possible extensions and applications in Section 7.

2 A Theory of Focusing

We now formulate our model of focusing, first specifying a static version in Section 2.1, and then extending it to dynamic choice in Section 2.2. In this paper, we model only riskless choice; in Kőszegi and Szeidl (2011, hopefully), we extend our model to risky choice in a way closely related to Bordalo et al. (2010).

2.1 Focus-Weighted Utility

We model choices from a finite set C. As will be clear below, in our model manifestly irrelevant alternatives in C can affect a person's focus and hence her behavior. Hence, to avoid some seemingly unrealistic predictions we (realistically) allow C to be different from the entire choice set, thinking of it as the decisionmaker's set of reasonable options or "consideration set." While for the bulk of the paper we take C as exogenous, in Section 6 we propose a way to define it endogenously. Our theory of choice from consideration sets can also be combined with any other model of consideration-set determination.

The decisionmaker's *consumption utility* from option $c \in C$ is $U(c) = \sum_{k=1}^{K} u_k(c)$, the sum of the utilities it provides on K "attributes." Consumption utility corresponds to the decisionmaker's experienced utility from the product. In many applications, it will be natural to assume that one attribute corresponds to one consumption dimension, so that $c = (c_1, \ldots, c_K) \in \mathbb{R}^K$ and $U(c) = \sum_{k=1}^{K} u_k(c_k)$. But in other settings—such as when consumption in one period of an intertemporal choice problem affects utility in multiple periods—there is no such one-to-one correspondence.

The central assumption of our model is that the consideration set C affects the relative focus on different attributes, and thereby also perceived utility and choice. In particular, we assume that the decisionmaker's focus on, and the weight she attaches to, an attribute when making her choice is a function of the dispersion of consumption-utility outcomes in that attribute. Formally, the *focus-weighted utility* of option c from set C is

$$\tilde{U}(c, C) = \sum_{k=1}^{K} g(\Delta_k(C)) \cdot u_k(c), \tag{1}$$

where $\Delta_k(C) = \max_{c \in C} u_k(c) - \min_{c \in C} u_k(c)$ and $g(\Delta_k(C))$ is the focus weight on attribute k.[2]

Much of our paper will be based on the following basic hypothesis:

Assumption 1. The function $g(\Delta)$ is strictly increasing in Δ.

Assumption 1 says that the decisionmaker focuses more on attributes in which her options generate a greater range of consumption utility. This assumption is directly supported by some

[2] Our formulation uses a particular tractable functional form to capture dispersion in consumption utility. We are not aware of evidence that would allow us to tie down this functional form precisely. The logic of our results in this paper, however, does not seem to depend on the specific formulation.

evidence on what people focus on in specific situations.[3] For example, Dunn et al. (2003) examine freshmen's predicted and upperclassmen's actual happiness with their randomly assigned dorms at a major university. Consistent with our hypothesis, predicted happiness depends greatly on features (e.g. location) that vary a lot between dorms, and not on features (e.g. social life) that vary little between dorms—whereas actual happiness does not show the same patterns. Similarly, Schkade and Kahneman (1998) find that both Midwesterners and Southern Californians incorrectly predict Californians to be more satisfied with life because they focus on the main differences (climate and cultural opportunities) between the two locations and underweight important other determinants of life satisfaction. But rather than the direct evidence, our main motivation for Assumption 1 is that a key prediction below on the "bias toward concentration," a prediction that is equivalent to an increasing $g(\cdot)$, helps understand a variety of facts documented in the literature. In addition, our key assumption is similar to a central assumption in Bordalo, Gennaioli, and Shleifer's (2010) model of salience in choice under uncertainty, which the authors use to endogenously derive a weighting function akin to the probability weighting function in prospect theory.

While focus-weighted utility determines choice, we assume that the decisionmaker's welfare is given by her consumption utility. This modeling approach reflects our view (consistent with the psychology literature above) that focusing distorts the individual's perception of preferences and perhaps her well-being near the moment of choice, but does not alter the much more important experienced utility emanating from the choice.

Given the psychological evidence and economic examples motivating our model, the most natural interpretation of our formulation is that the decisionmaker is at some level aware of the attributes of her options, but does not focus equally on them. We refer to this phenomenon as "focusing" partly to distinguish our model from theories of attention in economics, where the issue is more about what information a person becomes aware of. Nevertheless, when it seems semantically appropriate and does not cause confusion, we also refer to the phenomenon in terms of attention.

Note that unless $g(\cdot)$ is linear, affine transformations of consumption utility do not in general result in affine transformations of the attention-weighted utility function. While this raises no

[3] The evidence we discuss pertains to predicted happiness rather than choice—but presumably individuals would make choices in the above situations corresponding to their predicted happiness.

problem in applying our model once both $u_k(\cdot)$ and $g(\cdot)$ are specified, it does mean that when deriving our model from a classical model based on consumption utility alone, the specification of $g(\cdot)$ must be sensitive to the scaling of consumption utility. In Appendix B, we show how to elicit the model's ingredients from behavior; this elicitation guarantees that $g(\cdot)$ will be scaled to $u_k(\cdot)$.[4]

2.2 Focusing in Intertemporal Choices

To explore the implications of focusing in intertemporal decisions, we now formulate a dynamic model. Our starting assumption is that consumption utilities realized on different dates are evaluated as separate attributes. This view is motivated in part by Thaler and Johnson (1990), who argue that temporal separation of payments or receipts facilitates separation of these monetary transactions in a person's mind. But even beyond this basic assumption, application of our framework requires us to posit how a person conceptualizes her choice set in a period. Our key assumption is that the agent represents her consideration set in a period as the set of lifetime consumption profiles associated with her current options, given her beliefs regarding her future behavior. In a consumption-savings decision, for instance, the individual would have beliefs regarding how a dollar consumed today affects consumption on each future date. Through these beliefs, today's consumption possibilities generate a set of lifetime consumption profiles, and this set determines the individual's focus.

Formally, we consider a framework with T periods, $t = 1, \ldots, T$. Each period t, the agent must make a choice c_t from the deterministic finite consideration set $C_t(c_1, \ldots, c_{t-1})$. Continuing with the consumption-savings example, c_t can represent the bundle of goods to be consumed today, and C_t the budget set given the path of past consumption and possible credit constraints. We abstract away from any time inconsistency or time discounting, and assume that the decisionmaker's consumption

[4] Because the weight $g(\Delta_k)$ on an attribute depends only on the utility dispersion $\Delta_k(C)$ in that attribute, literally interpreted Equation (1) says that focusing more on one attribute does not reduce the focus on other attributes. But choices are determined by the *relative* focus across attributes, and increasing the weight on one attribute does reduce the relative weight on other attributes. To make the role of relative weights more transparent, we could normalize each attribute's weight by $\sum_{j=1}^{K} g(\Delta_j(C))$, which would ensure that weights sum to one. This adjustment only multiplies $\tilde{U}(c, C)$ by a constant for all options, and hence does not change the choice implications of the model.

Equation 1 also implies that the relative focus $g(\Delta_i)/g(\Delta_j)$ on two attributes i and j depends only on attributes i and j of the available options. This property is analytically convenient and does not seem to eliminate important economic forces in the applications we have explored.

utility in period s is $\sum_{t=s}^{T} u_t(c_t, c_{t-1}, \ldots, c_1)$ where u_t is the possibly history-dependent per-period utility function.

Besides requiring that utilities realized on different dates are evaluated as separate attributes, we also allow for κ different attributes at each date t, and assume that the instantaneous utility function u_t is additively separable over these κ attributes. We formalize the individual's beliefs at time t about the consequences of her actions with the functions $\{c_\tau(c_t)\}_{\tau=t+1,\ldots,T}$, which specify future choices as a function of the current decision c_t. We will return to the question of how the individual might formulate her beliefs below. For any history $c_1, \ldots, c_{t-1}$, the decisionmaker's beliefs induce a set of lifetime consumption profiles as a function of the current action c_t:

$$\overline{C}_t = \{ (c_t, c_{t+1}(c_t), \ldots, c_T(c_t)) |\, c_t \in C_t(c_1, \ldots, c_{t-1}) \}.$$

We assume that $\overline{C}_t$ is the consideration set that determines the decisionmaker's focus in period t, so that she applies the model of Section 2.1 to $\overline{C}_t$.

An alternative model to that above would be one in which the entire dynamic choice set determines the agent's focus. Our formulation reflects the idea, broadly consistent with evidence on narrow bracketing (e.g. Tversky and Kahneman (1981) and Rabin and Weizsäcker (2009)), that people make decisions one-by-one rather than by integrating them into a grand lifetime decision problem.

Our framework can be combined with any theory of the agent's beliefs regarding future behavior. Following standard economic methodology, in this paper we assume that the agent formulates "rational" rational, i.e. she correctly predicts her future behavior. This implies that we can derive behavior in any decision problem using backward induction. In Appendix C, we formulate a "naive" theory, in which the decisionmaker believes that she will act in the future as she would now commit herself to do. The insights in this paper hold for both theories of how the agent forms beliefs about the future.

3 Bias Toward Concentration

This section identifies the key behavioral implication of our model: because the decisionmaker weights a few big advantages more heavily than many small disadvantages, she is biased toward options whose advantages are more concentrated than its disadvantages. In Section 3.1, we develop theoretical results that formalize this insight, and in Section 3.2 argue that it explains evidence in a number of domains.

3.1 Balanced and Unbalanced Choices

Balanced choices. We begin our analysis of the bias towards concentration by showing that in decisions where neither alternative has more concentrated advantages, the agent makes the consumption-utility-maximizing choice. As a simple example, consider the following two alternatives in utility space: $(a, 0, 0)$ and $(0, b, 0)$ with $a, b \geq 0$. This is an example of a "balanced" choice, where the gains of the two alternatives are equally concentrated: they are both on a single dimension. Using Equation (1), the consumer strictly chooses the first option if and only if $g(a)a > g(b)b$. If Assumption 1 (that $g(\cdot)$ is increasing) holds, this inequality is equivalent to $a > b$, the condition under which the first option is the unique utility-maximizing choice. While applying the focusing weights changes the agent's perceived utility of both options, it does so in a way that reinforces the original utility ordering.

To generalize this example, we say that a choice set $C = \{c^1, c^2\}$ has *balanced tradeoffs* if $u_1(c^1) - u_1(c^2) = u_2(c^1) - u_2(c^2) = \cdots = u_{K'}(c^1) - u_{K'}(c^2) \equiv a > 0$, $u_{K'+1}(c^2) - u_{K'+1}(c^1) = \cdots = u_{2K'}(c^2) - u_{2K'}(c^1) \equiv b > 0$, and $u_k(c^1) = u_k(c^2)$ for $k > 2K'$. That is, option 1 is better by a on K' attributes and worse by b on the same number of attributes. Then:

Proposition 1 (Rationality in Balanced Tradeoffs)**.** *Suppose Assumption 1 holds. In a balanced tradeoff, the decisionmaker chooses the first option if $a > b$ and the second option if $a < b$.*

Extending the logic of Proposition 1, the agent chooses rationally whenever the vector of utility differences in which c^1 is better than c^2 is a constant multiple of the vector of utility differences in which c^1 is worse than c^2—that is, when the advantages of c^1 relative to c^2 have the same pattern

across attributes as its disadvantages.

Beyond identifying a class of situations in which the agent acts rationally, Proposition 1 has a convenient implication: it allows us to elicit all ingredients of our model from choice behavior. To see how a consumer's monetary valuation for a one-attribute product can be elicited, for instance, consider the choice between utility vectors $(u, -v(p))$ and $(0, 0)$, where the former option is buying the product with consumption utility u at price p and the latter option is not buying. This is a balanced choice with $K' = 1$, $a = u$ and $b = v(p)$. Thus, at the point where the consumer is indifferent, we must have $u = v(p)$, so p is the monetary value of the product. The same idea can be applied when a good has multiple attributes: we can elicit the value of each attribute by varying it and measuring the change in the consumer's financial value for the good. Building on these insights, Appendix B shows how to also elicit the focus weights from choice behavior.

Note that once the ingredients of our theory are obtained in the above way, the theory provides a complete mapping from consideration sets to choices and welfare. As a result, our theory satisfies the spirit of the revealed-preference criterion for economic theories, that one be able to identify the theory's full set of testable predictions as well as its welfare implications from behavior.

Bias toward concentration. The main result of this section is that agents are biased toward an option if its advantages are more concentrated than its disadvantages. To illustrate the idea, recall that the consumer makes a consumption-utility-maximizing choice if the two alternatives are $(a, 0, 0)$ and $(0, b, 0)$. Now consider instead the unbalanced choice between $(a, 0, 0)$ and $(0, b/2, b/2)$. Applying (1), the agent chooses the first option if $g(a)a > g(b/2)b$. This inequality is no longer a monotone transformation of the consumption utility order: in particular, for $a = b$, the consumer chooses the first option if $g(a) > g(a/2)$, which always holds under Assumption 1. Even though consumption utility is the same, the agent strictly prefers the first alternative because the large payoff a tilts her focus toward the first attribute.

To extend and formalize this insight, we introduce some terminology. For two alternatives c^1, c^2, let I be the set of attributes in which c^1 is better and J the set of attributes in which c^2 is better. When these two options generate the same consumption utility, $\sum_k u_k\left(c^1\right) = \sum_k u_k\left(c^2\right)$, we say that c^1 *has more concentrated advantages than disadvantages relative to* c^2 if $\min_{i \in I} \Delta_i >$

$\max_{j \in J} \Delta_j$, i.e., if the advantages are uniformly greater than the disadvantages. For example, this condition holds when $c^1 = (a, 0, 0)$ and $c^2 = (0, a/2, a/2)$, because the advantage a in the first attribute exceeds the disadvantages $a/2$ in the other two attributes. Then:

Proposition 2 (Bias Toward Concentration)**.** *Suppose that $g(\cdot)$ is continuous. Then the following are equivalent:*

1. Assumption 1 holds.

2. If c^1 and c^2 have the same consumption utility, and c^1 has more concentrated advantages than disadvantages relative to c^2, then the agent strictly prefers to choose c^1 from $\{c^1, c^2\}$.

We also formalize a slightly different version of the bias toward concentration. For a third option c', we say that c' combines two advantages of c^1 relative to c^2 if c' can be obtained from c^1 by merging its advantages in two attributes into a single attribute. Formally, this holds if there are attributes $k, l \in I$ such that $u_k(c') - u_k(c^2) = \big[u_k(c^1) - u_k(c^2)\big] + \big[u_l(c^1) - u_l(c^2)\big]$, $u_l(c') = u_l(c^2)$, and $u_n(c') = u_n(c^1)$ for all $n \neq k, l$, , i.e., the advantages of c^1 in k and l are merged to obtain c'. We say that c' *combines some advantages of c^1 relative to c^2* if there is a sequence $c(1) = c^1, c(2), ..., c(n) = c'$ such that $c(i)$ combines two advantages of $c(i-1)$ relative to c^2 for each i. The condition when c' combines some *dis*advantages of c^1 relative to c^2 can be formulated analogously.

Corollary 1 (Bias Toward Concentration)**.** *Suppose Assumption 1 holds. If c' combines some advantages of c^1 relative to c^2, and the agent chooses c^1 from $\{c^1, c^2\}$, she strictly prefers to choose c' from $\{c', c^2\}$. Conversely, if c' combines some disadvantages of c^1 relative to c^2, and the agent chooses c^2 from $\{c^1, c^2\}$, she strictly prefers to choose c^2 from $\{c', c^2\}$.*

Intuitively, contrast a situation in which an option has advantages of 10 utils on each of two dimensions with a situation in which it has an advantage of 20 utils on one dimension. In the former case, the difference of 10 utils on a dimension attracts some attention to that dimension and hence to that advantage, but it does not attract attention to the other 10 utils of advantage. In the latter case, however, each 10 utils of advantage attracts attention not only to itself, but also to the other 10 utils of advantage on the same dimension. As a result, concentrating the advantages of an option makes the option seem more favorable.

Small stakes yield rational choices. To conclude this section, we point out that if $g(0) > 0$, the decisionmaker maximizes consumption utility—and hence her bias toward concentrated advantages disappears—in the limit as stakes become small:

Proposition 3 (Rationality with Small Stakes)**.** *Suppose $g(0) > 0$ and $g(\cdot)$ is continuous at zero. For any $u = (u_1, \ldots, u_K)$ and $\delta = (\delta_1, \ldots, \delta_K)$, there is an $\bar{\epsilon} > 0$ such that if $0 < \epsilon < \bar{\epsilon}$, the decisionmaker makes a consumption-utility-maximizing choice from the choice set $\{u, u + \epsilon \cdot \delta\}$.*

Intuitively, with small stakes, the focus weights on all attributes are near their minimal level $g(0)$, and hence the agent does not perceive much of a difference in relative concentration across alternatives. Both because small-stakes decisions are all else equal less interesting and because we do not know whether $g(0) > 0$ holds in general, Proposition 3 is economically the least important of our predictions. But the result does highlight an interesting way our model is different from theories, such as Reis (2006), Gabaix et al. (2006), Sims (2010), and Salant (forthcoming), in which the agent makes rational choices taking into account the cognitive or effort cost associated with attention or calculation. In these models the agent is likely to pay more attention, and hence to make better decisions, in choices with larger stakes, where the potential benefits of choosing correctly are more worth the cognitive costs. This prediction is in sharp contrast with our model, where focusing distorts behavior in non-trivial unbalanced decisions, while rationality may obtain in small stakes.[5]

3.2 Evidence on the Bias Toward Concentration

In this section, we discuss a variety of evidence that can be explained by Proposition 2's prediction that the decisionmaker is biased toward concentrated advantages. In light of the equivalence between the bias toward concentration and Assumption 1, this evidence can be interpreted as directly supporting our main assumption, and serves as our key empirical motivation.

Most of the evidence we review here is about the evaluation of monetary payment bundles. Because consumption-utility-maximizing behavior would treat different sources of money in the

[5] If the focus weight function $g(\Delta)$ has a maximum, our model also implies rational decisions if stakes are very large in *all* attributes. Even then, the bias toward concentration remains in choices which contrast one large gain with several small losses in different attributes.

same way, this kind of evidence makes it particularly clear when choices are distorted relative to the optimum. To explain the evidence using our model, we will assume that individuals separate, in specific ways, different components of monetary payments or receipts into different focus-relevant attributes, and hence think of them as separate components of utility. The hypothesis that people separate money into different "mental accounts" of wealth, income, and expenditures has long been considered a stylized fact, and forms the basis of Thaler's (1985, 1999) theory of mental accounting. Unfortunately, however, there is little evidence on exactly how individuals compartmentalize money. Here we make what we believe are natural assumptions, but it is important to emphasize that our analysis does depend on these assumptions.

Some of the evidence can be organized around the following immediate corollary of Part 2 of Proposition 2:

Corollary 2. *Consider the choice between two K-dimensional vectors of consumption utility, $(v, ..., v)$ and $(Kv, 0, ..., 0)$. The decisionmaker strictly prefers the first alternative if $v < 0$, and the second if $v > 0$.*

Concentrated payment streams. Jones (2010) documents that nearly 80% of taxpayers receive a lump-sum tax refund when filing their taxes, even though they could ask their employer to pay them the same amount over the course of the previous year by withholding less from their paychecks. Hence, the choice to receive a lump-sum tax refund amounts to saving at a 0% interest rate. Previous research has identified three potential reasons for tax refunds: asymmetry in penalties around zero refunds (Highfill, Thorson and Weber 1998); a commitment to save (Thaler 1994); and inertia (Jones 2010). Our model provides a different explanation: taxpayers simply like the tax refund because it comes as a single attention-grabbing big payment rather than as a sequence of less-noticed small payments. For example, letting $u = \$200$ and $K = 12$, Corollary 2 says that a person likes a yearly payment of \$2,400 more than monthly payments of \$200.

Johnson, Hershey, Meszaros and Kunreuther (1993) find a similar preference in the context of insurance: in a hypothetical choice, the majority of subjects prefer a five-year disability policy with a \$20 higher monthly premium that pays a refund of \$1,200 if no disability claim is made. Abstracting away from the possibility that a disability claim might be made and thinking of the

higher premium and no refund as the baseline, this is exactly what Corollary 2 says for $u = \$20$ and $K = 60$.[6]

Prices and taxes. There is also empirical evidence consistent with Corollary 1. Chetty et al. (2009) document that demand for a taxable item is lower if the tax is included in the posted price than if it is added at the register. To see how our model explains this phenomenon, suppose a product has three consumption attributes: the consumption utility from the product u, the "posted price," and "further monetary costs." Suppose, for instance, that the product has a pre-tax price of \$99 and is subject to a tax of \$8.75. Then, buying the product when the posted price does not include tax is equivalent to choosing $(u, -\$99, -\$8.75)$, buying when the tax is included in the posted price is equivalent to choosing $(u, -\$107.75, \$0)$, and not buying is equivalent to choosing $(0, \$0, \$0)$. A direct implication of Corollary 1 is that a consumer is less likely to buy when the tax is included in the posted price.

Product design and pricing. The prediction of our model that splitting up the price of a product into many components increases demand for it also explains the popularity of such pricing attempts by firms.[7] For instance, Woodward and Hall (2010) note that mortgage-broker compensation is artificially split into many small fees, likely contributing to the consumer confusion surrounding these products. Similarly, many low-cost airlines split the price of flying into a number of components, including the ticket price as well as taxes, baggage fees, seat reservation fees, and credit-card transaction fees. In the same spirit, our result may also explain why firms frequently offer financing for large consumer purchases, such as durable goods. A laptop which costs "\$29 a month" may seem less expensive than one with a price of \$899. The main existing explanation for consumers' preference for financing is liquidity constraints, possibly exacerbated by present bias. Our model provides an additional explanation, which also explains why consumers sometimes resort to financing even when they have liquid funds available (Bertaut, Haliassos and Reiter 2009, Stango and Zinman 2009).[8] And although it is probably not the main explanation, the bias toward concen-

[6] Allowing for disability to occur makes our point stronger, because the refund is conditioned on no disability but the monthly premium payments up to the disability event are not.

[7] We explore the implications of the bias toward concentration for firm pricing in more detail in Section 5.

[8] Stango and Zinman (2011) propose a "fuzzy math" model of borrowing which also predicts that quoting the price of a loan in terms of monthly payments makes the loan seem cheaper to borrowers. The mechanism driving this effect, however, is completely different from ours, and only holds for short-term loans.

tration may help understand why consumers run up large bills from various small penalties and fees service providers impose. For instance, Agarwal et al. (2007, 2008) and Stango and Zinman (2009) document that consumers pay very significant amounts in banking and credit-card fees, and the Federal Communication Commission is considering regulating the mobile-phone market to help consumers avoid the "bill shock" from unexpectedly large bills due to many minor charges (Federal Communication Commission 2010). It is likely that in many of these situations, consumers are simply unaware that what they are doing is subject to a fee. But our model generates a kind of bill shock even when consumers are aware, and reinforces the tendency otherwise: as predicted by Corollary 2 ($v = -\$0.50, K = 200$), while fifty-cent charges for each of many text messages to a new partner while roaming do not seem like much, a $100 texting charge after an international trip does.

Finally, the bias toward concentration and the attention-based logic generating it are reminiscent of views commonly held in the marketing literature about the benefits of concentrating on a core product attribute. For example, Aaker (1991) argues that "a positioning strategy which involves too many product attributes can result in a fuzzy, sometimes contradictory, confused image. In part, the problem is that the motivation and ability of the audience to process a message involving multiple attributes is limited."

3.3 Bias Toward Concentration and Diminishing Sensitivity

Before moving on to applications, we briefly discuss the connection between our model and diminishing sensitivity, the basic hypothesis of prospect theory (Kahneman and Tversky 1979) that an initial deviation from a reference point is felt more strongly than an increase in the deviation. As emphasized by Thaler and Johnson (1990) and others, diminishing sensitivity predicts that individuals should prefer to segregate gains and integrate losses. This prediction is in direct contradiction with our model's prediction of a bias toward concentration, and most notably with Corollary 2.

The most compelling evidence for diminishing sensitivity comes from comparing attitudes toward changes starting from outcomes closer to versus further from the reference point.[9] Combined

[9] For instance, people are risk-averse over gains and risk-loving over losses (Kahneman and Tversky 1979), people are more willing to work to save a given amount when it comes off a small payment rather than a large payment

with diminishing sensitivity in consumption utility within each attribute, our model would accommodate all this evidence. Within the combined framework, diminishing sensitivity's prediction about separating gains and combining losses would be weakened or reversed by the bias toward concentration, explaining why there is limited evidence for this prediction.[10] More interestingly, our theory seems useful for making sense of some examples commonly used to illustrate diminishing sensitivity in the literature. One popular example argues that consumers are more willing to buy add-ons such as a car stereo after buying the car itself, due to diminished sensitivity to additional expenditures following a big payment (Thaler 1985, Chiu and Wu 2009, for example). This example actually *contradicts* diminishing sensitivity: due to diminishing sensitivity on the "car" attribute, the consumer should be more willing to make an unrelated purchase. In contrast, our model explains why the two purchases must be related: a consumer is more likely to purchase a car stereo because the high-value car tilts her focus toward the "car" attribute, which increases the perceived value of the add-on as well.

4 Intertemporal Choice

This section explores the implications of our model for choice over time. A fundamental insight of our model is that the concentration of an action's costs and benefits over time, rather than their temporal placement, determines the weights a person places on the consequences of the action. In addition, because a dynamic choice situation can generate different concentrations of costs and benefits when looked at from a lifetime perspective—with many decisions integrated—than when looked at from a single day's perspective—with only that day's decision to be made—our model predicts a form of time inconsistency in behavior. Below we flesh out the specific implications of these basic insights, and argue that our predictions help explain a number of stylized facts about intertemporal choice.

(Tversky and Kahneman's (1981) famous calculator example), and—gambling that the price will bounce back—people are more likely to hold on to losing assets than to winning assets (Odean 1998, Genesove and Mayer 2001).

[10] Thaler and Johnson (1990) and Linville and Fischer (1991), find that subjects do prefer to separate gains, but—in contrast to the prospect-theory prediction—they also prefer to separate losses. Lehenkari (2009) documents that individual investors in the Finnish stock market do not integrate sales of stocks that have lost money or segregate sales of stocks that have made money.

By endogenizing time preference based on the concentration of the costs and benefits of available actions, our theory contributes to the literature in three main ways. First, while in many circumstances our model—like models of present bias—implies an ex-post underinvestment into the future relative to a person's ex-ante preference, it also identifies plausible circumstances under which a person overinvests into the future, and keeps making additional overinvestments as time goes on. Second, in contrast to the existing literature on time-inconsistent behavior, our theory provides an unambiguous welfare measure without a-priori assuming that the ex-ante choice is welfare relevent, and identifies a simple way to determine whether ex-ante or ex-post choices better reflect welfare. Third, these theoretical results provide a new perspective about several empirical phenomena discussed by researchers in different contexts.

In Section 4.1, we develop the key ideas of this section through a stylized model of an investment decision. In Section 4.2, we discuss how our results relate to previous results and insights in the literature. Finally, in Section 4.3, we identify general conditions for when different forms of time inconsistency occur. For the rest of the paper, we suppose that Assumption 1 holds.

4.1 A Simple Investment Decision

We develop the central ideas of this section using a simple dynamic decision problem. There are $T_i + T_b$ periods, $t = 1, \ldots, T_i + T_b$. The agent makes an investment or effort decision $e_t \in \{0, 1\}$ in each of the the first $T_i \geq 1$ periods, and her investments provide benefits in the next $T_b \geq 1$ periods. The consumption-utility cost of investment in each of the first T_i periods is $e_t \cdot B$, and the consumption-utility benefit in each of the last T_b periods is $u_t = A\left(B \sum e_t\right)/T_b$. The variable A measures the efficiency of the investment: investing maximizes consumption utility if and only if $A \geq 1$.

This framework can be interpreted as a stylized model of many intertemporal problems, including (i) exercise, where u_t are the health benefits of exercise; (ii) work, where u_t derive from the compensation or other results from work; and (iii) consumption-savings, where u_t are the extra future consumption made possible by saving today. Of course, our simple framework abstracts away from many aspects of these settings, including possible overlaps between the investment and

benefit periods and non-binary choices. Adding these considerations to our framework would affect the focus weights and hence the precise predictions, but does not seem to eliminate the basic forces we identify. In addition, Section 4.3 develops general versions of our most important predictions about time-inconsistent behavior.

A key question we explore is whether the decisionmaker behaves in a way that she would commit to ex ante. The *ex-ante* or commitment problem is a choice problem in which the consumer makes all investment decisions at time 1, choosing from the set of all lifetime profiles $C^*_{ante} = \{0,1\}^{T_i}$. Applying the dynamic model of focusing introduced in Section 2.2, the individual's focus in period 1 is determined by the range of consumption utilities generated by all possible effort paths. This problem is effectively a static decision problem whose attributes are defined by consumption utility over different dates. In the *ex-post* or no-commitment problem, in contrast, the dynamic model of focusing introduced in Section 2.2 does not reduce to a static problem.[11]

Because of linearity and symmetry, both the ex-ante and ex-post problems have the property that either the consumer strictly prefers to exert effort on all dates, she strictly prefers never to exert effort, or (in a knife-edge case) she is indifferent between all effort paths. Let A^*_{ante} and A^*_{post} be the cutoff levels of A above which the agent chooses to exert effort in the ex-ante and ex-post problems, respectively. A consumption-utility-maximizing agent would set $A^*_{ante} = A^*_{post} = 1$ for all T_i and T_b. Hence, if $A^*_i > 1$ ($i = ante, post$), the decisionmaker is biased toward the present: to make the investment she requires it to be more efficient than a consumption-utility maximizer would. Similarly, if $A^*_i < 1$, the decisionmaker is biased toward the future. Proposition 4 identifies the key patterns in the individual's behavior and bias.

Proposition 4. *Suppose $g(\cdot)$ satisfies Assumption 1, and is continuous.*

1. *The cutoff A^*_{post} does not depend on T_i, is strictly increasing in T_b, and equals 1 if $T_b = 1$.*
2. *The cutoff A^*_{ante} is strictly increasing T_b/T_i, and equals 1 if $T_b = T_i$.*
3. *If $T_i = 1$, $A^*_{ante} = A^*_{post}$, and if $T_i > 1$, $A^*_{ante} < A^*_{post}$.*

[11] In general, our dynamic model requires a theory of the agent's beliefs regarding her own future behavior. However, in the current model, because actions in other periods do not affect the costs and benefits of investment today, all beliefs about future behavior generate the same predictions.

We illustrate the intuition and economic implications of Proposition 4 using a few simple examples.

Example 1 (Exercise)**.** Suppose a consumer is deciding in each of the periods 1 through 100 whether to exercise. Exercising in a period generates pain of 80 in that period and health benefits of 1 in each period 101 through 200. ($T_i = T_b = 100, B = 80, A = 5/4$)

Because $A > 1$, in this example the consumption-utility-maximizing choice is to make the investment, i.e., to exercise. Now consider what a consumer would do in our model with no commitment. Her decision of whether to exercise on any given day can be represented as a choice between $(-80, 1, ..., 1)$ and $(0, ..., 0)$, where 80 is the pain from exercising today, and the 1's represent the health benefits in the 100 future periods. Then, although $A > 1$, the consumer does not exercise on any day if and only if $g(80)/g(1) > 1/0.8 = 1.25$, which may well hold given that $g(\cdot)$ is increasing. Because exercising has a big attention-grabbing current pain and dispersed easy-to-neglect future benefits, the consumer is biased toward the present.

More generally, Part 1 of Proposition 4 says that with no commitment the decisionmaker is biased toward the present for any $T_b > 1$, and is the more so the greater is T_b. In addition, our model never generates future-biased behavior in ex-post choice: for $T_b = 1$, the agent's behavior is consumption-utility-maximizing. Intuitively, for larger T_b the future benefit—but not the current cost—of effort is distributed over more dates and hence is less concentrated, attracting less focus. As a result, the consumer focuses relatively less on the future, creating a present bias in her choice relative to utility maximization.

Now consider what the consumer would commit to in Example 1. As mentioned above, the linearity and symmetry of the problem implies that the relevant comparison is between a lifetime of exercising every day and a lifetime of never exercising. This is equivalent to choosing between $(-80, \ldots, -80, 100, \ldots, 100)$ and $(0, \ldots, 0)$, where the number of both -80's and 100's in the former option is 100. Because each alternative is better than the other on exactly 100 attributes ($T_i = T_b$), we have a balanced choice, so that the consumer commits to always exercising. Intuitively, looking at exercising from the perspective of a lifetime rather than a single day leads the consumer to integrate the incremental health benefits of exercise. With the benefits and costs of exercising now

equally concentrated, the consumer's focus does not distort her choice.

Our next example considers choices when investment is directed toward a concentrated goal.

Example 2 (Writing a book)**.** Suppose a scholar decides in each of the first 100 periods whether to expend effort to write a book, whose benefit is realized in period 101. A period's effort has a utility cost of 50 in that period and increases the utility from the book by 40. ($T_i = 100, T_b = 1, B = 50, A = 0.8$)

The consumption-utility-maximizing choice is not to write the book. This is also the ex post decision, which is now a balanced choice: the cost today is compared with the benefit which accumulates on a single future date. However, judged from an ex-ante perspective, the scholar may commit to writing the book. Intuitively, in her ex-ante choice she focuses on the large concentrated gain from having a completed book, and pays relatively less attention to the dispersed everyday costs. Thus, in contrast to the exercise example, the ex-ante choice is now biased toward the future.

Generalizing Examples 1 and 2, Part 2 of Proposition 4 says that the ex-ante choice is biased toward the future if and only if $T_b < T_i$, and the more so the smaller is T_b/T_i, i.e., the more concentrated are the future benefits relative to the investment costs.

In both the exercise and book examples, the consumer is less future-oriented in her ex-post choice then what she would like to commit to ex ante. As Part 3 of Proposition 4 shows, this time inconsistency is a general feature of the investment model of this section, arising whenever $T_i > 1$. Intuitively, time inconsistency arises because the broader perspective of the ex-ante choice changes the range of costs and benefits. Since (if $T_i > 1$) the benefits from multiple investments accumulate, the range of possible future benefits is larger in the ex-ante than in the ex-post decision. At the same time, the range of investment costs is the same in both choice perspectives. As a result, ex ante the agent focuses relatively more on the future benefits.

The above mechanism generating time-inconsistent behavior is different from that in models of present bias, where time inconsistency arises because the relative weighting of costs and benefits in a period changes once the person is in that period. In particular, unlike present bias, our model predicts that the decisionmaker's commitment and no-commitment choices over any *single decision* are identical. For example, a person who would strongly prefer to commit to exercising

over her entire life may (if she cannot make any other commitment decisions today) commit to *not* exercising on December 12, 2018, because this one-day commitment decision leads to the same focus weights as her decision on that day would. Commitment behavior in single choices thus offers a way to empirically distinguish our theory from existing models of present bias. Although we believe that individuals often have time-inconsistent preferences even for single decisions, we also feel that our time inconsistency based on the shifting breadth of thinking captures an important aspect of intertemporal choice in practice.

By relaxing the assumption that the range of investment costs is the same in the ex-ante and the ex-post problem, our model also predicts intuitively plausible circumstances in which the consumer is *more* future-oriented in her ex-post decision.

Example 3 (Adding Achievements)**.** Consider the decisionmaker from Example 2, and suppose that in period 91, she can decide whether to commit to writing an editorial for the newspaper. Writing has an effort cost of 5 for the next 10 periods, and a benefit of 45 in period 102.

Because $45 < 50$, not writing the editorial is the rational choice. In our model, when looking at the two decisions together ex ante, the effort dimension looms large relative to the benefit of the editorial, as the range of possible efforts in periods $t = 91, ..., 100$ is 0 to 55. Hence, while the scholar wants to commit to writing the book, she would commit to not taking on the editorial assignment. Now suppose that she has committed to writing the book, and makes the decision regarding the editorial in period 91. At this point, whether to take on the new assignment is equivalent to choosing between $(-50, -50, \ldots, -50, 0)$ and $(-55, \ldots, -55, 45)$. Since the *incremental* per-period costs of the editorial are small relative the the large benefit, the agent might commit to the consumption-utility-decreasing option of writing editorial as well.

4.2 Connections to Stylized Facts and Welfare Measures

Lifestyle choices versus achievements. Proposition 4 predicts a time-inconsistent underinvestment into the future when $T_i \approx T_b \gg 0$, and a time-inconsistent overcommitment to investing in the future when $T_b \ll T_i$. The former type of situation matches a number of lifestyle choices, including consumption-savings and borrowing behavior (Laibson 1997, Laibson, Repetto

and Tobacman 2007, Shui and Ausubel 2004, Skiba and Tobacman 2008, Meier and Sprenger 2010, for example), the consumption of harmful products (O'Donoghue and Rabin 1999, Gruber and Kőszegi 2001, Gruber and Kőszegi 2004, O'Donoghue and Rabin 2006, and others), and exercise (DellaVigna and Malmendier 2004, DellaVigna and Malmendier 2006, Acland and Levy 2010), which the literature on present bias has invoked as prime examples of situations in which people act in an overly present-oriented way. The latter prediction is consistent with the view of several researchers, including Scitovsky (1976), Loewenstein, O'Donoghue and Rabin (2003), Kahneman et al. (2006), and Hamermesh and Slemrod (2008), that many people choose overly ambitious careers relative to what would make them happiest. Most closely related to the mechanism in our model, Kahneman et al. (2006) argue that people are too keen to pursue career advancement because they overweight the big potential achievements, such as a big raise or a promotion, relative to the many everyday costs.

Beyond reconciling the above seemingly contradictory patterns, our model makes a number of additional predictions, including (i) that people are more likely to exhibit excessive commitment to working when the costs are more dispersed (e.g., over more days), and the goals are concentrated on fewer attributes; (ii) that people may agree to more such tasks as time goes on; and (iii) that after having made the commitment, people may not feel the daily investments to be worth the cost. Although we are not aware of systematic evidence on these predictions, they are consistent with casual observation. For example, consistent with (i), people often commit to months of hard work for a salary increase, but may not leave their family vacation one day early for a $100 one-time job. Similarly, consistent with (ii) and (iii), many people—including academics—appear to pile on tasks, only to bitterly complain later that this has made them worse off.

Welfare. In addition to providing a unified account of time-inconsistent choice in different domains, our model also yields a simple guide about the proper measure of welfare in different situations. There are two main approaches to measuring welfare in the existing literature on present bias. The most common approach, advocated for instance by DellaVigna and Malmendier (2004), Gruber and Kőszegi (2004), and O'Donoghue and Rabin (2006), is to ignore the taste for immediate gratification and evaluate welfare using long-run preferences. While this criterion has

some good justifications, it ignores that the immediate pleasure at the moment of consumption may be very important. A more cautious approach, proposed originally by Laibson (1997) and further explored by Bernheim and Rangel (2008) and Asheim (2008), is to use the Pareto criterion as applied to the intertemporal incarnations of the individual. Unfortunately, this criterion is often silent about the welfare-optimal choice.

Going beyond both of these approaches, our model yields an unambiguous welfare ranking in all situations without a-priori assuming that the ex-ante perspective is the welfare-relevant one. The guiding principle emerging from our theory is that decisions in more balanced choices are more likely to reflect a person's true well-being. To illustrate how this principle differs from existing approaches, recall Examples 1 and 2 on exercising and writing a book, respectively. In both of these examples, the agent is time-inconsistent in that her ex-post choices are more present-oriented than her ex-ante choices. But while in the exercise example the agent's welfare is aligned with her ex-ante preferences, in the book-writing example it is aligned with her ex-post preferences. This is because the perspective that leads to a balanced choice is different in the two examples. Formally, Part 2 of Proposition 4 says that the ex-ante welfare criterion is appropriate if $T_i = T_b$—a condition approximately satisfied for lifestyle choices—and Part 1 says that the ex-post criterion is appropriate if $T_b = 1$—a condition often approximately satisfied for achievement-type tasks.

Horizon effects. Part 2 of Proposition 4 also implies that the taste for immediate gratification is smaller for T_b low, i.e., if the number of periods in which the consequences of current misbehavior are absorbed is smaller. This prediction distinguishes our model from hyperbolic discounting, in which the temporal concentration of future costs has no effect on present bias. Suppose, for instance, that an employee must decide whether to perform a five-hour job today or put it off to the future. Our model says that she is more likely to perform the job if the alternative is a seven-hour job in six weeks than if the alternative is 10 minutes of extra work each day for the next six weeks. Hyperbolic discounting says that this might not make much difference.[12]

Mental accounting. Finally, our model sheds some light on how mental accounting (Thaler 1985, Thaler 1999) might help overcome the tendency for immediate gratification. As defined by Thaler,

[12] In fact, because the bulk of work is sooner in the latter case, hyperbolic discounting might typically predict that the employee is more likely to do the five-hour job if the alternative is 10 minutes per day for six weeks.

mental accounting is a set of cognitive operations used by individuals to organize and keep track of income, wealth, and expenditure. A central component of mental accounting for our purposes is the assignment of funds to dedicated budgets such as an entertainment budget or food budget. If a consumer keeps such a budget, then the future consequences of spending in a period can be quite concentrated, mitigating present bias. For example, if a consumer has a \$150 monthly budget for entertainment, spending \$30 on a dinner today may mean that she has to forego another dinner relatively soon. Hence, the consequence of saving on the dinner today turns into a single benefit, leading to consumption-utility-maximizing decisions regarding when to go out.[13]

4.3 General Conditions for Time-Inconsistent Behavior

We conclude this section by showing that our central insights about time-inconsistent behavior extend to more general settings. To do so, we introduce some notation and concepts. Let the utility functions in the κ dimensions at time t be $u_{t1}(\cdot)$ through $u_{t\kappa}(\cdot)$, so that $u_t = \sum_{k=1}^{\kappa} u_{tk}$. We denote the history of consumption choices up to period t by $\bar{c}_t = \{c_t, ..., c_1\}$. We allow $u_{tk}(c_t, \bar{c}_{t-1})$ to depend on both current and past consumption choices, but require that the choice set C_t does not depend on previous choices.

For any dynamic problem defined by the choice sets C_1 through C_T, we define the commitment problem, as before, to be the choice of committing to a lifetime consumption profile from the ex-ante perspective. Formally, the agent chooses from the entire path of choices $\tilde{C} = C_1 \times \cdots \times C_T$, expecting to follow through that behavior in each future date. For any given ex-ante committed plan $(c_1^*, \ldots, c_T^*)$, we will also ask whether the decisionmaker would prefer an *ex-post deviation* from the plan in a period t, assuming that in all other past and future periods she follows the plan. Formally, we will ask whether given history $c_1^*, \ldots, c_{t-1}^*$, the decisionmaker would deviate from the choice c_t^* in the continuation problem $C_t \times \{c_{t+1}^*\}, \ldots, \times \{c_T^*\}$.

Our first result identifies a general set of circumstances under which the agent prefers to behave in a more present-oriented way ex post than she would have committed to ex ante. We say that the

[13] Our model does not say whether and how such mental-account-based consumption rules can be credible; it merely identifies a way in which they can be useful for overcoming the taste for immediate gratification *if* they are credible.

current utility of a current action is not controlled by previous choices if, whenever u_{tk} depends on c_t, it does not depend on the history $\bar{c}_{t-1}$. We say that period-t choice has a unidirectional effect on future utility if for any $c_t, c_t' \in C_t$, either $u_{tk}(c_t', \bar{c}_{t-1}) \geq u_{tk}(c_t, \bar{c}_{t-1})$ for all k and $u_{\tau k}(\bar{c}_{\tau,-t}, c_t') \leq u_{\tau k}(\bar{c}_{\tau,-t}, c_t)$ for all $\tau > t$, k, and feasible c_{t+1} through c_T, or vice versa. Both of these conditions are satisfied for example by the investment problem of Section 4.1.

Proposition 5 (Present-Oriented Time Inconsistency)**.** *Suppose $g(\cdot)$ satisfies Assumption 1. Take any $t \in \{2, \ldots, T\}$, and suppose that (i) the current utility of current actions in period t is not controlled by previous choices; and (ii) period-t choice has a unidirectional effect on future utility. Then the decisionmaker would prefer an ex-post choice in period t which yields at least as high period-t utility as her ex-ante plan.*

Because of (i), the decisionmaker pays the same attention to the period-t utility consequences of c_t when she commits ex ante and when she makes her choice in period t. Her attention on utility in future periods $\tau > t$ is determined by C_τ in the ex-ante choice, but may be determined by a smaller range of options at time t. By (ii), her choice at time t is a clear trade-off between the present and the future. Since she pays relatively less attention to the future at time t than ex ante, it follows that she will act in a more present-oriented manner.

We next identify conditions under which the decisionmaker can be prone to time inconsistency in future-oriented behavior, whereby she keeps becoming more and more ambitious in her quest for achievements. We say that a future attribute τ^*, k^* is fully controlled by period-t choice if $u_{\tau^* k^*}$ depends only on c_t. This attribute is said to have a unidirectional effect on the utility of other attributes if, for any c_t and c_t' such that $u_{\tau^* k^*}(c_t') > u_{\tau^* k^*}(c_t)$, for any other τ, k with $\tau \geq t$ it must be that $u_{\tau,k}(c_\tau, c_{\tau-1}, \ldots, c_{t+1}, c_t', c_{t-1}, \ldots, c_1) \leq u_{\tau,k}(c_\tau, c_{\tau-1}, \ldots, c_{t+1}, c_t, c_{t-1}, \ldots, c_1)$.

Proposition 6 (Future-Oriented Time Inconsistency)**.** *Suppose $g(\cdot)$ satisfies Assumption 1. Consider any $t \in \{1, \ldots, T-1\}$, and suppose that (i) there is a future attribute (τ^*, k^*) fully controlled by period-t choice, and (ii) this attribute has a unidirectional effect on the utility of other attributes. Then the decisionmaker would prefer an ex-post choice in period t which yields at least as high utility in attribute (τ^*, k^*) as her ex-ante plan.*

This result generalizes Example 3. Thinking of the future utility dimension (τ^*, k^*) as an achievement, the statement means that people are too prone to take on an additional task than they would have preferred earlier. Intuitively, because of (i) the range of utility from the achievement is the same ex ante and ex post, while the range of costs might be smaller ex post. Thus the benefit from the achievement looms larger in relative terms ex post than ex ante. Since by (ii) the individual faces a clean trade-off between the benefit and the costs, she will be more willing to take on the additional task ex post.

5 Product Design and Pricing

The idea that firms design and position products to attract consumer attention is central to both academic and applied marketing research. By endogenizing consumers' focus as a function of available products, our model provides a way to incorporate this key marketing idea into economics. In this section we begin to develop this approach by analyzing the behavior of rational profit-maximizing firms when consumers' focus is affected by the bias toward concentration. While a full analysis of the issues we raise is beyond the scope of this paper, we formalize and qualify a number of insights from marketing in a stylized setting, and derive several new predictions.

5.1 Product Design Under Monopoly

We first explore how a monopolist designs and prices products to take advantage of consumers' bias toward concentration. Suppose that a monopolist can sell exactly one item of one product to a consumer with focus-dependent behavior. The product is represented by a vector $(v_1, ..., v_{N_1}, -p_1, ..., -p_{N_2})$, where the first N_1 dimensions measure consumption value over different product attributes, and the last N_2 dimensions are price (or other cost) attributes to the consumer. We assume that consumption utility over all attributes is linear, and that the outside option of the consumer is not purchasing the product, which is represented by the vector $(0, ..., 0)$. The monopolist has discretion in how it designs and prices its product, and chooses $v_i, p_j \geq 0$ to maximize

$$\sum_{i=1}^{N_2} p_i - \phi\left(\sum_{i=1}^{N_1} v_i\right).$$

Hence, the firm can produce each of the value dimensions at the same marginal cost, and cares about the total price the consumer pays. We assume that the cost function $\phi(\cdot)$ is twice continuously differentiable and satisfies $\phi'(0) = 0, \phi''(c) > 0$ for all $c \geq 0$, as well as $\lim_{c\to\infty} \phi'(c) = \infty$.

Before turning to results, we define, for any utility difference Δ, the attention-weighted difference $G(\Delta) \equiv g(\Delta)\Delta$. Under Assumption 1, $G(\Delta)$ is strictly increasing and convex near zero. For Proposition 7 below, we will make the natural assumption that the convexity of $G(\Delta)$ extends to the relevant range of medium stakes. Intuitively, when the available options differ more rather than less in a given attribute, how a further increase in the differences in this attribute affects the difference in perceived utility is determined by three effects. First, since focus on the attribute is greater, the increase in the difference counts more heavily in perceived utility. Second, the increase in attention applies to a greater utility difference, again increasing the difference in perceived utility. Third, depending on the shape of $g(\cdot)$, the increase in attention may be smaller, decreasing the effect on perceived utility. The assumption that $G(\cdot)$ is convex over a range amounts to assuming that in as much as there is such a negative effect, the former effects outweigh it. This is especially plausible since for moderate-range stakes, one might expect the first direct effect to dominate.

Proposition 7 identifies how the monopolist responds to the bias toward concentration:

Proposition 7. *Suppose $g(\cdot)$ satisfies Assumption 1 and is differentiable, and $G(\Delta)$ is convex. The optimal product has $v_i > 0$ for one value attribute and $v_j = 0$ for all other $j \neq i$ value attributes; and has $p_1 = p_2 = \cdots = p_{N_2} > 0$. If $N_2 = 1$, the level of value production is efficient ($\phi'(v) = 1$), and if $N_2 > 1$, there is overproduction of value ($\phi'(v) > 1$).*

Concentrating the value of the product into a single attribute attracts attention to the quality of the product; while evenly distributing the cost into many dimensions dilutes the focus on its price. To understand the efficiency implications, first note that with a single price dimension ($N_2 = 1$), the consumer faces a balanced choice, and because she acts rationally in these situations, the monopolist has an incentive to choose the efficient quality level. With at least two price dimensions, however, the consumer focuses more on the single large value than the many small prices, and in response the monopolist produces "too much" value.[14]

[14] While the sharp form of the results in Proposition 7 relies on our assumptions that attributes are symmetric

5.2 Competition and Relative Positioning

We next explore how competition shapes product design and product positioning relative to competitors. One strategy firms often seem to attempt in practice is to build products with a unique attribute in which other firms are perceived to be weak. For example, Volvos are known for their safety and BMWs for their performance. This strategy echoes the result of Proposition 7 that a firm should concentrate its effort on one distinguishing attribute, and is consistent with common advice on brand positioning. At the same time, some firms seem to attempt the exact opposite strategy—they try to copy the identifying attribute of the competitor. For instance, since the introduction of the iPhone, touch-screen mobile phones have become ubiquitous. This leads to the natural question: when should a firm copy the flagship attribute of its competitor, and when should it differentiate by emphasizing an attribute in which the competition is weak?

To address the above question, we consider a simple model of how a firm responds to a competitor's exogenously given product. We assume that products have two quality attributes ($N_1 = 2$) and a single price dimension ($N_2 = 1$). Firm 1 chooses v_1, v_2, p, where v_1 and v_2 are the attributes and p is the price. The total cost of producing qualities v_1 and v_2 is $\phi(v_1 + v_2)$. Firm 2 sells an exogenously given competing product with qualities $\bar{v}_1$ and $\bar{v}_2 < \bar{v}_1$ and price $\bar{p}$. The consumer's consideration set includes these two products, as well as the option of buying neither product.

Proposition 8 characterizes how firm 1 designs its product depending on how its overall quality compares to firm 2's product.

Proposition 8. *Suppose $g(\cdot)$ satisfies Assumption 1 and is differentiable.*

(1) If $v_i < \bar{v}_i$ for $i = 1, 2$, then $v_1 > v_2$.

(2) If $v_1 + v_2 > \bar{v}_1 + \bar{v}_2$, $g(\cdot)$ is unbounded, and either $v_1 + v_2$ is large enough or $\bar{v}_2$ is small enough, then $v_1 < v_2$.

Part (1) says that when firm 1's product is dominated by firm 2's product, then firm 1 chooses as its stronger attribute firm 2's stronger attribute. Intuitively, in this case firm 2's product determines the consumer's focus, so firm 1 is effectively maximizing profits with the consumer's focus held fixed.

and perfectly substitutable in production and consumption, the basic mechanism of concentrating advantages and separating costs to bias focus toward buying the product is independent of these assumptions.

Hence, it is optimal for firm 1 to concentrate on the attribute to which the consumer pays more attention, which is firm 2's stronger dimension.

Part (2) of Proposition 8 says that if firm 1 is sufficiently better than firm 2, it will focus more on the attribute in which its opponent is weaker. To understand the intuition, suppose that firm 2 only produces on attribute 1 ($\bar{v}_2 = 0$), and consider firm 1's choice of whether to allocate a given value v to attribute 1 or attribute 2. If $v < \bar{v}_1$, then allocating to attribute 1 is optimal: it maximizes the focus on firm 1's product without affecting the focus on the opponent. If $v > \bar{v}_1$, however, then allocating the value v to attribute 2 becomes optimal: this draws attention to the weakness of its opponent and (by lowering the weight on attribute 1 from $g(v)$ to $g(\bar{v})$) reduces focus on the opponent's strength.[15]

This result provides a simple explanation for why firms sometimes position products with respect to competitors' weaknesses and sometimes copy competitors' strengths. Firms who are not good enough to produce value that shifts consumer attention should copy attributes where the competition is strongest, because those attributes enjoy the highest attention. To state the logic in the context of cell phones, if a firm cannot shift attention with a new attribute, it should concentrate on having a touch-screen, which is what consumers focus on. In contrast, good firms, who can provide value that shifts attention, should concentrate on the weakness of the competitor.

Having analyzed a firm's incentive to differentiate its product from those of competitors, to conclude this section we explore the resulting quality levels in "attentionally differentiated" and "attentionally undifferentiated" competition. In the former case, $N_1 = 2$, and firm $i = 1, 2$ can only produce attribute i. In the latter case, there is a single value dimension in which firms compete ($N_1 = 1$).[16] In both cases, we assume that there is only one payment dimension ($N_2 = 1$), and that the consumer's consideration set includes both options as well as not buying. Firms simultaneously

[15] The additional assumptions in Part 2 of the proposition address the possibility that firm 2's value on attribute 2 is positive ($\bar{v}_2 > 0$), but "just under the radar" of consumer attention. In this case, switching to that attribute draws attention to firm 2 as well. However, when $\bar{v}_2$ is close to zero this effect is small, while when $v_1 + v_2$ grows the effect in the text becomes large, as long as the focusing weight is unbounded.

[16] We have also explored the case when firms choose not only what quality level to produce, but also which dimension to produce on. In this case, there is no pure-strategy equilibrium. If firms choose the same dimension, then by the Proposition quality would be relatively low—but then switching to the other dimension and increasing quality is profitable. If firms produce on different dimensions, then quality would have to be high, but then switching to the other firm's dimension and lowering quality is a profitable deviation.

choose quality levels and price. Then:

Proposition 9. *Suppose $g(\cdot)$ satisfies Assumption 1 and is differentiable. For any symmetric pure-strategy equilibria, value is higher with attentionally differentiated than with attentionally undifferentiated competition, and in either case higher than under monopoly.*

The difference between differentiated and undifferentiated competition arises from how an increase in a firm's quality affects the consumer's focus. With differentiated competition, an increase in firm 1's quality draws attention to attribute 1, benefitting firm 1 and hurting firm 2. With undifferentiated competition, however, an increase in firm 1's quality either has no effect on the consumer's attention (if firm 2's product is superior), or it draws more attention to quality (if firm 1's product is superior), but in both cases it has the same effect on the two firms. As a result, firm 1 has a greater incentive to increase quality with attentionally differentiated competition. The second part of Proposition 9, that both types of competition generate higher quality than a monopoly, and therefore also higher quality than is optimal, is due to the effect of competition on prices. Since competition lowers prices, it shifts focus to quality, increasing firms' incentives to produce higher quality.

More generally, this result says that in industries in which product differentiation is easier, there will be higher production of value because of the added gain from directing consumers' attention toward one's own strength. While we are not aware of a formal test of this prediction, it seems both intuitively reasonable and consistent with anecdotal examples. For instance, traditional telephones were difficult to differentiate, and the technology did not change for decades. However, following the introduction of mobile phones—which turned the phone into a personal item and opened a greater scope for differentiation—telephone technology improved rapidly.

6 A Model of Consideration Sets

In this section, we return to an important but so far underemphasized assumption of our model: that the decisionmaker's focus and choices are based on an exogenously given consideration set that could be different from her choice set. Our motivation for allowing the consideration set to

be different from the choice set is that manifestly irrelevant options in the consideration set have seemingly unrealistic effects on behavior. As a simple example, consider a choice set consisting of three options with two attributes each: $(0, 10)$, $(11, 0)$, and $(-1, -1000)$. If the consideration set was the entire choice set, the decisionmaker would put a very large weight on attribute 2, possibly leading her to choose the first option. It seems unlikely, however, that a person would focus so much on attribute 2 just because she has a terrible option on that attribute. For instance, a consumer who is looking to buy a new computer with a monitor will not focus more on the quality of monitors just because there are computers available without monitors. Hence, it seems more realistic to think of the decisionmaker's consideration set as being made up of only the first two options.

More generally, in most economic settings a person has "self-destructive" choices that lower her utility in all respects, and these typically do not enter her consideration set—or often even her consciousness. While we feel that the approach of basing our model on an exogenously given consideration set consisting of the decisionmaker's reasonable alternatives is satisfactory for many applications, for a complete theory of focus-driven choice it is necessary to define consideration sets endogenously. Because we are not aware of much evidence that would help guide such a formulation of consideration sets, we tentatively propose a theory as a first pass. But our theory above of how a person chooses from a consideration set can be combined with any theory of how the consideration set is formed.

The basic idea behind our formulation is simple, and builds on the premise of our theory that focus will be drawn to large differences. We amend this force by also assuming that a person's attention is drawn to better rather than worse options. The consideration set, then, will be determined as the set that maximizes a combination of consumption utility and tradeoffs—that is, generates the most attractive tradeoffs.

Formally, suppose that the decisionmaker's full choice set is $\overline{C}$, and let $\alpha \geq 1$. We normalize the consumption-utility-maximizing option in $\overline{C}$ such that it has a consumption utility of zero. We

assume that the consideration set is given by the set $C \subset \overline{C}$ that solves

$$\max_{C \subset \overline{C}} \frac{1}{|C|} \left[\alpha \sum_{c \in C} U(c) + \sum_{k=1}^{K} \Delta_k(C) \right]. \tag{2}$$

This definition has a number of potentially attractive properties. First, there always exists a consideration set, and generically, it is unique. Second, if $\alpha > 1$, c^1 dominates c^2, and c^2 is in the consideration set, then so is c^1; otherwise, replacing c^2 by c^1 strictly increases the maximand in Equation 2. This also implies that if $\alpha > 1$, dominated choices can never constitute the consideration set. Third, if $\alpha = 1$ and there are two dimensions, the consideration set always includes the extreme options of the Pareto set.

Our theory of consideration-set determination is distantly related to that of Eliaz, Richter and Rubinstein (2010), who axiomatically characterize procedures for choosing a two-element consideration set (or "finalists"). Their "top two" criterion selects the best options according to some ordering, while their "extremes" criterion selects the two extremes according to some ordering. In trading off high utility and differences, our specification can be thought of as a combination of these two criteria.[17]

With the endogenous theory of consideration sets, our model has a feature that makes it easier and methodologically less problematic to apply: once the function $g(\cdot)$ and the weight α are specified, applying our model to a new economic setting requires only the specification of relevant attributes in that setting in addition to assumptions (regarding the choice set and utility function) that a classical model would already make. Hence, our theory provides a way to translate a classical model into one with focus-sensitive choice with very little degrees of freedom.

Context effects in purchasing. The following example illustrates the use of our consideration-set definition, and also identifies a new implication of our model. Consider a consumer who makes a purchase decision with five attributes, four product attributes and money. Suppose first that there are two options, not buying and buying one particular product: $\overline{C} = \{(0,0,0,0,0), (20,20,0,0,-35)\}$. For any $\alpha < 15$, the consideration set contains both options ($C = \overline{C}$). As a result, money looms large, and the consumer might not buy the product—even though buying it would maximize con-

[17] In an approach orthogonal to ours, the marketing literature, including Hauser and Wernerfelt (1990) and Roberts and Lattin (1991), has explored search-theory based models for consideration set formation.

sumption utility. By a similar argument, the consumer might not buy if she has a choice between not buying and buying a more expensive product: $\overline{C} = \{(0,0,0,0,0),(0,0,22,22,-40)\}$. But now suppose that *both* products are in her choice set: $\overline{C} = \{(0,0,0,0,0),(20,20,0,0,-35),(0,0,22,22,-40)\}$. In this case, for any $\alpha \geq 1$ the consideration set consists of only the last two options. Since money does not loom as large in this comparison, the decisionmaker might choose the more expensive option—even though in terms of consumption utility it is not worth the extra money.

Intuitively, when there are multiple products with sufficiently different attributes, the consumer's attention is drawn to the tradeoffs between the products rather on the tradeoffs between buying and not buying. With the consumer considering *what* to buy rather than *whether* to buy, our model of focus implies that she will tend to care less about money, so that she is more likely to buy an expensive product.

7 Conclusion

By virtue of defining focus-dependent utility based on consumption utility and the decisionmaker's choice set, our model opens the way for analyzing the role of focusing in many economic settings using one generally applicable model. In addition to the questions in intertemporal choice and product design we have started to address in this paper, there seem to be many additional potential applications. An employee may be motivated by some features of her employment contract—such as a bonus, major promotion, or other large goal—not only because they can generate higher consumption utility, but also because of her disproportionate focus on these features. A firm may recognize and take advantage of consumers' distorted focus in specific situations, such as when selling an add-on to a consumer who is focusing too much on her current purchase. And just like marketers, political parties may attempt to manipulate voters by the positioning of their candidates in part through attempting to attract focus to their candidates' strengths.

Being defined over riskless choices only, our model is not applicable to situations in which uncertainty is a central part. In ongoing work (Kőszegi and Szeidl 2011, hopefully), we develop an extension of our framework to uncertainty by thinking of states as different attributes, resulting in a model similar to that of Bordalo et al. (2010). Incorporating risk into our framework will allow us

to apply it to a number of additional applications. For instance, because a large possible monetary loss in case of an accident attracts focus more than insurance payments in other states, a person may be too prone to buy insurance.

In addition, in some situations the psychological principle motivating our theory—that big differences attract attention—leads to a different reduced-form model than ours. For example, the agent's focus might be affected by options (such as her status quo or salient alternative) that are not in her choice set. More distantly related to our model, focus seems to depend on recent changes in circumstances—that is, comparisons between a person's previous and current situation. In particular, consistent with evidence by DellaVigna and Pollet (2007) that investors do not notice slowly-moving changes and with the model of (Madarász and Tasoff 2009), larger changes in expectations regarding an attribute attract disproportionately more attention.

References

Aaker, David A., *Managing Brand Equity*, New York: The Free Press, 1991.

Acland, Dan and Matthew R. Levy, "Habit Formation, Naivete, and Projection Bias in Gym Attendance," 2010. Working Paper, Harvard University.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson, "The Age of Reason: Financial Decisions Over the Lifecycle," 2007. MIT Department of Economics Working Paper No. 07-11.

___ , John C Driscoll, Xavier Gabaix, and David Laibson, "Learning in the Credit Card Market," 2008. NBER Working Paper #13822.

Asheim, Geir B., "Procrastination, Partial Naivete, and Behavioral Welfare Analysis," 2008. Working Paper, University of Oslo.

Bernheim, B. Douglas and Antonio Rangel, "Beyond Revealed Preferences: Choice Theoretic Foundations for Behavioral Welfare Economics," January 2008. NBER Working Paper No. 13737.

Bertaut, Carol C., Michael Haliassos, and Michael Reiter, "Credit Card Debt Puzzles and Debt Revolvers for Self Control," *Review of Finance*, 2009, *13* (4), 657–692.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience Theory of Choice Under Risk," 2010. Working Paper, Harvard University.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

Chiu, Andrew and George Wu, "Prospect Theory," 2009. Entry for *Wiley Encyclopedia of Operations Research and Management Science*.

DellaVigna, Stefano and Joshua M. Pollet, "Demographics and Industry Returns," *American Economic Review*, 2007, *97* (5), 1667–1702.

___ **and Ulrike Malmendier**, "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, 2004, *119* (2), 353–402.

___ **and** ___ , "Paying Not to Go to the Gym," *American Economic Review*, 2006, *96* (3), 694–719.

Dunn, Elizabeth W., Timothy D. Wilson, and Daniel T. Gilbert, "Location, Location, Location: The Misprediction of Satisfaction in Housing Lotteries," *Personality and Social Psychology Bulletin*, 2003, *29*, 1421–1432.

Eliaz, Kfir, Michael Richter, and Ariel Rubinstein, "Choosing the Two Finalists," *Economic Theory*, 2010, *46* (2), 211–219.

Federal Communication Commission, "Comment Sought on Measures Designed to Assist U.S. Wireless Consumers to Avoid "Bill Shock"," Technical Report 2010. Public Notice May 11 2010. CG Docket No. 09-158, `http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-10-803A1.pdf`.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg, "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 2006, *96* (4), 1043–1068.

Genesove, David and Chris Mayer, "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 2001, *116* (4), 1233–1260.

Gruber, Jonathan and Botond Kőszegi, "Is Addiction 'Rational?' Theory and Evidence," *Quarterly Journal of Economics*, 2001, *116* (4), 1261–1305.

___ **and** ___ , "Tax Incidence when Individuals are Time-Inconsistent: The Case of Cigarette Excise Taxes," *Journal of Public Economics*, 2004, *88* (9-10), 1959–1987.

Hamermesh, Daniel S. and Joel B. Slemrod, "The Economics of Workaholism: We Should Not Have Worked on This Paper," *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy*, 2008, *8* (1).

Hauser, John R. and Birger Wernerfelt, "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 1990, *16* (4), 393–408.

Highfill, Jannett, Douglas Thorson, and William V. Weber, "Tax Overwithholding as a Response to Uncertainty," *Public Finance Review*, 1998, *26* (4), 376–391.

Johnson, Eric J., John C. Hershey, Jacqueline Meszaros, and Howard Kunreuther, "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty*, 1993, *7* (1), 35–51.

Jones, Damon, "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds," 2010. NBER Working Paper No. 15963.

Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone, "Would You Be Happier If You Were Richer? A Focusing Illusion," *Science*, 2006, *312* (5782), 1908–1910.

___ **and Amos Tversky**, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Kőszegi, Botond and Ádám Szeidl, "Attention and Compensation," 2011, hopefully. In progress.

Laibson, David I., "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 1997, *112* (2), 443–477.

___ **, Andrea Repetto, and Jeremy Tobacman**, "Estimating Discount Functions with Consumption Choices over the Lifecycle," 2007. NBER Working Paper #13314.

Lehenkari, Mirjam, "The Hedonic Editing Hypothesis: Evidence from the Finnish Stock Market," *Journal of Behavioral Finance*, 2009, *10* (1), 9–18.

Linville, Patricia W. and Gregory W. Fischer, "Preferences for Separating or Combining Events," *Journal of Personality and Social Psychology*, 1991, *60* (1), 5–23.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin, "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 2003, *118* (4), 81–123.

Madarász, Kristóf and Joshua Tasoff, "A Model of Attention and Anticipation," 2009. Working Paper, Claremont Graduate University.

Meier, Stephan and Charles Sprenger, "Present-Biased Preferences and Credit Card Borrowing," *American Economic Journal: Applied Economics*, 2010, *2* (1), 193–210.

Odean, Terrance, "Are Investors Reluctant to Realize Their Losses?," *Journal of Finance*, 1998, *53* (5), 1775–1798.

O'Donoghue, Ted and Matthew Rabin, "Addiction and Self-Control," in Jon Elster, ed., *Addiction: Entries and Exits*, Russel Sage Foundation, 1999.

___ **and** ___ , "Optimal Sin Taxes," *Journal of Public Economics*, 2006, *90* (10-11), 1825–1849.

Rabin, Matthew and Georg Weizsäcker, "Narrow Bracketing and Dominated Choices," 2009. forthcoming, American Economic Review.

Reis, Ricardo, "Inattentive Producers," *Review of Economic Studies*, 2006, *73* (3), 793–821.

Roberts, John H. and James M. Lattin, "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 1991, *28* (4), 429–440.

Salant, Yuval, "Procedural Analysis of Choice Rules with Applications to Bounded Rationality," *American Economic Review*, forthcoming.

Schkade, David A. and Daniel Kahneman, "Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction," *Psychological Science*, 1998, *9* (5), 340–346.

Scitovsky, Tibor, *The Joyless Economy: The Psychology of Human Satisfaction*, Oxford University Press, US, 1976.

Shui, Haiyan and Lawrence M. Ausubel, "Time Inconsistency in the Credit Card Market," 2004. Available at http://ssrn.com/abstract=586622.

Sims, Christopher A., "Rational Inattention and Monetary Economics," in Benjamin A. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3A, Elsevier, 2010, chapter 4.

Skiba, Paige Marta and Jeremy Tobacman, "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default," 2008. Vanderbilt Law and Economics Research Paper No. 08-33.

Stango, Victor and Jonathan Zinman, "What Do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs," *American Economic Review Papers and Proceedings*, 2009, *99* (2), 424–429.

___ and ___ , "Fuzzy Math, Disclosure Regulation, and Market Outcomes: Evidence from Truth-in-Lending Reform," *Review of Financial Studies*, 2011, *24* (2), 506–534.

Thaler, Richard, "Mental Accounting and Consumer Choice," *Marketing Science*, 1985, *4* (3), 199–214.

Thaler, Richard H., "Psychology and Savings Policies," *American Economic Review*, 1994, *84* (2), 186–192.

___ , "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 1999, *12* (3), 183–206.

___ **and Eric J. Johnson**, "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science*, 1990, *36* (6), 643–660.

Tversky, Amos and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 1981, *211*, 453–458.

Woodward, Susan E. and Robert E. Hall, "Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence," 2010. Working Paper, Stanford University.

A Proofs (Preliminary and Sketchy!)

Proof of Proposition 1. Immediate. □

Proof of Proposition 2. $1 \Rightarrow 2$. The attention-weighted utility difference between $c^1$and c^2 is

$$\begin{aligned}\tilde{U}\left(c^1, C\right) - \tilde{U}\left(c^2, C\right) &= \sum_{i \in I} g(\Delta_i) \cdot \Delta_i - \sum_{j \in J} g\left(\Delta_j\right) \cdot \Delta_j \\ &\geq g(\min_I \Delta_i) \cdot \sum_{i \in I} \Delta_i - g(\max_J \Delta_j) \cdot \sum_{j \in J} \Delta_j > g(\min_I \Delta_i) \cdot \sum_{i \in I} \Delta_i - g(\min_I \Delta_i) \cdot \sum_{j \in J} \Delta_j = 0.\end{aligned}$$

The key step is the strict inequality, where we use that $\min_I \Delta_i > \max_J \Delta_j$, i.e., that c^1 has more concentrated advantages than disadvantages relative to c^2.

$2 \Rightarrow 1$. Suppose there are $\Delta \geq 0$ and $\Delta' > \Delta$ rational numbers such that $g(\Delta') \leq g(\Delta)$. Then, for a sufficiently large K we can construct a c^1 and c^2 such $\Delta_i = \Delta'$ for $i \in I$ and $\Delta_j = \Delta$ for $j \in J$. This pair contradicts the statement of Part 2. By continuity, $g(\cdot)$ is everywhere strictly increasing. □

Proof of Corollary 1. It is sufficient to prove that for any positive Δ and Δ', we have

$$g(\Delta + \Delta')(\Delta + \Delta') > g(\Delta)\Delta + g(\Delta')\Delta'.$$

This is obvious since $g(\cdot)$ is increasing, so that $g(\Delta + \Delta') > g(\Delta), g(\Delta')$. □

Proof of Proposition 3. The attention-weighted utility difference between the two options is

$$\epsilon \cdot \sum_{k=1}^{K} g(\epsilon \cdot \delta_k) \cdot \delta_k.$$

Dividing by ϵ and letting $\epsilon \to 0$ gives

$$g(0) \cdot \sum_{k=1}^{K} \delta_k.$$

Since the sum of δ_k determines the consumption-utility-maximizing choice, we conclude that for ϵ small, the decisionmaker maximizes consumption utility. □

Proof of Proposition 4.

1. Attention on dimension t is set by $\Delta_t = B$, while attention on dimensions $s = T_i{+}1, \ldots, T_i{+}T_b$ is determined by $\Delta_s = AB/T_b$. The consumer strictly prefers to invest if $g(B) < ABg(AB/T_b)$. Since the right-hand side is continuous and increasing in A, the value A^*_{post} is defined by $g(B) = A^*B \cdot g(A^*B/T_b)$. It is immediate that A^*_{post} is independent of T_i. Given that $g(\cdot)$ is strictly increasing we also have that A^*_{post} is strictly increasing in T_b and equals 1 for $T_b = 1$.

2. Since all possible effort paths are considered, focus on the investment dimensions is set by $\Delta_t = B$, $t = 1, \ldots, T_i$ while focus on the benefit dimensions is defined by $\Delta_s = AB \cdot T_i/T_b$ for $s = T_i+1, \ldots, T_i+T_b$. Effort is strictly preferred if and only if $g(B) < AB \cdot g(AB \cdot T_i/T_b)$, and A^*_{ante} is defined by $g(B) = A^*B \cdot g(A^*B \cdot T_i/T_b)$. The claims now all follow from the strict monotonicity of $g(\cdot)$.

3. Immediate from the above expressions. □

Proof of Proposition 5. Consider $c'_t \in C_t$ such that $u_t(c'_t, \bar{c}^*_{t-1}) \leq u_t(c^*_t, \bar{c}^*_{t-1})$. We show that in the ex-post decision of period t, the decisionmaker prefers c^*_t to c'_t. This in turn implies the statement of the proposition that the decisionmaker would prefer at least as high period-t utility ex post as she does ex ante.

Denote by $\tilde{K}$ the set of attributes of period t utility which are affected by c_t. Both ex ante and ex post, the choice of c_t only affects period-t attributes in $\tilde{K}$ and future attributes in periods $\tau > t$. The fact that ex ante the agent chooses c^*_t rather than c'_t means that

$$\sum_{k \in \tilde{K}} g(\Delta_{tk}) \left[u_{tk}(c^*_t, \bar{c}^*_{t-1}) - u_{tk}(c'_t, \bar{c}^*_{t-1})\right] \geq \sum_{k,\tau>t} g(\Delta_{\tau k}) \left[u_{\tau k}(c'_t, \bar{c}^*_{\tau,-t}) - u_{\tau k}(c^*_t, \bar{c}^*_{\tau,-t})\right].$$

By (ii), all the utility differences are non-negative on both sides. Now consider how the focusing weights change in this inequality in the ex-post choice. By (i), for each $k \in \tilde{K}$, the range of u_{tk} and hence the value of $g(\Delta_{tk})$ is the same in the ex-ante and the ex-post problem. In contrast, the values $\Delta_{\tau k}$ may be smaller ex post, when not all ex ante feasible consumption paths are considered. Thus the left hand side is unchanged while the right hand side is weakly smaller in the ex-post decision, and the inequality continues to hold. □

Proof of Proposition 6. Consider $c'_t \in C_t$ such that $u_{\tau^* k^*}(c'_t) \leq u_{\tau^* k^*}(c^*_t)$. We show that in the ex-post decision of period t, the decisionmaker prefers c^*_t to c'_t. This in turn implies the statement of the proposition that the decisionmaker would prefer at least as high utility in attribute (τ^*, k^*) ex post as she does ex ante.

The fact that ex ante the agent chooses c^*_t rather than c'_t means that

$$g(\Delta_{\tau^* k^*}) \left[u_{\tau^* k^*}(c^*_t) - u_{\tau^*, k^*}(c'_t)\right] \geq \sum_{(\tau,k) \neq (\tau^*, k^*), \tau > t} g(\Delta_{\tau k}) \left[u_{\tau k}(c'_t, \bar{c}^*_{\tau,-t}) - u_{\tau k}(c^*_t, \bar{c}^*_{\tau,-t})\right].$$

By (ii) all terms are positive. By (i), the left-hand side is unchanged in the ex-post choice. Since the right-hand side is weakly smaller, the claim follows. □

Proof of Proposition 7. The monopolist maximizes

$$\sum_{i=1}^{N_2} p_i - \phi \left(\sum_{i=1}^{N_1} v_i \right)$$

subject to

$$\sum_{i=1}^{N_2} G(p_i) \leq \sum_{i=1}^{N_1} G(v_i).$$

Denote $v = \sum_{i=1}^{N_1} v_i$. For any fixed v, Proposition 2 part 3 shows that the right-hand side of the constraint is maximized by setting $v_i = v$ for some i and $v_j = 0$ for $j \neq i$. The convexity of $G(\cdot)$ implies that for a fixed value of $\sum_{i=1}^{N_2} p_i$, the left hand side of the constraint is minimized when $p_i = p/_j$ for all prices. It follows that the monopolist's problem is equivalent to maximizing

$$N_2 \cdot p - \phi(v)$$

subject to

$$G(v) = N_2 \cdot G(p).$$

Because $\phi'(0) = 0$, producing some value is better than producing no value. Because ϕ' is unbounded, it is not hard to see that producing a very large value is not optimal. It follows that this problem has an interior solution. Differentiability of the objective function and the constraint implies that the optimum must satisfy the first order condition.

Because $g(\cdot)$ is strictly increasing, the constraint implies that $v = p$ if $N_2 = 1$ and $v > p$ if $N_2 > 1$. Using the Lagrangian, the first order condition of this problem is

$$\phi'(v) = \frac{G'(v)}{G'(p)}.$$

This expression equates the marginal cost and marginal benefit of increasing value. The right hand side is the marginal benefit, because it measures by how much can prices be increased to keep the consumer indifferent, accounting for the differences in focus between the value and price attributes. For $N_2 = 1$ and $v = p$ this equation gives the efficient choice of value, while for $N_2 > 1$ and $v > p$ it implies overproduction since $G(\cdot)$ is strictly increasing. □

Proof of Proposition 8 (i) By assumption, the focusing weights are set by $\bar{v}_1$ and $\bar{v}_2$, and hence the consumer focuses more on attribute 1. Suppose that $v_1 < v_2$, and consider the deviation of increasing the first attribute and reducing the second by ϵ. This does not change the cost to firm 1, and if ϵ is small, also does not affect the focusing weights and hence the perceived value of firm 2's product. But this change does moves more of firm 1's value on the attribute which enjoys more attention, increasing the perceived value of the product.

(ii) Suppose by way of contradiction that $v_1 < v_2$. We will distinguish three cases. (a) Firm 1 is better only in the first attribute: $v_1 > \bar{v}_1$ but $v_2 \leq \bar{v}_2$. (b) Firm 1 is better only in the second attribute: v_1 $leq\bar{v}_1$ and $v_2 > \bar{v}_2$. (c) Firm 1's product dominates firm 2: $v_1 > \bar{v}_1$ and $v_2 > \bar{v}_2$. We will only work out case (a) here; the other two cases are simpler. In all three cases, we consider the deviation where zero value is allocated to the first attribute and $v_1 + v_2$ is allocated to the second. This deviation has the same cost as before, and it is straightforward to check that it is profitable in cases (b) and (c) as long as $v_1 + v_2 > \bar{v}_1 + \bar{v}_2$. Consider case (a). The perceived value difference between the two firms' products before the deviation is

$$g(v_1)(v_1 - \bar{v}_1) + g(\bar{v}_2)(v_2 - \bar{v}_2)$$

while after the deviation it equals

$$g(\bar{v}_1)(-\bar{v}_1) + g(v_1 + v_2)(v_1 + v_2 - \bar{v}_2).$$

The difference can be written as

$$(\bar{v}_1 - \bar{v}_2)(g(v_1) - g(\bar{v}_1)) + \bar{v}_2(g(\bar{v}_2) - g(\bar{v}_1)).$$

The first term is positive and the second is negative. When $\bar{v}_2$ goes to zero, the second term vanishes, and hence the first term dominates. Similarly, when $v_1 + v_2$ is large enough, we must have v_1 large since $v_2 < \bar{v}_2$, and hence the first term grows without bound when $g(\cdot)$ is unbounded. It follows that the deviation is profitable under the conditions of the proposition. □

Proof of Proposition 9 Consider a candidate symmetric equilibrium (v, p) for attentionally undifferentiated competition. Due to competition, we must have $p = \phi(v)$. We solve for conditions under which firm 1 would not want to deviate from this candidate equilibrium. Take first a deviation to $v^1 > v$ and $p^1 > p$. For the consumer to be indifferent between the two products, we must have

$$g(v^1)v^1 - g(p^1)p^1 = g(v^1)v - g(p^1)p,$$

so that

$$g(v^1)(v^1 - v) - g(p^1)(p^1 - p) = 0.$$

Differentiating totally with respect to v^1 and evaluating at $v^1 = v, p^1 = p$ gives

$$\left.\frac{\mathrm{d}p^1}{\mathrm{d}v^1}\right|_{v^1=v} = \frac{g(v)}{g(p)}.$$

For (v, p) to be an equilibrium, it must be the case that by this local deviation, the firm cannot make a positive profit; that is, $\phi'(v) \geq g(v)/g(p)$.

A similar calculation for a local deviation to $v^1 < v, p^1 < p$ shows that we must also have $\phi'(v) \leq g(v)/g(p)$, so that equilibrium requires

$$\phi'(v) = \frac{g(v)}{g(\phi(v))}. \tag{3}$$

Now consider attentionally differentiated competition. Consider again a candidate equilibrium (v, p) with $p = \phi(v)$. If firm 1 deviates to $v^1 > v$, $p^1 > p$, then for the consumer to be indifferent between the two products we must have

$$g(v^1)v^1 - g(p^1)p^1 = g(v)v - g(p^1)p.$$

Unlike in the case of attentionally undifferentiated competition, here a deviation by firm 1 to increase value *decreases* rather than increases attention on the competitor's product. Rewriting the above expression gives

$$g(v^1)v^1 - g(p^1)(p^1 - p) = g(v)v.$$

Differentiating totally with respect to v^1 and evaluating at $v^1 = v$, $p^1 = p$ gives

$$\left.\frac{\mathrm{d}p^1}{\mathrm{d}v^1}\right|_{v^1=v} = \frac{g(v) + g'(v)v}{g(p)},$$

and the same expression holds for downward deviations. As a result, an equilibrium must satisfy

$$\phi'(v) = \frac{g(v) + g'(v)v}{g(\phi(v))}. \tag{4}$$

The statements of the proposition follow from comparing (3) and (4) with the analogous expression in the proof of Proposition 7. □

B Eliciting Model Ingredients from Behavior

Since our model has ingredients that are not based directly on observed behavior, in this section we outline an algorithm for eliciting these ingredients from choices. Once the ingredients are elicited, our model provides a prediction on behavior for any choice finite choice set. Note that from the perspective of falsifiability, this is logically equivalent to providing axiomatic foundations for the model.

As we have noted in Section 2, our model has no behavioral implications if there are at most two dimension-state pairs. For our algorithm, we assume that there are at least three consumption dimensions; minor modifications work also if there are at least three dimension-state pairs. We normalize $u_k(0) = 0$ for each k, $g(0) = 1$, and $u_1'(0) = 1$. The first two of these normalizations is without loss of generality. Holding the attention function $g(\cdot)$ fixed, the third normalization is not without loss of generality—but our procedure elicits the $g(\cdot)$ that is appropriate for this normalization.

The first step in our algorithm elicits the curvature of the utility function for each dimension k. Focusing only on dimensions 1 and k, consider choice sets of the form $\{(0, x + \delta(p)), (p, x)\}$ for any $x \in \mathbb{R}$ and $p > 0$. For any $p > 0$ we can find the $\delta(p)$ that makes the decisionmaker indifferent between the two options. Hence, we have

$$g(u_1(p) - u_1(0))(u_1(p) - u_1(0)) = g(u_k(x + \delta(p)) - u_k(x))(u_k(x + \delta(p)) - u_k(x)).$$

Dividing the above by p and letting $p \to 0$ gives

$$g(0)u_1'(0) = g(0)u_k'(x)\delta'(p).$$

Hence, this procedure elicits $u_k'(x)$.

The second step in our procedure elicits the attention weights $g(\cdot)$. Since we have now elicited the utility function, here we work directly with utilities. Looking only at dimensions 1, 2, and 3, consider choice sets of the form $\{(p, 0, x), (0, x + \delta(p), 0)\}$. Again, for each p we find the $\delta(p)$ that makes the decisionmaker indifferent between the two options. Hence

$$g(p)p = g(x + \delta(p))(x + \delta(p)) - g(x)x.$$

Dividing by p and letting $p \to 0$ gives

$$g(0) = G'(x)\delta'(p),$$

where $G(y) \equiv g(y)y$. Hence, we have elicited $G'(x)$, which allows us to fully identify $g(\cdot)$.

C Naive Dynamic Model

To be typed in.