FUTILE ATTEMPTS AT SELF-CONTROL

Paul Heidhues
University of Bonn and CEPR

Botond Kőszegi
University of California, Berkeley

Abstract
We investigate costly yet futile attempts at self-control when consumption of a harmful product has a binary breakdown/no-breakdown nature and individuals tend to underestimate their need for self-control. Considering time-inconsistent preferences as well as temptation disutility, we show that becoming more sophisticated can decrease welfare and investigate what kind of mistaken beliefs lead to low welfare. With time-inconsistent preferences, being close to perfectly understanding one's preferences but assigning zero probability to true preferences induces the worst outcome. (JEL: D03, D11, D91)

1. Introduction

People spend tremendous energy and resources on trying to improve their behavior. An industry lives off designing dangerous yet typically ineffective methods to lose weight. Addicts attempt to control their desire to consume through medicines or limiting access to harmful substances, only to go out of their way to overcome these barriers later. And in a well-known example, DellaVigna and Malmendier (2006) document that most exercise "enthusiasts" who buy an expensive gym membership hardly use the membership.

This paper investigates costly yet futile attempts at self-control when consumption of a harmful product has a binary breakdown/no-breakdown nature. Although it is difficult to conclusively rule out that the above futile attempts are due solely to uncertainty, we instead base our model on the assumption that individuals are imperfect at predicting their future preferences or behavior.[1] Generalizing existing formulations of partial naivete with respect to time inconsistency by O'Donoghue and Rabin (2001), Eliaz and Spiegler (2006), and Asheim (2007),

Acknowledgments: We thank Geir Asheim, Eddie Dekel, Erik Eyster, Sebastian Kranz, and Ran Spiegler for useful discussions. Heidhues gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through SFB/TR-15. Kőszegi gratefully acknowledges financial support from the National Science Foundation under Award 0648659.

E-mail address: Heidhues: heidhues@uni-bonn.de; Kőszegi: botond@econ.berkeley.edu

1. For evidence, see for instance Ausubel (1991), Read and van Leeuwen (1998), Shui and Ausubel (2004), DellaVigna and Malmendier (2006), Giordano et al. (2005), and Skiba and Tobacman (2007).

Journal of the European Economic Association April–May 2009 7(2–3):423–434

we investigate what kinds of mistaken beliefs tend to generate low welfare. If a person does not fully appreciate her future taste, she does not take sufficiently forceful steps to prevent herself from indulging in harmful consumption, making her attempt at self-control futile. Furthermore, because a better understanding of her future taste may lead her to make more aggressive, but still insufficient, attempts at self-control, an increase in such sophistication often decreases welfare. Being very close to perfectly understanding one's preferences but assigning zero probability to true preferences is the worst possible outcome.[2] Hence, educating consumers about the extent of their taste for immediate gratification can be welfare decreasing. Finally, after proposing an extension of Gul and Pesendorfer (2001) to allow for incorrect beliefs about future temptation disutility, we investigate how naivete in that model affects attempts at self-control.

2. Model

In this section, we introduce our model of consumer choice for products for which the consumer has self-control problems. There are two periods, $t = 0, 1$. Consumption can only occur in period 1 and the person's preferences regarding consumption in period 1 change between periods 0 and 1. To highlight this change, we call the person's period-0 incarnation "self 0," and her period-1 incarnation "self 1." Our formulation is motivated by consumers' attitude toward harmful products with hyperbolic discounting. The same formalism is also consistent with systematic mistakes that lead to overconsumption in period 1, as in Bernheim and Rangel (2004).

In period 1, the consumer chooses whether to consume the product: $c \in \{0, 1\}$. We normalize the price of the product to zero; its price can be included in its utility. Self 1's utility from consuming the product—gross of the self-control costs we introduce below—is $u(c, \beta)$, whereas self 0's gross utility function is $u(c, 1)$, where $0 \leq \beta \leq 1$. To capture the idea that selves 0 and 1 disagree about consumption—self 1 wants to consume whereas self 0 does not—we assume that $u(1, 1) < u(0, 1)$ and $u(1, \beta) > u(0, \beta)$. And to capture that β is a measure for their disagreement, we suppose that $u(1, \beta) - u(0, \beta)$ is decreasing in β.

A time-inconsistent taste for immediate gratification, as captured by Laibson's (1997) model of hyperbolic discounting, satisfies our assumptions if consumption in period 1 generates sufficiently large harm in period 2. Alternatively, one can think of the consumer as being in a Bernheim–Rangel (2004) "hot" state in period 1 and crave consumption to an extent she disapproves of ex ante.

2. Although this result does not necessarily carry over to models with continuous consumption choices, Heidhues and Kőszegi (2008) show that profit-maximizing firms may choose discontinuous pricing policies—and hence artificially induce a kind of discreteness in consumption—because that is the most effective way of taking advantage of consumers' partial naivete.

Consistent with these interpretations, we equate welfare with long-run, period-0 preferences and put zero weight on self 1's taste.[3]

The consumer has access to a commitment technology: She can impose an instantaneous cost $k \geq 0$ on herself from consuming in period 1. The price of imposing this commitment is $p(k)$, where $p(0) = 0$ and $p(k)$ is strictly increasing and strictly convex in k, and has to be paid in period 1. The self-imposed cost and its price are additively separable from the utility from consumption for both selves, and are evaluated identically by the two selves.

As a stylized example consider a gambler trying to prevent herself from gambling too much. She can put herself on the exclusion list for any number of casinos, and has to choose the range of her home within which she cannot enter a casino. In this scenario, k is the cost of traveling to the nearest non-excluded casino and $p(k)$ is the cost of putting oneself on the exclusion list of casinos within travel cost k.

There are two ranges of self-control costs that lead to somewhat different implications. Let $k^* = u(1, \beta) - u(0, \beta)$; k^* is the minimal cost that will induce self 1 not to consume. If $p(k^*) < u(0, 1) - u(1, 1)$, then self-control is available at a sufficiently low price to be valuable. In the opposite case, it is objectively undesirable. Because we find the former case economically more interesting, we focus on it in the current paper.

Our main purpose is to investigate behavior and welfare when agents are possibly overoptimistic about their future behavior. We must therefore make assumptions regarding what self 0 thinks about self 1's preferences. There are two extreme assumptions one can make in this regard. Under naivete, the decisionmaker is completely unaware that she will have different preferences in the future, and under sophistication, she perfectly predicts how her preferences will change. Much of the growing literature that has examined the effect of awareness of one's future tendency for immediate gratification focuses on these two polar cases. Yet the question of failed attempts at self-control calls for a model with *intermediate* levels of sophistication: a completely naive individual does not make an attempt at self-control, and (without uncertainty) a completely sophisticated individual does not break down. Hence, we assume self 0 has beliefs $F(\hat{\beta})$ distributed on $[0, 1]$ regarding her future β. For much of our analysis, we will assume that the support of F is contained in $[\beta, 1]$, so that the consumer is weakly optimistic as to how much future preferences coincide with the current one. In

3. For justifications of this assumption in the context of hyperbolic discounting, see for instance DellaVigna and Malmendier (2004), Gruber and Kőszegi (2004), and O'Donoghue and Rabin (2006), as well as Theorem 11 of Bernheim and Rangel (2008); in our model this also coincides with the welfare approach proposed by Asheim (2007). And in Bernheim and Rangel's (2004) model, consumption in the hot state might be motivated by a misprediction of the resulting pleasure, so that ex ante utility is relevant for welfare.

this case, we can define an increase in sophistication as a first-order stochastic downward shift in F.

To the best of our knowledge, two metrics for the degree of sophistication have been proposed, and our specification encompasses both as specific cases. O'Donoghue and Rabin (2001) suppose that self 0 believes with certainty that self 1's utility function will be $u(c, \hat{\beta}) - p(k) - ck$. The parameter $\hat{\beta}$ reflects self 0's beliefs about β, so that $\hat{\beta} = \beta$ corresponds to perfect sophistication, $\hat{\beta} = 1$ corresponds to complete naivete, and a decrease in $\hat{\beta}$ corresponds to an increase in sophistication.

Eliaz and Spiegler (2006) and Asheim (2007) approach the question by restricting period-0 beliefs to binary distributions that put positive weights on time consistency and on the truth. Slightly generalizing this, we consider binary beliefs that put positive weight on the two utility functions $u(c, \beta) - p(k) - ck$ and $u(c, \beta') - p(k) - ck$, where $\beta' > \beta$. In this framework, an increase in sophistication corresponds to an increase in the subjective probability of the correct utility function.

To determine behavior, we assume that self 0 chooses k to maximize period-0 utility given her beliefs about future behavior, assuming that if self 1 is indifferent between consuming and not consuming, she does not consume. This corresponds to a "perception-perfect strategy" as defined by O'Donoghue and Rabin (2001).

3. Analysis

We next analyze our model. To build intuition for our more general results and to compare them to those under previous theories, we first analyze outcomes under the O'Donoghue-Rabin and Eliaz-Spiegler parameterizations of naivete.

3.1. Degenerate Beliefs about Future Preferences

We begin by analyzing the specific case of a degenerate distribution F that puts unit mass on some $\hat{\beta} \geq \beta$, as in O'Donoghue and Rabin (2001). Because self 0 is certain she will have type $\hat{\beta}$, she believes she will refrain from consumption if

$$u(0, \hat{\beta}) - p(k) \geq u(1, \hat{\beta}) - p(k) - k.$$

Hence, self 0 believes it will take a self-control cost of

$$k = \max\{0, u(1, \hat{\beta}) - u(0, \hat{\beta})\}$$

to make sure that self 1 does not consume.

Because self-control is desirable for a fully sophisticated consumer, for any $\hat{\beta} \geq \beta$ self 0 believes it is desirable for her as the perceived costs are lower and

the benefits are the same. Hence, self 0 imposes the perceived necessary cost that achieves self-control. Unless $\hat{\beta} = \beta$, however, this will be insufficient to deter self 1 from consuming. Hence, perfect sophistication is necessary to reap the benefits of the self-control technology. Intuitively, the consumer may realize that she needs to take steps to achieve self-control, but utility maximization implies that she would like to do this at the lowest possible cost. Hence, she imposes the consumption penalty k that she thinks is just sufficiently large to ensure she does not consume. Thus, all unsophisticated consumers—no matter how close they are to sophistication—endogenously select an ineffective level of punishment.

Not only is full sophistication necessary to benefit from the self-control technology, among unsophisticated consumers welfare is *decreasing* in sophistication. Figure 1 illustrates welfare as a function of $\hat{\beta}$. For high $\hat{\beta}$, the consumer believes she will not consume anyhow, so she chooses $k = 0$, and her utility is constant in $\hat{\beta}$. For lower $\hat{\beta}$, the consumer realizes that she needs the self-control technology, but has false beliefs about how much she needs, and her utility is lower than in the former case. Moreover, because a more pessimistic consumer buys more—yet still insufficient amounts of—self-control, she decreases her welfare by imposing a larger futile cost on herself. This decrease in welfare is not due to less self-control, but due to failed attempts at self-control.

Once the consumer reaches full sophistication, her welfare jumps up because she now imposes a sufficiently large penalty for consumption so that she does not indulge. Figure 1 also illustrates what happens when $\hat{\beta} < \beta$, so that the consumer is overly pessimistic about her future self-control. So long as she is not extremely pessimistic, she buys sufficient self-control, but her welfare is lower than that of

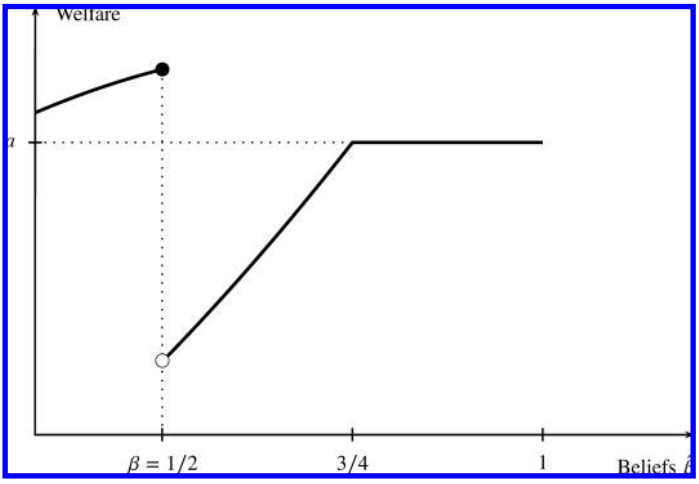

FIGURE 1. Welfare as a function of degenerate beliefs $\hat{\beta}$.

a sophisticated consumer because she pays too high a price. If she is extremely pessimistic, she may decide that self-control is too expensive and not buy any.

Although our model is very simplistic, the above basic intuition fits many examples of futile attempts at self-control. In some situations, people try to make it more difficult to overconsume by withdrawing less cash, buying less unhealthy food for home, avoiding parties, and imposing other restrictions on their own choice set. But if a person chooses a suboptimal commitment level, she may end up both overconsuming and paying the cost of breaking the commitment. For instance, if she buys too small a package of potato chips for the soccer game, she might drive to the supermarket at half time, both gaining the weight and missing the beginning of the second half. Similarly, some drug users take pharmaceuticals or engage in "aversive therapy" to make it unpleasant to consume the drug in the future. But if these measures are insufficient, they both consume the drug and suffer the aversive costs of the treatment. The same logic holds for many internal commitment devices, such as internal rules of spending, no flirtation, and so on, or asking friends to stop overconsumption by acting as a "punisher."

The above results make use of the fact that the punishment k is a continuous variable while the consumption decision is discrete. If the set of possible commitment levels is discrete, then a *nearly* sophisticated consumer—unable to fine-tune k to the level she believes is just sufficient—often chooses a sufficiently stiff punishment to deter consumption. Furthermore, if the consumption decision is continuous, even a lower-than-optimal level of punishment can greatly reduce the person's time-inconsistency problem. Still, within the range in which she chooses an insufficient level of punishment, becoming more pessimistic can make her worse off.

3.2. Binary Beliefs about Future Preferences

We now consider situations where self 0's beliefs about her future preferences are binary, with weight q on the true conflict β and weight $1 - q$ on $\beta' > \beta$.

Because intermediate levels of k would constitute a waste of resources that do not change behavior in any state of the world, the consumer chooses one of three possible self-control costs: $k = 0$, $k = k^*$, and the level $k = k^*(\beta')$ that is just sufficient to prevent self 1 from consuming if she has short-term impatience β'. To simplify the exposition, suppose that $k^*(\beta') > 0$; the argument for $k^*(\beta') = 0$ is almost identical. Because the consumer believes she always needs self-control, $k = k^*$ dominates $k = 0$. Hence, the consumer is effectively choosing between $k = k^*$ and $k = k^*(\beta')$. Clearly, for $q = 0$ she chooses $k^*(\beta')$. As q increases, self 0 views it as less valuable to impose the intermediate level of self-control, as this becomes less likely to be sufficient and leads to a greater probability of paying an ineffective punishment. This means that there is some cutoff value $\bar{q}$

such that the consumer chooses $k = k^*(\beta')$ for $q < \bar{q}$, and $k = k^*$ for $q > \bar{q}$. An increase in sophistication therefore benefits the consumer.

While Eliaz and Spiegler (2006) have a more complicated setting where the self-control technology is supplied by a firm through endogenous contracts, this result is analogous to their insights that (i) an increase in sophistication increases welfare; (ii) an increase in sophistication increases self-control; and (iii) sufficiently sophisticated consumers receive perfect self-control. As we show subsequently, points (i) and (iii) extend to more general changes in F that shift weight to the true preference parameter, but point (ii) does not.

3.3. Some General Results

This section identifies ways in which the above results generalize. We show a sense in which believing the exact truth is crucial for welfare: An increase in the weight on the true preference parameter β increases utility, but an increase in the weight near but above β often decreases utility, with a lot of weight in that region being close to welfare-minimizing.

To start, we consider situations in which self 0's belief about the preference parameter β is distributed with a density $f(\hat{\beta})$. We establish that if the density is high enough, the consumer imposes a penalty she thinks is just sufficient to always deter her from consuming in period 1. Suppose this is not the case. Then, we can define a cutoff β type $b(k)$ such that if the consumer has this type in period 1, she is indifferent between consuming and not consuming: $u(1, b(k)) - u(0, b(k)) = k$. Hence, self 0's problem is

$$\max_k F(b(k))[u(1, 1) - k] + (1 - F(b(k)))u(0, 1) - p(k).$$

The first-order condition for an interior k is

$$f(b(k))(-b'(k))[k + u(1, 1) - u(0, 1)] = p'(k). \tag{1}$$

If $f(\hat{\beta})$ is sufficiently large everywhere on the support of $\hat{\beta}$, equation (1) cannot be satisfied. Hence, we have the following.

PROPOSITION 1. *Let F be distributed on the support $[\underline{\beta}, \bar{\beta}]$ with a density $f(\hat{\beta})$. If $f(\hat{\beta})$ is sufficiently large, then the consumer imposes an amount of punishment k that she deems just sufficient to always deter her from consuming, $b(\underline{\beta})$. If $\beta < \underline{\beta}$, she nevertheless consumes in period 1. In this range, her welfare is increasing in $\underline{\beta}$.*

For densely distributed beliefs, the consumer does not choose an intermediate level of punishment: If choosing a k that prevents her from consuming in some

states of the world is worthwhile, choosing a slightly higher k that always does so is also worthwhile. She is therefore effectively choosing between $k = 0$ and the lowest k that always prevents her from consuming. Because self-control is valuable and she believes she needs a cost lower than k^* to ensure zero consumption, she chooses the latter option. As the consumer becomes more pessimistic, therefore, she will choose a higher punishment k, decreasing her welfare because this punishment is still insufficient to reign her in. This result generalizes our insights for degenerate beliefs in that an increase in sophistication will often lower welfare, and a consumer who has beliefs very close to sophisticated that do not include β on their support has close to the lowest possible welfare.

Once the support of the consumer's beliefs reaches the true conflict β ($\underline{\beta} = \beta$), she is sufficiently pessimistic to buy enough self-control, maximizing welfare. If the support shifts even lower, she chooses an overly costly punishment, decreasing welfare. Even so, in this range the commitment technology benefits her.

The conclusions that as long as $\underline{\beta} > \beta$ an increase in sophistication lowers welfare may not hold for more widely distributed beliefs. In fact, in that case there is a force for an increase in sophistication to decrease k and thereby increase welfare. Intuitively, for dispersed beliefs the consumer chooses an intermediate k to save on its cost. This means that she foresees a chance of consuming despite the punishment. As she becomes more pessimistic, she believes she will pay the futile cost more often. This leads her to decrease the punishment.

The second part of this section's main message is that if the consumer's beliefs shift to β from other regions of the distribution, her welfare must increase.

PROPOSITION 2. *Suppose F and G are distributed on $[\beta, 1]$ with $F(\beta) > G(\beta)$ and for any interval $I \subset (\beta, 1]$, $\Pr_F(I) \leq \Pr_G(I)$. Then, the consumer's welfare is at least as high if she believes F as if she believes G, and strictly higher if behavior is different. If period-0 behavior is different, the consumer chooses some $0 < k < k^*$ with G, and either a lower k or k^* with F.*

To see why the proposition is true, notice that for any $k < k' < k^*$, the perceived gain from increasing k to k' is greater if the consumer believes G than if she believes F. Because F puts a lower weight on any range of high $\hat{\beta}$'s, this increase in punishment prevents fewer types from consuming. In addition, the subjective probability of having to pay the punishment is greater under F than under G, providing a further reason not to choose a high punishment. As a consequence, a more pessimistic consumer either chooses a lower punishment, or chooses k^*.

Finally, to illustrate an effect missing from our propositions, we consider a "hybrid" case in which self 0 puts positive probability on β, and becomes more sophisticated by moving the rest of the distribution closer to β. As a simple

possibility, consider a binary distribution with weights q and $1-q$ on β and $\beta' > \beta$, respectively. Suppose q is sufficiently small so that the consumer chooses the punishment $k^*(\beta')$ just sufficient to prevent her from consuming with parameter β'. We analyze what happens as β' decreases toward β. Similar to the logic in the previous section on binary distributions, the consumer chooses either $k = k^*$ or $k = k^*(\beta')$. As β' shifts left, the cost premium for choosing the higher punishment decreases, and carries the increased benefit of avoiding a higher punishment if self-control fails. Hence, there is some threshold β' above which the consumer chooses $k^*(\beta')$, and below which she chooses k^*. Her welfare is decreasing in sophistication in the former range, and higher and constant in sophistication in the latter range.

4. Naivete in Temptation Disutility

Gul and Pesendorfer (2001) introduce a model in which a time-consistent person may prefer committing not to consume because she anticipates unpleasant temptation if the good is available. Here, we propose a slight modification of their framework that allows for individuals to underestimate future temptation, and consider the effects of naivete about temptation on self-control.

Suppose that similarly to our model above, in period 1 the consumer chooses whether to consume, and in period 0 she chooses a punishment $k \geq 0$ to impose on herself for consumption. In Gul and Pesendorfer's model, utility is defined over period-1 choice sets as

$$\max_{c \in \{0,1\}} u(c) - \left[\max_{c' \in \{0,1\}} v(c') - v(c) \right]. \tag{2}$$

The interpretation of these preferences is the following. The first term $u(c)$ is the person's "commitment utility" from c, the utility she would get from c if she could commit to it in period 0. The second term is the "temptation disutility" from choosing c rather than the most tempting option in the choice set, where temptation is captured by the function v. The model is predicated on the assumption that when committing to consumption ex ante, the decisionmaker does not experience temptation, but when confronted with an immediate choice, she does. We define u and v by

$$u(c) = 1 - c - ck - p(k); \quad v(c) = \alpha c(1-k),$$

where $\alpha > 1$ is a constant that captures the strength of temptation.

This specification shares some features of the time-inconsistent model of the previous section. If it was possible, the consumer would commit herself not to consume. But without commitment and $k = 0$, she consumes in period 1. Furthermore, it may be optimal to impose a consumption penalty on herself. To see

this, note that the utility from consuming is $-p(k) - k$, and the utility from not consuming is $1 - p(k) - \alpha(1 - k)$. Clearly, the optimal choice of k is zero if the decisionmaker consumes, and it is the $k^*(\alpha)$ defined by $p'(k^*(\alpha)) = \alpha$ if she consumes. If $1 - p(k^*(\alpha)) - \alpha(1 - k^*(\alpha)) > 0$, the consumer chooses a penalty equal to $k^*(\alpha)$ and does not consume in period 1. Like in the time-inconsistent model, this means that imposing a punishment is utility-decreasing unless it prevents consumption. But unlike in the time-inconsistent model, the optimal k is often larger than the minimum required to prevent consumption. Even if the consumer is already refraining from consumption, she may want to increase k because that mitigates the utility-decreasing temptation to consume. Analogously to the time-inconsistent case, we assume that the self-control technology is valuable: $1 - p(k^*(\alpha)) - \alpha(1 - k^*(\alpha)) > 0$.

We extend the model of Gul and Pesendorfer (2001) to allow for naive beliefs about v. Specifically, let $\hat{\alpha} \leq \alpha$ be self 0's belief about the strength of temptation disutility α.[4] Figure 2 shows how an increase in sophistication—an increase in $\hat{\alpha}$—impacts welfare in this model. For low levels of $\hat{\alpha}$ self 0 believes that she will refrain from consumption even for $k = 0$. Nevertheless, to decrease temptation she chooses a positive level of punishment. Because a higher $\hat{\alpha}$ means she is more concerned about temptation, she chooses a higher k. But because a relatively low

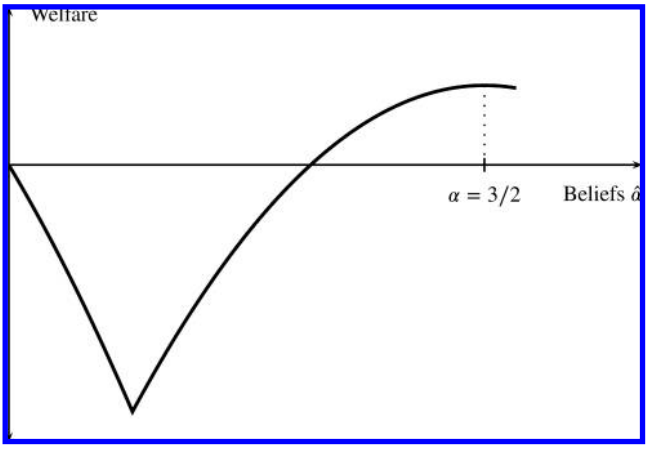

FIGURE 2. Welfare as a function of degenerate beliefs $\hat{\alpha}$.

4. Although Gul and Pesendorfer (2001) do not allow for naivete, doing so is not counter to the spirit of their revealed-preference framework. If a person's period-0 behavior satisfies their axioms, there are commitment utility and temptation disutility functions that represent this behavior. The inferred temptation disutility can be interpreted as the person's beliefs about her future temptation disutility. If the decisionmaker's period-1 behavior does not correspond to maximizing the sum of commitment and temptation utilities—for instance, because she behaves as if placing double the relative weight on temptation disutility—she can be interpreted as being naive.

punishment is insufficient to deter consumption, this decreases her utility. For some $\hat{\alpha} < \alpha$, $k^*(\hat{\alpha})$ is sufficient to prevent consumption in period 1. From this point on, an increase in k increases utility because it decreases the temptation to consume, and for high levels of $\hat{\alpha}$ welfare is higher than for $\hat{\alpha} = 0$.

These results replicate the main insight of the previous section: If an increase in sophistication does not lead a person to choose a sufficiently high punishment to deter consumption, the increase in sophistication is actually welfare decreasing. With temptation disutility, however, the consumer does not need to be fully sophisticated to buy sufficient self-control to benefit her, so that there is no discontinuity at full sophistication. Intuitively, with dynamic inconsistency the only purpose of the punishment is to deter consumption, so the consumer chooses the lowest level she thinks does so. With temptation disutility, punishment can increase utility even if it does not change behavior, so the consumer never chooses such knife-edge punishments. Thus, she benefits when being near sophisticated.[5]

Finally, consider binary beliefs that put probability q on the true strength of temptation disutility α and probability $1-q$ on some $\hat{\alpha} < \alpha$, where $k^*(\hat{\alpha})$ is insufficient to deter self 1 from consuming. As q increases from zero, the consumer recognizes that a low amount of punishment is less likely to deter consumption. She then either has to choose a discretely higher punishment to prevent consumption in both contingencies, or she has to accept indulging with probability q. For low levels of q, the higher costs of a discretely higher punishment are not worth it, so she chooses the latter possibility. In this range, an increase in k is a waste if she later breaks down, so an increase in q leads to a decrease in k and hence an increase in welfare. At a critical level of q, the consumer invests into a sufficiently high punishment to deter consumption, leading to an upward jump in welfare. If she attaches even higher probabilities to having high temptation disutility, she increases k to decrease temptation if it happens to be high and thereby further increase her welfare.

References

Asheim, Geir B. (2007). "Procrastination, Partial Naivete, and Behavioral Welfare Analysis." Working paper, University of Oslo.

Ausubel, Lawrence M. (1991). "The Failure of Competition in the Credit Card Market." *American Economic Review*, 81(1), 50–81.

5. The lack of discontinuity at sophistication in the temptation model is a manifestation of a more general difference between time-consistent and time-inconsistent models of a preference for commitment. With time consistency, changes in period-1 behavior cannot lead to jumps in welfare because self 1 changes her behavior when she is approximately indifferent. A fundamental property of time-inconsistent models, however, is exactly the existence of such discontinuities: Because selves 0 and 1 do not agree on what self 1 should do, as self 1 switches her action self 0's utility typically jumps.

Bernheim, B. Douglas, and Antonio Rangel (2004). "Addiction and Cue-Triggered Decision Processes." *American Economic Review*, 94(5), 1558–1590.

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

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

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

Eliaz, Kfir, and Ran Spiegler (2006). "Contracting with Diversely Naive Agents." *Review of Economic Studies*, 73(3), 689–714.

Giordano, Louis A., Warren K. Bickel, Eric A. Jacobs, George Loewenstein, Lisa Marsch, and Gary J. Badger (2005). "Altered States: Addicts Underestimate Future Drug Preferences." Working paper, University of Vermont, Duke University, Southern Illinois University, and Carnegie Mellon University.

Gruber, Jonathan, and Botond Kőszegi (2004). "A Theory of Government Regulation of Addictive Bads: Optimal Tax Levels and Tax Incidence for Cigarette Taxation." *Journal of Public Economics*, 88(9–10), 1959–1987.

Gul, Faruk, and Wolfgang Pesendorfer (2001). "Temptation and Self-Control." *Econometrica*, 69(6), 1403–1435.

Heidhues, Paul, and Botond Kőszegi (2008). "Exploiting Naivete about Self-Control in the Credit Market." Working paper, University of California, Berkeley.

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

O'Donoghue, Ted, and Matthew Rabin (2001). "Choice and Procrastination." *Quarterly Journal of Economics*, 116(1), 121–160.

O'Donoghue, Ted, and Matthew Rabin (2006). "Optimal Sin Taxes." *Journal of Public Economics*, 90(10–11), 1825–1849.

Read, Daniel, and Barbara van Leeuwen (1998). "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes*, 76(2), 189–205.

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

Skiba, Paige Marta, and Jeremy Tobacman (2007). "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default." Working paper, Vanderbilt University and the University of Pennsylvania.