UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

**WORKING PAPERS IN ECONOMICS**

**No 515**

# Naïve and Capricious: Stumbling into the Ring of Self-Control Conflict

**by**

**Kristian Ove R. Myrseth**

**Conny Wollbrant**

**August 2011**

**ISSN 1403-2473 (print)**
**ISSN 1403-2465 (online)**

---

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se    info@handels.gu.se

# Naïve and Capricious:

# Stumbling into the Ring of Self-Control Conflict*

*Kristian Ove R. Myrseth, ESMT European School of Management and Technology, Germany*[1]

*Conny Wollbrant, University of Gothenburg, Sweden*[2]

**Abstract**

We model self-control conflict as a stochastic struggle of an agent against a visceral influence, which impels the agent to act sub-optimally. The agent holds costly pre-commitment technology to avoid the conflict altogether and may decide whether to procure pre-commitment or to confront the visceral influence. We examine naïve expectations for the strength of the visceral influence; underestimating the visceral influence may lead the agent to exaggerate the expected utility of resisting temptation, and so mistakenly forego pre-commitment. Our analysis reveals conditions under which higher willpower—and lower visceral influence—reduces welfare. We further demonstrate that lowering risk aversion could reduce welfare. The aforementioned results call into question certain policy measures aimed at helping people improve their own behavior.

**JEL Classification:** D01, D03, D69, D90

**Keywords:** self-control, temptation, inter-temporal choice, pre-commitment

---

* Financial support from the *Jan Wallander and Tom Hedelius foundation* and The Swedish Research Council (*Vetenskapsrådet)* is gratefully acknowledged. We are grateful for helpful comment and suggestions on this project from Tore Ellingsen, Olof Johansson-Stenman, Peter Martinsson, Katarina Nordblom, Clara Inés Villegas Palacio, Robert Östling and participants at the 4th Nordic conference on Behavioral Economics and participants at the 2009 Meeting of the Society for Judgment and Decision Making.

[1] ESMT European School of Management and Technology, Schlossplatz 1, 10178 Berlin, Germany; Ph +49 (0) 30 212 31 1529; Fax +49 (0) 30 212 31 1281 ; E-mail myrseth@esmt.org.

[2] Department of Economics, School of Business, Economics and Law, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 26 15; Fax +46 31 786 10 43; E-mail conny.wollbrant@economics.gu.se

## 1. Introduction

Characteristic of the human condition is the struggle between reason and passion—the duel between better judgment and temptation. While some wield a will of iron, and so overcome temptation without much ado,[3] others do not even find it worth their while.[4] Most of us, however, are somewhere in between, trying to determine when, whether, and how to resist (for a review, see e.g. Fredericks et al., 2002). In this endeavor, we are often endowed with the aid of foresight and the technology of pre-commitment; we may anticipate that the cookie is approaching, and we may know how to avoid it—by taking the alternative route away from the bakery. However, foresight is limited (e.g., O'Donoghue and Rabin, 1999; Giordano et al., 2005), and pre-commitment technology is costly (e.g., Thaler and Shefrin, 1981; DellaVigna and Malmendier, 2006).

Attempting a modeling approach that is as simple as possible, yet true to the psychological spirit of the self-control problem, we consider an agent who possesses partial naïveté in her appraisal of visceral influence—or strength of temptation—and for whom effort to resist temptation is costly.[5] Specifically, we examine the decision to confront temptation versus procure pre-commitment. Our analysis reveals conditions under which higher willpower may yield welfare loss; the naïve individual, underestimating the strength of the impending temptation, may with higher willpower mistakenly choose to take on the temptation over procuring pre-commitment—where she would not with lower willpower. Following a similar logic, our analysis reveals conditions under which a lower visceral influence—or weaker

[3] "I am indeed a king, because I know how to rule myself." – Pietro Aretino

[4] "I can resist anything but temptation." – Oscar Wilde

[5] For a discussion on the role and properties of visceral influences, see Loewenstein (1996; 2000) and O'Donoghue and Loewenstein (2007).

temptation—yields welfare loss; the naïve individual may upon anticipating a lower visceral influence mistakenly choose to take on the temptation over procuring pre-commitment—where she would not had the visceral influence been stronger. Because we model self-control conflict as a stochastic process with a dichotomous outcome (win = resist, lose = succumb), the decision to forego pre-commitment may be seen as a choice of a gamble over a certain outcome (procuring pre-commitment). Hence, reducing risk aversion may lead to welfare loss; the partially naïve individual may mistakenly forego pre-commitment and take on the conflict –where she would not with greater risk aversion.

Of psychological interest, we find that optimal self-control effort follows an inverted U-shape with respect to the strength of temptation and that raising risk aversion reduces optimal effort to be spent in a struggle against temptation.

## 2. A Model of Self-Control Conflict

### *2.1. The Basic Model*

This section outlines the basic model of self-control conflict, first by considering the agent's decision problem, second by considering the agent's resulting maximization problem, and third by deriving her reaction function. Our model features a two-period decision problem, and the agent is assumed to hold the following properties:

ASSUMPTION 1: The agent is an expected utility maximizer, with a utility function given by

$$U(x,e) = u(x) - c(e),$$

$$u(x) = x^{\alpha}$$

$$c(e) = e / \omega$$

$$\alpha \leq 1,$$

$$\omega \geq 0,$$

$$e \geq 0,$$

where $x$ is the payoff; the restriction on $\alpha$ implies that the agent is risk neutral or risk averse; $\omega$ is a willpower parameter that augments the cost *c(e)* of self-control effort *e*, none of which can be negative.

We next consider the agent's choice alternatives. We assume two mutually exclusive decision outcomes, *g* and *a*, with *g* representing payoff from the goal—not consuming the cookie--and with *a* representing payoff from indulgence or temptation—consumption of the cookie. Moreover, we assume that *U(g, 0) > U(a, 0)*, such that the utility of achieving the goal *g,* in the absence of effort *e,* is greater than that from indulging in the tempting alternative *a*, also in the absence of effort *e*. This assumption defines the domain of the self-control problem, wherein the agent would prefer the goal payoff *g*, but—due to visceral influences of the tempting alternative—the cravings triggered by the smell and sight of the cookie—she would in the absence of a mitigating force—the effort to resist or pre-commitment technology— instead choose the tempting alternative, the payoff of which we denote *a*.

Figure 1>>>

The agent's decision problem (Figure 1) consists of two stages. At the pre-commitment stage (Stage 1), the agent anticipates the visceral influence of the temptation soon to come. That is, the agent is not yet starring down the temptation and so visceral influences are not yet interfering with choice. But the agent knows with certainty that the temptation will arise, and she is equipped at this stage to apply a pre-commitment technology (*PT)*[6] that ensures successful goal pursuit[7]. In other words, the cookie is approaching, but the agent knows an alternate route that will keep her a safe distance from the bakery. However, the pre-commitment technology is costly—the alternate route is more time-consuming—and we denote this cost $c \geq 0$. Hence, the payoff associated with applying pre-commitment is $g - c$.

Should the agent not apply pre-commitment (*NPT*), she proceeds to the conflict stage (Stage 2), where she may either attempt to exercise restraint (*R*) or indulge (*I*)[8]. Given that the visceral influence is positive, she will have to apply costly effort $e \geq 0$ to attempt restraint (*R*). This takes us to a definition of self-control conflict:

DEFINITION: *A self-control conflict occurs if and only if the following two conditions hold:*

(1) $$U(g,0) > U(a,0)$$

(2) $$v > 0$$

[6] Note that PT also denotes "personal trainers," who are paid handsomely to function as pre-commitment technologies. An acquaintance of one of the authors is known to book PT-sessions on Sunday mornings to prevent himself from going out drinking on Saturday nights.

[7] The first stage of our decision problem roughly corresponds to the "planner's" decision problem in Thaler and Shefrin's (1981) intrapersonal, two-agent "planner-doer" model.

[8] The second stage of our decision problem roughly corresponds to the "doer's" decision problem in Thaler and Shefrin's (1981); in their model, however, the doer always goes for the tempting alternative.

The first self-control conflict condition in (1) states that in the absence of visceral influences (where zero effort is required), the utility from *g*—associated with exercising restraint (*R*)— is larger than that from *a*—associated with indulging (*I*). The second self-control condition (2) states that the visceral influence must be positive. Conditions (1) and (2) yield in isolation conflicting prescriptions for behavior, and we apply the following assumption to resolve the conflict.

ASSUMPTION 2: *Given that (1) and (2) hold, the conflict between the agent, exerting effort* $e$ *and the visceral influence* $v$ *is resolved stochastically using the contest success function* $e/(e+v) = p_{win}$.[9]

The probability of successful restraint, denoted $p_{win}$, increases in the agent's own effort $e$ and falls in visceral influence $v$.

We solve the decision problem using backward induction, with the conflict stage as our point of departure. At the conflict stage, the agent first decides how much effort to exert, leading to an expected utility of conflict, which we denote $A$ . Second, at the pre-commitment stage, and knowing the value of $A$ at the conflict stage, the agent has to decide whether or not to apply pre-commitment. Hence, we start by solving (3) the maximization problem of the agent at the conflict stage, leading us to Proposition 1, below:

[9] For a discussion of this and the logit form of contest success functions, see Skaperdas (1996) and Hirshleifer (1989).

PROPOSITION 1: *Given risk neutrality ($\alpha = 1$), optimal self-control $e^*$ is given by the reaction function in (5)*

(3) $$\max_e \frac{e}{e+v} g + \left(1 - \frac{e}{e+v}\right) a - \frac{e}{\omega},$$

(4) $$s.t.\ e \geq 0,$$

(5) $$e^* = -v + \sqrt{\omega v (g-a)}.$$

All proofs are in Appendix A.

B. *The extended model: Underestimating visceral influence*

Knowing that an individual in a "cold" state (about to enter the bakery) tends to underestimate the influence of a "hot" state (standing inside the bakery),[10] we extend the basic model to account for naïve perception of the conflict stage from the vantage point of the pre-commitment stage. We denote by $s$, $0 \leq s \leq 1$, the agent's degree of sophistication in correctly anticipating the strength of the visceral influence experienced at the conflict stage. When the parameter of sophistication equals one, the agent has full appreciation of the strength of the visceral influence; when the parameter equals zero, the agent's naiveté is complete, and she expects no visceral influence at all. The extended decision problem, incorporating the sophistication parameter, is given in Figure 2.

Figure 2 >>>

[10] See e.g., Loewenstein (2000), O'Donoghue and Rabin (2001), Eliaz and Spiegler (2006) and Asheim (2008).

The maximization problem and reaction function for the extended decision problem are given below:

PROPOSITION 2: *Given* $s<1$, *naïve optimal self-control* $e^*_N$ *is given by the reaction function in (8)*

(6) $$\max_e \ \frac{e}{e+sv}g+\left(1-\frac{e}{e+sv}\right)a-\frac{e}{\omega},$$

(7) $$s.t.\ e\geq 0,$$

(8) $$e^*_N=-sv+\sqrt{\omega sv\left(g-a\right)}.$$

## 3. Analysis

We first explore in the basic model the association between key variables. We next investigate in the extended model welfare implications of naïve expectations of visceral influences and the implications of changes in willpower, visceral influence, risk preferences, and in costs of pre-commitment.

### *3.1. The Basic Model*

The reaction function obtained in (3) allows us to examine the agent's optimal amount of effort exerted in the face of temptation as a function of the strength—or visceral influence—of said temptation:

PROPOSITION 3: *Optimal self-control* $e^*$ *follows an inverted U with respect to visceral influence v.*

Figure 3 >>>

As the visceral influence increases (Figure 3), the optimal level of effort to be exerted in the self-control conflict increases to a maximum, after which it declines to zero. This result is of psychological significance. The psychological literature has often presumed that an individual who yields to temptation either has failed to perceive the self-control conflict in the first place (Myrseth & Fishbach, 2009) or flat-out failed at her best attempts at resisting (e.g., Baumeister et. al, 2002). Our result implies a third possibility—that the individual, out of her own best interest, may have not have exerted much effort against the temptation.

We next consider the association between optimal effort and other key variables, namely the utility from the goal, the utility from the tempting alternative (to be distinguished from the visceral influence of the tempting alternative, discussed above), and the level of willpower:

COROLLARY 1: *Effort rises in the utility from the goal, falls in the utility from the tempting alternative, rises in willpower, and falls in risk aversion*

Naturally, a higher degree of effort implies a greater chance of successful restraint:

COROLLARY 2: *Increases in the payoff from the goal and increases in willpower raise the probability of success, while increases in the utilities from the tempting alternative and the visceral influence reduce it; choice probabilities are monotonic in payoffs.*

Having analyzed the association between the variables conditional on the agent reaching the second stage of the decision problem, we now define condition under which the agent will choose to employ pre-commitment technology—and hence avoid the second stage altogether:

PRE-COMMITMENT CONDITION 1: *When the following condition holds, the agent will prefer to apply pre-commitment technology (PT) to facing the temptation*

$$u(g-c) \geq A, \tag{9}$$

where $A$ , the expected utility from conflict, is attained by maximizing with respect to effort $e$, i.e., using (5) in (3). That is, the agent prefers pre-commitment when the utility of applying pre-commitment technology exceeds or equals the expected utility from conflict.

### *3.2. The extended model: Underestimating visceral influence*

We next consider how the naïve agent at the pre-commitment stage treats the decision problem:

PRE-COMMITMENT CONDITION 2: *The naïve agent will prefer to forego available pre-commitment to applying pre-commitment if the following condition holds*

$$N \geq u(g-c), \tag{10}$$

where $N$ , the naive expected utility of conflict, is attained from naïve maximization with respect to effort, i.e., using (8) in (6). The distinction between naïve and "sophisticated" maximization is important because the naïve expected utility from conflict (*N*) is weakly greater than the actual expected utility from conflict (*A*):

LEMMA 1: *Given* $s < 1$ *, the naive expected utility from conflict (N) always is greater than the actual expected utility from conflict (A)*

In turn, naïve estimates of the visceral influence will lead the agent to suboptimal decisions under the following conditions:

WELFARE LOSS CONDITION: *The naïve agent makes suboptimal decisions whenever both pre-commitment conditions (9 and 10) hold (with strict inequality for one)*

Pre-commitment conditions 1 and 2 define the range for which the naïve agent makes suboptimal decisions and consequently suffers welfare loss. We rely on these conditions throughout our subsequent analysis.

Moreover, for $s < 1$ , there exist values of willpower $\omega$ , costs of pre-commitment *c*, values of risk aversion $\alpha$ , and of visceral influence $v$ , such that the agent makes suboptimal decisions—as summarized below:

LEMMA 2 (WILLPOWER): *Given* $s < 1$, *there exist values of willpower* $\omega$ *such that (9) and (10) hold, so that the agent makes suboptimal decisions and suffers a welfare loss (Figure 4).*

(11) $$sv\Lambda \le \omega \le v\Lambda ,$$

where $\Lambda = \left[ c(g-a) + 2a\sqrt{c(g-a)} + 2a^2 \right] / c^2(g-a)$.

Figure 4 >>>

LEMMA 3 (PRE-COMMITMENT COST): *Given* $s < 1$ *, there exist values of cost of pre-commitment c such that (9) and (10) hold, so that the agent makes suboptimal decisions and suffers a welfare loss (Figure 5).*

(12) $$g - A \geq c \geq g - N$$

Figure 5 >>>

LEMMA 4 (RISK AVERSION): *Given* $s < 1$ *, there exist values of risk aversion* $\alpha$ *such that (9) and (10) hold, so that the agent makes suboptimal decisions and suffers a welfare loss*

Figure 6>>>.[11]

[11] This condition is given by the solution to a cubic function, which is analytically cumbersome. Hence we provide numerical examples to prove that there exist values of risk aversion such that the naïve agent makes suboptimal decisions.

EXAMPLE 3: *Assume the following values for the model g=4, a=2, v=.2, ω=.3,s=.1 and c=.5. Then the agent makes suboptimal decisions and suffers a welfare loss when the risk aversion parameter is approximately between .97 and .73.*

LEMMA 5 (VISCERAL INFLUENCE): *Given* $s<1$ *, there exist values of visceral influence v such that (9) and (10) hold, so that the agent makes suboptimal decisions and suffers a welfare loss*

Figure 7>>>[12]

EXAMPLE 4: *Assume the following values for the model g=20, a=5, α=1, ω=1,s=.3 and c=10. Then the agent makes suboptimal decisions and suffers a welfare loss when the visceral influence is approximately between 2.5 and 7.5.*

Having established that naïve agents may make suboptimal decisions, we turn our attention to the relationship between the parameters of the model and the agent's level of welfare, summarized below:

PROPOSITION 4*: When the agent underestimates visceral influences v, increases in (i) willpower* $\omega$ *and (ii) cost of pre-commitment* $c$ *, and reductions in (iii) risk aversion α and (iv) visceral influence v, can all lead to suboptimal decisions and welfare loss .*

[12] This condition is given by the solution to a cubic function, which is analytically cumbersome. Hence we provide numerical examples to prove that there exist values of visceral influence such that the naïve agent makes suboptimal decisions.

A particularly noteworthy result is that (i) raising an agent's willpower may lead the agent to make suboptimal decisions and hence suffer welfare loss. That is, higher willpower for the naïve agent may not necessarily be a good thing. The intuition is straightforward (see Figure 4). The naïve agent, who at the pre-commitment stage underestimates the visceral influence, may still perceive that the expected value of procuring pre-commitment exceeds that of taking on conflict. However, higher willpower may alter this calculus, such that the perceived expected value of conflict exceeds that of pre-commitment—but such that the true expected value of conflict does not.

More in line with intuition is the result that (ii) higher cost of pre-commitment leads to welfare loss (see Figure 5). This result is obtained because the naïve agent, who initially would correctly procure pre-commitment, upon learning of a higher cost of pre-commitment might perceive the expected utility of taking on conflict as higher than that of procuring pre-commitment—even when the true expected utility of conflict remains lower than that of procuring pre-commitment. Hence, as a result of higher cost of pre-commitment, the naïf might choose to take on conflict even when she should not.

Similarly, (iii) upon experiencing reductions in risk aversion, the naïve agent, who initially would correctly procure pre-commitment, might perceive the expected utility of taking on conflict as higher than that of procuring pre-commitment—even when the true expected utility of conflict remains lower than that of pre-commitment (see Figure 6). Hence, as a result of lower risk aversion, the naïf might choose to take on conflict even when she should not.

Possibly least intuitive of all is our result (iv) that reducing the visceral influence might lead to welfare loss (see Figure 7). Upon observing that the visceral influence has dropped, the naïve agent, who initially would correctly procure pre-commitment, might perceive the expected utility of taking on conflict as greater than that of procuring pre-commitment—even when the true expected utility of conflict remains lower than that of pre-commitment. Hence, as a result of a lower visceral influence, the naïf might choose to take on conflict even when she should not. In other words, lowering visceral influence raises both the perceived and the true expected utility of conflict, the former of which always is greater than the latter; when the expected utility of pre-commitment is higher than both, there is a possibility that the reduction in visceral influence lifts the perceived expected utility of conflict above that of pre-commitment—all while the true expected utility of conflict remains below that of pre-commitment.

## 4. Related literature

### *4.1. Economics*

Although our model consists of a single decision maker, it may be derived as a reduced form of the dual-self model by Fudenberg and Levine (2006). They present a generalized model of that by Thaler and Shefrin (1981). The models by both Fudenberg and Levine (2006) and Thaler and Shefrin (1981) assume a "long-run self" ("planner") , who gains utility from the discounted sum of the utility of all "short-run selves" ("doers") . The agent in our model could be seen as containing the operating procedures of the planner, and the visceral influence as representing a function of the doer's opportunity cost of restraint.

An important similarity between the Fudenberg and Levine (2006) model and ours is their specification of self-control cost as "opportunity-based:" in a given situation, raising the opportunity cost of restraint for the short-run self, raises self-control cost. While our cost of self-control does not directly depend on preferences by a short-run self, a higher visceral influence raises the cost of self-control (in the form of effort to resist). Because the magnitude of visceral influences can be thought of as measuring the difference between the short-run utility from restraint and the short-run utility from indulgence, visceral influences can be interpreted as measuring the opportunity cost of the short-run self, as in the model of Fudenberg and Levine (2006). Furthermore, the preferred specification of self-control costs of Fudenberg and Levine (2006) is convex, to account for cognitive load. In our model, costs are linear, but—due to the non-linear nature of the contest-success function in the maximization problem—the benefit function is concave. Our concave benefit function and linear cost function, taken together, is technically equivalent to the linear benefit function and convex cost of Fudenberg and Levine (2006).

Alternatively, taking the Gul and Pesendorfer (2001, 2004) axiomatic model as a starting point, one could also derive a stochastic model similar to ours. The authors consider a single player who has preferences for choice sets that limit the available alternatives. That is, the agent may prefer a subset of alternatives to the set itself. These different choice sets are motivated by the presence of visceral influences, such as hunger. Under various axioms of choice over menus of lotteries, including the "set betweenness axiom," they show that the decision process can be represented by a utility function with a cost of self-control, or a disutility from the presence of a tempting alternative. The result is a set of preferences that explains behavior, in the face of temptation, as the outcome of a rational decision process; the agent benefits from pre-commitment technologies to reduce future temptations. Bénabou and Pycia (2002) show that Gul and Pesendorfer's (2001) representation result can be re-expressed as a costly stochastic intra-personal conflict between a planner and doer, as in Thaler and Shefrin (1981); the probabilities that determine successful restraint are in turn determined by the relative strengths of the (long-run) utility from restraint and the (short-run) utility from indulgence. Though our specification is slightly different, the visceral influence that we employ can be thought of as resulting from temptation disutility, and our probabilities determining successful restraint could also be derived in the manner that Bénabou and Pycia (2002) demonstrate.

A prominent alternative to the approaches by Fudenberg and Levine (2006) and Gul and Pesendorfer (2001;2004) is the family of hyperbolic discounting models. These account for the apparent preference reversals that have been documented since Strotz (1956); individuals may prefer $50 today to $100 in a year, while simultaneously preferring $100 in six years to $50 in five—even though these decisions are normatively identical. [13] Laibson (1997) studied the class of discount functions that lead to such "time-inconsistent preferences," in particular the ($\beta$, $\delta$) functional form of hyperbolic discounting. This discounting function has the property of being steeper than the conventional exponential discount function—the standard in neoclassical economics. Our model also allows for the preference reversals observed under hyperbolic discounting; the visceral influence is stronger for decisions closer in time (e.g., see Loewenstein, 1996), and temporal proximity of a given consumption opportunity is therefore associated with more indulgence (and less restraint).

Reminiscent of the hyperbolic discounting approach, Schelling (1978, 1984) discusses the dichotomy between what one wants ex-ante and what one wants ex-post. In particular, Schelling discusses the different strategies one might undertake to "game" oneself ex-ante. For example, a woman about to give birth might request that anesthesia be made unavailable during delivery if she knows that she will use it. Our approach does not speak to the explicit nature of these strategies, but it allows for them, and it refers to them collectively as pre-commitment technology.

[13] See also Ainslie (1975) and Fredericks et al., (2002)

As originally conceived, the aforementioned models all rely on sophisticated agents. Our emphasis, however, is on naïve agents, who underestimate visceral influences. Heidhues and Kőszegi (2010) describe two approaches to modeling naïve expectations of successful self-control (or “time-consistent”) behavior. The first is (i) to assume that the individual holds two utility functions, one time-consistent and one time-inconsistent; each individual holds a belief that she with a certain probability will behave according to one or the other utility functions. When the subjective probability measure puts more weight on the time-inconsistent utility function, the agent is said to be more sophisticated (see e.g., Eliaz and Spiegler, 2006; Asheim, 2008). The second is (ii) to assume that the planner has a belief about her β-parameter and that the belief takes values between the true β-parameter and 1 (see O’Donoghue and Rabin, 2001). When the belief is equal to her true β-parameter, the agent is fully sophisticated, and when the belief is equal to 1, she is fully naïve; an increase in the belief parameter corresponds to a decrease in sophistication. Our own analysis is similar to the latter approach; we use a belief parameter $s$, which takes values between 0 and 1, such that the anticipated value of the visceral influence $v$ becomes $sv$. The agent is fully sophisticated when $s = 1$ and fully naïve when $s = 0$; an increase in $s$ corresponds to an increase in sophistication.

Notably, the main result of Hedihues and Kőszegi (2010) is not preserved within our framework. Heidhues and Kőszegi (2010) demonstrate that higher degrees of sophistication can lead to welfare loss when self-control behavior has the dichotomous nature of utter failure or complete success. As sophistication of the agent increases, she invests more resources in order to improve her behavior. Because she is not completely sophisticated, however, she fails to invest just enough to secure optimal behavior. Consequently, when the temptation presents itself, the agent may indulge and incur more losses than she would have done, had she been more naïve and invested less. Our framework does not preserve this result because self-control behavior at the conflict stage in our model is probabilistic. Raising naïveté about the visceral influence increases the naïve expected utility of conflict (*N*) above that of the true expected utility of conflict (*A*). Raising the parameter *s* (measuring sophistication) reduces the difference between these two values and hence contracts the space for suboptimal behavior.

### *4.2. Psychology*

Our model captures the psychological distinction between "system 1" and "system 2" cognition—between "cold," rational, and volitional cognition, on one side, and "hot," emotional, automatic cognition, on the other (see e.g., Sloman, 1996; Stanovich and West, 2000; Kahneman, 2003). The agent in our model can be said to represent system 2 cognition, and the visceral influence—the lure of the temptation that acts upon the agent—can be said to represent system 1. The self-control conflict thus becomes the effort from system 2 to override system 1, as similarly suggested by Metcalfe and Mischel (1999) and Hofmann et al. (2009). A notable property of system 2 is that it relies on limited cognitive resources. That is, it gets tired. Accordingly, some psychologists have employed a muscle metaphor to understand the self-control problem; the individual holds limited resources to draw on in the struggle against temptation, and having used some to resist one temptation, there will be fewer available for the next (Baumeister et al., 1998; Muraven and Baumeister, 2000; Baumeister, 2002).[14] Our model, utilizing a concave benefits function with a linear cost function (which is technically the same as having a linear benefit function with a convex cost function), is isomorphic to a model that treats the agent's resources as limited.

[14] However, with time, the resources get replenished—as a tired muscle regains strength.

Critical for our results is the partial naïveté of the agent in appraising anticipated visceral influences. This particular form of naïveté—underestimating future visceral influences—is well-documented in the psychological literature and referred to as the empathy gap (see e.g., Loewenstein, 2000). Individuals in a "cold" state—when not subject to a visceral influence—fail to appreciate the force that the visceral influence will exert upon them in a "hot" state—when subject to it (Loewenstein et al., 1997; VanBoven et al., 2000; VanBoven and Loewenstein, 2003). For example, prior to passing the bakery, the dieter likely will underestimate the attractive forces as she walks by, triggered by the scent and sight of fresh pastries.

The psychological literature also offers examples of self-control strategies that may not be conscious and effortful (see e.g., Fujita, 2011), and our model could easily be augmented to account for such mechanisms. For example, Fishbach et al. (2010) argue that individuals facing self-control conflict employ automatic, effortless strategies to reduce the value of temptation and to increase the value of the goal (see also Fishbach and Shah, 2006). To accommodate such effects, we could add parameters to the payoff from the goal $g$ and to that from the tempting alternative $a$. Our analysis would otherwise remain the same.

## 5. Conclusion

We have proposed a model that explores the self-control conflict—how an agent determines whether or not to procure pre-commitment technology and, if not, how much effort the agent invests in the struggle against temptation. In so doing, we have attempted to incorporate psychological principles—that the expenditure of willpower is costly and that individuals tend to be naïve about the strength of anticipated visceral influence that is associated with temptation. Our model yields a number of insights, the perhaps most noteworthy of which are that the naïve agent may not benefit from (i) having greater willpower or (ii) facing a weaker visceral influence (or strength in temptation). The intuition for both results is that the naïve agent, who possesses higher willpower, or who faces a weaker visceral influence, may mistakenly confront temptation—rather than procure pre-commitment.

Our results call into question certain policy measures designed to help individuals improve their behavior. It is often thought that reducing the intensity of tempting stimuli—be they fatty foods, alcohol, cigarettes, or drugs—will help individuals restrain consumption. For example, the idea behind the nicotine patch, and nicotine replacement therapy more generally, is to reduce the cravings associated with withdrawal. [15] We show that this indeed may not be the case. In fact, reducing the cravings to smoke might well raise consumption and reduce welfare.[16] Moreover, policy makers should be aware that this measure might raise consumption for some individuals even if it were to reduce aggregate consumption. That is to say—consumers and policy makers should be aware that a potential side-effect for every consumer is the risk that the measure reduces their welfare. The same is true for measures designed to improve willpower.[17]

In a world of naïve consumers who possess pre-commitment technology, attempting to manipulate the visceral influence of temptation, or the self-control of an individual, raises

[15] The nicotine patch releases nicotine through the skin and into the blood stream; the patch is supposed to reduce cravings for cigarettes. Other forms of nicotine replacement therapy include the nicotine gum and the nicotine inhaler.

[16] The effectiveness of over-the-counter (OTC) nicotine replacement therapy (NRT) is still debated; Hughes et al. (2011) review non-randomized tests of the effectiveness of OTC NRT; while some of the studies present evidence for the effectiveness of OTC NRT, "the most rigorous studies do not find greater quitting among users."

[17] The Mighty Oak Inner Strength Program To Stop Smoking Marijuana (2011) tries to help individuals help themselves in building willpower to overcome Marijuana addiction—for the price of $40. http://www.marijuana-addiction-help.net/?hop=gladngreen (Available August 29, 2011).

distributional concerns, as these measures may raise the welfare of some while reducing that of others. To avoid distributional questions altogether, our analysis suggests that policy should focus on increasing perceived benefits of goal pursuit (e.g., staying healthy); increasing perceived costs associated with indulging in temptation (e.g., smoking cigarettes)—and, perhaps most importantly, reducing the individual's naïveté. In so doing, policy would not risk inadvertently pushing naïve consumers into the ring of self-control conflict—consumers whom with better foresight would have gladly paid to avoid the fight.

## References


Ainslie, George (1975). "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin,* 82(4), 463-496.

Asheim, Geir B. (2008). "Procrastion, Partial Naïveté and Behavioral Welfare Analysis." Memorandum 02/2007, Department of Economics, University of Oslo, Revises August 2008 (http://folk.uio.no/gasheim/procra02.pdf)

Baumeister, Roy F. (2002). "Yielding to Temptation: Self-Control Failure, Impulsive Purchasing, and Consumer Behavior." *Journal of Consumer Research,* 28(4), 670-676.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice (1998). "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology,* 74(5), 1252 – 1265.

Baumeister, Roy F., Todd F. Heatherton, and Dianne M. Tice (1994). "Losing Control: How and Why People Fail at Self-Regulation." San Diego, CA: Academic Press.

Bénabou, Roland, and Marek Pycia (2002). "Dynamic Inconsistency and Self-Control: A Planner–Doer Interpretation," *Economic Letters,* 77(3), 419–424.

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 Naïve Agents." *Review of Economic Studies*, 73(3), 689-714.

Elster, Jon (1977). "Ulysses and the Sirens: A Theory of Imperfect Rationality." *Social Science Information XVI*, 16(5), 469-526.

Elster, Jon (1998). "Emotions and Economic Theory." *Journal of Economic Literature,* 36(March), 47-74.

Fishbach, Ayelet, and James Y. Shah (2006). "Self-Control in Action: Implicit Dispositions Toward Goals and Away from Temptations." *Journal of Personality and Social Psychology,* 90(5), 820-832.

Fredericks, Shane, George Loewenstein, and Ted O'Donoghue (2002). "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature,* 40(2), 351-401(51).

Fudenberg, Drew, and David K. Levine (2006). "A Dual Self Model of Impulse Control." *American Economic Review*, 96(5), 1449–1476.

Fujita, Kentaro (2011). "On Conceptualizing Self-Control as More than Effortful Inhibition of Impulses." *Personality and Social Psychology Review,* XX(X), 1-15.

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.

Gul, Faruk, and Wolfgang Pesendorfer (2001). "Temptation and Self-Control," *Econometrica,* 69(6), 1403-1436.

Gul, Faruk, and Wolfgang Pesendorfer (2004). "Self-Control and the Theory of Consumption." *Econometrica*, 72(1), 119-158.

Heidhues, Paul, and Botond Kőszegi (2009). "Futile Attempts at Self-Control." *Journal of the European Economic Association*, 7(2-3), 423-434.

Hirshleifer, Jack (1989). "Conflict and rent-seeking success functions: Ratio vs. difference models of relative success." *Public Choice,* 63(2), 101-112.

Hofmann, Wilhelm, Malte Friese and Fritz Strack (2009). "Impulse and Self-Control from a Dual-Systems Perspective." *Perspectives on Psychological Science,* 4(2), 162-176.

Hughes, John R., Erica N. Peters and Shelly Naud (2011). "Effectiveness of Over-the-Counter Nicotine Replacement Therapy: A qualitative Review of Nonrandomized Trials." *Nicotine and Tobacco Research*, 13(7), 512-522.

Kahneman, Daniel. (2003). "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist*, 58(9)**,** 697–720.

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

Loewenstein, George. (2000). "Preferences, Behavior and Welfare: Emotions in Economic Theory and Economic Behavior." *The American Economic Review,* 90(2), 426-432.

Loewenstein, George. (1996). "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes,* 65(3), 272–292.

Loewenstein, George, Daniel Nagin, and Raymond Paternoster (1997). "The Effect of Sexual Arousal on Predictions of Sexual Forcefulness." *Journal of Research in Crime and Delinquency,* 34(4), 443-473.

Metcalfe, Janet, and Walter Mischel (1999)."A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower." *Psychological Review,* 106(1), 3–19.

Muraven, Mark and Roy F. Baumeister (2000). “Self-Regulation and Depletion of Limited Resources: Does Self-control Resemble a Muscle?” *Psychological Bulletin,* 126(2), 257-259.

Myrseth, Kristian Ove R., and Ayelet Fishbach (2009). “Self-Control: A Function of Knowing When and How to Exercise Restraint.” *Current Directions in Psychological Science,*18(4), 247-252.

O’Donoghue, Ted, and Matthew Rabin (1999). “Doing it Now or Later?” *American Economic Review,* 89(1), 103-124.

O’Donoghue, Ted, and Matthew Rabin (2001). “Choice and Procrastination.” *Quarterly Journal of Economics,* 116(1), 121-160.

Schelling, Thomas C. (1984). “Self-Command in Practice, in Policy, and in a Theory of Choice,” *American Economic Review,* 74(2), 1-11.

Skaperdas, Stergios. (1996). “Contest Success Functions.” *Economic Theory,* 7(2), 283-290.

Sloman, S.A. (1996). “The Empirical Case for Two Systems of Reasoning.” *Psychological Bulletin,* 119(1), 3-22.

Stanovich, Keith E., and, Richard F. West (2000). “Individual Differences in Reasoning: Implications for the Rationality Debate.” *Behavioral and Brain Sciences*, 23, 645-665.

Strotz, Robert H. (1955-56). “Myopia and Inconsistency in Dynamic Utility Maximization,” *Review of Economic Studies,* 23(3), 165-180.

Thaler, Richard H. and Hersch M. Shefrin (1981). "An Economic Theory of Self-Control." *Journal of Political Economy,* 89(2), 392-406.

Trope, Yaacov, and Ayelet Fishbach (2000). "Counteractive Self-Control in Overcoming Temptation." *Journal of Personality and Social Psychology,* 79(4), 493-506.

Van Boven, Leaf, David Dunning, and George Loewenstein (2000). "Egocentric Empathy Gaps Between Owners and Buyers: Misperceptions of the Endowment Effect." *Journal of Personality and Social Psychology,* 79(1), 66-76.

Van Boven, Leaf, and George Loewenstein (2003). "Social Projection of Transient Visceral Feelings." *Personality and Social Psychology Bulletin,* 29(9), 1159–1168.

APPENDIX A: PROOFS

PROOF PROPOSITION 1

The Lagrangean for this problem becomes:

$$\Gamma = \frac{e}{e+v}\left(g-\frac{e}{\omega}\right)+\frac{v}{e+v}\left(a-\frac{e}{\omega}\right)-\lambda_1\left(-e\right) \quad .$$

First order conditions:

$$\frac{\partial\Gamma}{\partial e} = \left(\frac{e}{e+v}\right)\left(-\frac{1}{\omega}\right)+\left(g-\frac{e}{\omega}\right)\left(\frac{v}{\left(e+v\right)^2}\right)+ \left(\frac{v}{e+v}\right)\left(-\frac{1}{\omega}\right)+\left(a-\frac{e}{\omega}\right)\left(\frac{-v}{\left(e+v\right)^2}\right)-\lambda_1 = 0 \quad ,$$

$$\frac{\partial\Gamma}{\partial\lambda_1} = -e \le 0;\ \lambda_1 \ge 0;\ \lambda_1\left(-e\right)=0 \quad .$$

The interior solution is derived by solving the first order conditions for $e$ yielding the reaction function $e^*$ below:

$$e^* = -v+\sqrt{\omega v\left(g-a\right)}$$

PROOF PROPOSITION 2

Assume $s \leq 1$, pre-multiplying *v*. Then, the Lagrangean for this problem becomes:

$$\Gamma = \frac{e}{e+sv}\left(g-\frac{e}{\omega}\right)+\frac{v}{e+sv}\left(a-\frac{e}{\omega}\right)-\lambda_1(-e) \qquad .$$

First order conditions:

$$\frac{\partial\Gamma}{\partial e}=\left(\frac{e}{e+sv}\right)\left(-\frac{1}{\omega}\right)+\left(g-\frac{e}{\omega}\right)\left(\frac{sv}{(e+sv)^2}\right)+ \left(\frac{sv}{e+sv}\right)\left(-\frac{1}{\omega}\right)+\left(a-\frac{e}{\omega}\right)\left(\frac{-sv}{(e+sv)^2}\right)-\lambda_1=0 \quad ,$$

$$\frac{\partial\Gamma}{\partial\lambda_1}=-e\leq 0;\ \lambda_1\geq 0;\ \lambda_1(-e)=0$$

The interior solution is derived by solving the first order conditions for $e$ , yielding the reaction function $e^*_N$ :

$$e^*_N=-sv+\sqrt{\omega sv(g-a)}$$

PROOF PROPOSITION 3

Note that optimal effort $e^{*}=-v+\sqrt{\omega v\left(g-a\right)}$ is concave in *v*. Further, recall the definition of a concave function: A function $f:\mathbb{R}^{n}\supseteq \mathrm{X}\rightarrow \mathbb{R}$, where $\mathrm{X}$ is a convex set, is concave if, given any two points $x^{'}$ and $x^{"}$ in $\mathrm{X}$, we have

$$(1-\lambda)\,f\left(x^{'}\right)+\lambda f\left(x^{"}\right)\leq f\left[(1-\lambda)x^{'}+\lambda x^{"}\right]\equiv f\left(x^{\lambda}\right)\forall\lambda\in\left(0,1\right).$$

Now, let $v^{'}$ and $v^{"}$ be two visceral influences in the domain of $e^{*}$. Then, using $e^{*}$ we have

$$\begin{aligned}&(1-\lambda)\left(-v^{'}+\sqrt{wv^{'}\left(g-a\right)}\right)+\lambda\left(-v^{"}+\sqrt{wv^{"}\left(g-a\right)}\right)\leq\\&-\left((1-\lambda)v^{'}+\lambda v^{"}\right)+\sqrt{w\left((1-\lambda)v^{'}+\lambda v^{"}\right)\left(g-a\right)}\end{aligned}.$$

Solving for $\lambda$ yields the condition

$$\lambda\leq\frac{\left(v^{"}-v^{'}\right)\left(g-a\right)}{\left(v^{"}-v^{'}\right)\left(g-a\right)}\equiv 1,$$

which holds $\forall\lambda\in\left(0,1\right)$, hence proving that optimal effort is concave in *v*. Again using the expression for optimal effort $e^{*}=-v+\sqrt{\omega v\left(g-a\right)}$ , we see that optimal effort is zero whenever *v* = 0, as the agent has no incentive to exert any effort when there is no visceral influence. Furthermore, using $0=-v+\sqrt{\omega v\left(g-a\right)}$ reveals that effort is zero also when $v=\omega\left(g-a\right)$. In addition, the derivative $\frac{\partial e*}{\partial v}=-1+\frac{1/2}{\sqrt{wv(g-a)}}$ , which equals zero, reaches its maximum whenever $v=\frac{1/4}{\omega(g-a)}$, and it is positive (negative) when $v$ is less (more) than $\frac{1/4}{\omega(g-a)}$. This implies that optimal effort is zero when $v$ is zero, increasing in $v$ until $v=\frac{1/4}{\omega(g-a)}$, after which point optimal effort falls and reaches zero at $v=\omega(g-a)$. Effort then follows an inverted U shape, which proves the proposition.

PROOF COROLLARY 1

The derivates of effort with respect to $g$, $\omega$ and $a$ are positive, positive and negative, respectively.

$$\partial e^{*}/\partial\omega=\frac{gv-va}{2\sqrt{\omega(gv-va)}}>0,$$

$$\partial e^{*}/\partial g=\frac{\omega v}{2\sqrt{\omega(gv-va)}}>0,$$

$$\partial e^{*}/\partial a=-v-\frac{1}{1}\left[\omega v(g-a)\right]\omega v>0,$$

Finally, substitute $g$ with $g^{\alpha}$ and $a$ with $a^{\alpha}$ to get $e^{*}=-sv+\sqrt{\omega sv\left(g^{\alpha}-a^{\alpha}\right)}$. It is evident that increasing $\alpha$ (increasing risk loving) raises optimal effort. Conversely, reducing $\alpha$ (increasing risk aversion) lowers optimal effort.

PROOF COROLLARY 2

Using $p = e^*/(e^* + v)$ *and* $e^* = -v + \sqrt{\omega v(g-a)}$, we can write $p = 1 - \frac{\sqrt{v}}{\sqrt{\omega(g-a)}}$. Increasing *v* and *a* reduces *p* and increasing $\omega$ and *g* raises *p*--which demonstrates the corollary.

PROOF LEMMA 1

Because a smaller *s* leads the agent to anticipate a smaller *v*, to engage in conflict always appears less costly than it actually is, and so naïve expected utility of conflict must always be greater than the true expected utility of conflict. The proof is straightforward and omitted.

PROOF LEMMA 2

First we use the condition $N \geq g - c$ ,and the definition of $N$ , to solve the condition for the willpower parameter:

$$\omega \geq sv\left[c(g-a) + 2a\sqrt{c(g-a)} + 2a^2\right] / c^2(g-a),$$

We then use $g - c \geq A$ ,and the definition of *A*, to again solve for the willpower parameter:

$$v\left[c(g-a)+2a\sqrt{c(g-a)}+2a^2\right]/c^2(g-a)\geq\omega,$$

This allows us to write the condition in full as:

$$v\Lambda\geq\omega\geq sv\Lambda,$$

where $\Lambda=\left[c(g-a)+2a\sqrt{c(g-a)}+2a^2\right]/c^2(g-a)$.

This completes the proof of lemma 2.

# APPENDIX B: SIMULATIONS

- For Referees' use only -

## B.1: Risk aversion

The below spreadsheet provides the basis for the simulations in Figure 6

| alpha (risk aversion parameter) | Cost of pre-commitment c | utility c (cost of pre-commitment) | willpower (w) | v (visceral influence) | A (Expected utility from conflict) | payoff from exercising restraint | Utility from exercising restraint (R ) | e* (Optimal effort) | payoff from I | Utility from indulging (I) | probability of R | probability of I | s (sophistication parameter) | e*N (Naive optimal effort) | probability of R (naive) | Probability of I (Naive) | N (Naive expected utility from conflict) | Utility from pre-commitment (PC) | Payoff from pre-commitment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,5 | 1,50 | 0,3 | 0,2 | 2,84 | 5 | 5,00 | 0,22 | 2 | 2,00 | 0,53 | 0,47 | 0,1 | 0,11 | 0,85 | 0,15 | 4,17 | 3,50 | 3,5 |
| 0,95 | 1,5 | 1,47 | 0,3 | 0,2 | 2,61 | 5 | 4,61 | 0,20 | 2 | 1,93 | 0,50 | 0,50 | 0,1 | 0,11 | 0,84 | 0,16 | 3,83 | 3,29 | 3,5 |
| 0,9 | 1,5 | 1,44 | 0,3 | 0,2 | 2,40 | 5 | 4,26 | 0,18 | 2 | 1,87 | 0,47 | 0,53 | 0,1 | 0,10 | 0,83 | 0,17 | 3,52 | 3,09 | 3,5 |
| 0,85 | 1,5 | 1,41 | 0,3 | 0,2 | 2,21 | 5 | 3,93 | 0,16 | 2 | 1,80 | 0,44 | 0,56 | 0,1 | 0,09 | 0,82 | 0,18 | 3,24 | 2,90 | 3,5 |
| 0,8 | 1,5 | 1,38 | 0,3 | 0,2 | 2,05 | 5 | 3,62 | 0,14 | 2 | 1,74 | 0,40 | 0,60 | 0,1 | 0,09 | 0,81 | 0,19 | 2,98 | 2,72 | 3,5 |
| 0,75 | 1,5 | 1,36 | 0,3 | 0,2 | 1,91 | 5 | 3,34 | 0,12 | 2 | 1,68 | 0,37 | 0,63 | 0,1 | 0,08 | 0,80 | 0,20 | 2,74 | 2,56 | 3,5 |
| 0,7 | 1,5 | 1,33 | 0,3 | 0,2 | 1,78 | 5 | 3,09 | 0,10 | 2 | 1,62 | 0,32 | 0,68 | 0,1 | 0,07 | 0,79 | 0,21 | 2,53 | 2,40 | 3,5 |
| 0,65 | 1,5 | 1,30 | 0,3 | 0,2 | 1,67 | 5 | 2,85 | 0,08 | 2 | 1,57 | 0,28 | 0,72 | 0,1 | 0,07 | 0,77 | 0,23 | 2,33 | 2,26 | 3,5 |
| **0,6** | 1,5 | 1,28 | 0,3 | 0,2 | **1,57** | 5 | 2,63 | 0,06 | 2 | 1,52 | 0,23 | 0,77 | 0,1 | 0,06 | 0,76 | 0,24 | **2,15** | **2,12** | 3,5 |
| **0,55** | 1,5 | 1,25 | 0,3 | 0,2 | **1,49** | 5 | 2,42 | 0,04 | 2 | 1,46 | 0,17 | 0,83 | 0,1 | 0,06 | 0,74 | 0,26 | **1,98** | **1,99** | 3,5 |
| 0,5 | 1,5 | 1,22 | 0,3 | 0,2 | 1,41 | 5 | 2,24 | 0 | 2 | 1,41 | 0 | 1 | 0,1 | 0,05 | 0,72 | 0,28 | 1,83 | 1,87 | 3,5 |
| 0,45 | 1,5 | 1,20 | 0,3 | 0,2 | 1,37 | 5 | 2,06 | 0 | 2 | 1,37 | 0 | 1 | 0,1 | 0,04 | 0,69 | 0,31 | 1,70 | 1,76 | 3,5 |
| 0,4 | 1,5 | 1,18 | 0,3 | 0,2 | 1,32 | 5 | 1,90 | 0 | 2 | 1,32 | 0 | 1 | 0,1 | 0,04 | 0,66 | 0,34 | 1,58 | 1,65 | 3,5 |
| 0,35 | 1,5 | 1,15 | 0,3 | 0,2 | 1,27 | 5 | 1,76 | 0 | 2 | 1,27 | 0 | 1 | 0,1 | 0,03 | 0,63 | 0,37 | 1,46 | 1,55 | 3,5 |
| 0,3 | 1,5 | 1,13 | 0,3 | 0,2 | 1,23 | 5 | 1,62 | 0 | 2 | 1,23 | 0 | 1 | 0,1 | 0,03 | 0,59 | 0,41 | 1,37 | 1,46 | 3,5 |
| 0,25 | 1,5 | 1,11 | 0,3 | 0,2 | 1,19 | 5 | 1,50 | 0 | 2 | 1,19 | 0 | 1 | 0,1 | 0,02 | 0,53 | 0,47 | 1,28 | 1,37 | 3,5 |
| 0,2 | 1,5 | 1,08 | 0,3 | 0,2 | 1,15 | 5 | 1,38 | 0 | 2 | 1,15 | 0 | 1 | 0,1 | 0,02 | 0,46 | 0,54 | 1,20 | 1,28 | 3,5 |
| 0,15 | 1,5 | 1,06 | 0,3 | 0,2 | 1,11 | 5 | 1,27 | 0 | 2 | 1,11 | 0 | 1 | 0,1 | 0,01 | 0,36 | 0,64 | 1,13 | 1,21 | 3,5 |
| 0,1 | 1,5 | 1,04 | 0,3 | 0,2 | 1,07 | 5 | 1,17 | 0 | 2 | 1,07 | 0 | 1 | 0,1 | 0,00 | 0,19 | 0,81 | 1,08 | 1,13 | 3,5 |
| 0,05 | 1,5 | 1,02 | 0,3 | 0,2 | 1,04 | 5 | 1,08 | 0 | 2 | 1,04 | 0 | 1 | 0,1 | 0 | 0 | 1 | 1,04 | 1,06 | 3,5 |
| 0 | 1,5 | 1,00 | 0,3 | 0,2 | 1,00 | 5 | 1,00 | 0 | 2 | 1,00 | 0 | 1 | 0,1 | 0 | 0 | 1 | 1 | 1,00 | 3,5 |

## B.2: Visceral influence

The below spreadsheet provides the basis for the simulations in Figure 7

| Sophistication parameter s | payoff (and utility) from the goal g | payoff (and utility) from the tempting alternative a | Visceral influence v | cost of pre-commitment c | Naive expected utility from conflict N | Actual expected utility from conflict A | willpower parameter w | Probability of success (actual) | Ptobability of failure (actual) | Probability of success (naive) | Probability of failure (naive) | Optimal effort (actual) | Optimal effort (Naive) | Utility from pre-commitment u(g-c) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0,3 | 20 | 5 | 1 | 10 | 16,76 | 13,25 | 1 | 0,74 | 0,26 | 0,91 | 0,09 | 2,87 | 1,82 | 10 |
| 0,3 | 20 | 5 | **2** | 10 | 15,39 | **11,05** | 1 | 0,63 | 0,37 | 0,85 | 0,15 | 3,48 | 2,40 | **10** |
| 0,3 | 20 | 5 | **3** | 10 | 14,30 | **9,58** | 1 | 0,55 | 0,45 | 0,80 | 0,20 | 3,71 | 2,77 | **10** |
| 0,3 | 20 | 5 | 4 | 10 | 13,32 | 8,51 | 1 | 0,48 | 0,52 | 0,76 | 0,24 | 3,75 | 3,04 | 10 |
| 0,3 | 20 | 5 | 5 | 10 | 12,40 | 7,68 | 1 | 0,42 | 0,58 | 0,71 | 0,29 | 3,66 | 3,24 | 10 |
| 0,3 | 20 | 5 | 6 | 10 | 11,50 | 7,03 | 1 | 0,37 | 0,63 | 0,66 | 0,34 | 3,49 | 3,40 | 10 |
| 0,3 | 20 | 5 | **7** | 10 | **10,60** | 6,51 | 1 | 0,32 | 0,68 | 0,61 | 0,39 | 3,25 | 3,51 | **10** |
| 0,3 | 20 | 5 | **8** | 10 | **9,68** | 6,09 | 1 | 0,27 | 0,73 | 0,55 | 0,45 | 2,95 | 3,60 | **10** |
| 0,3 | 20 | 5 | 9 | 10 | 8,72 | 5,76 | 1 | 0,23 | 0,77 | 0,49 | 0,51 | 2,62 | 3,66 | 10 |
| 0,3 | 20 | 5 | 10 | 10 | 7,72 | 5,51 | 1 | 0,18 | 0,82 | 0,43 | 0,57 | 2,25 | 3,71 | 10 |

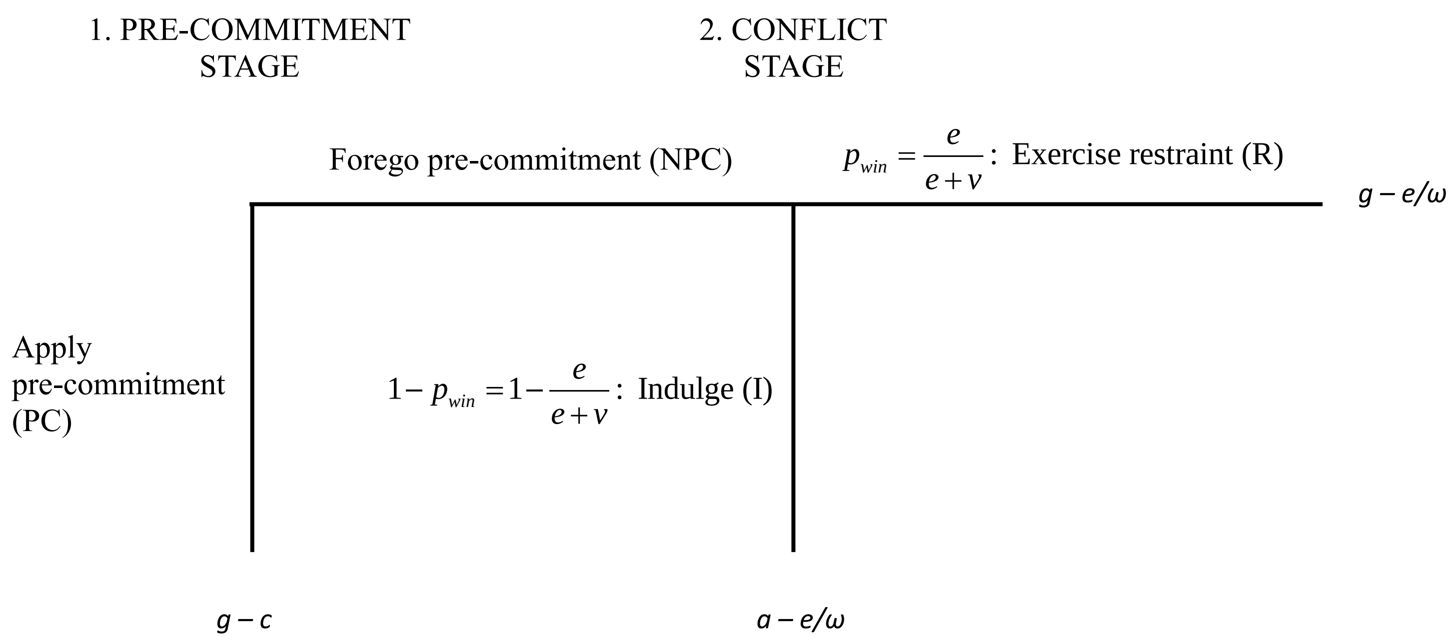


FIGURE 1. THE DECISION PROBLEM

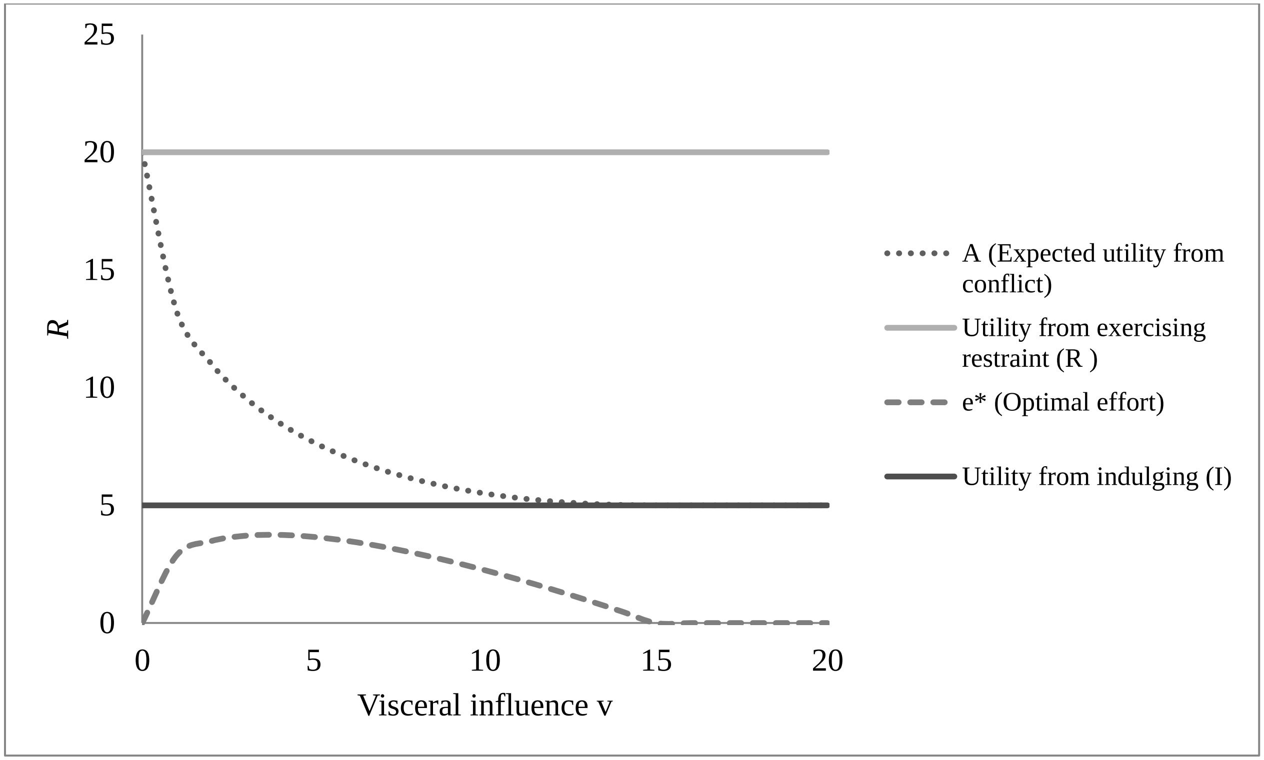


FIGURE 2: EXPECTED UTILITY OF CONFLICT, UTILITY FROM EXERCISING RESTRAINT, OPTIMAL EFFORT, AND UTILITY FROM INDULGING AS FUNCTIONS OF VISCERAL INFLUENCE

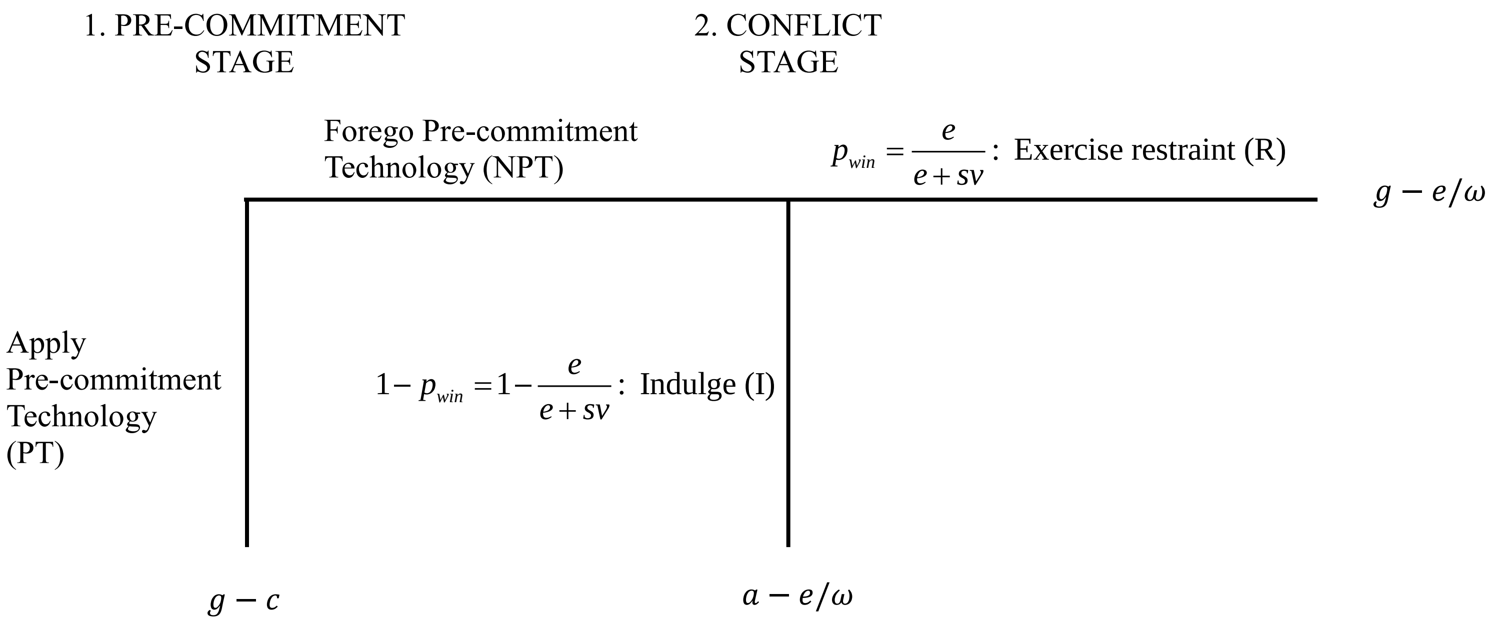


FIGURE 3. THE EXTENDED DECISION PROBLEM:
UNDERESTIMATING VISCERAL INFLUENCES ON BEHAVIOR

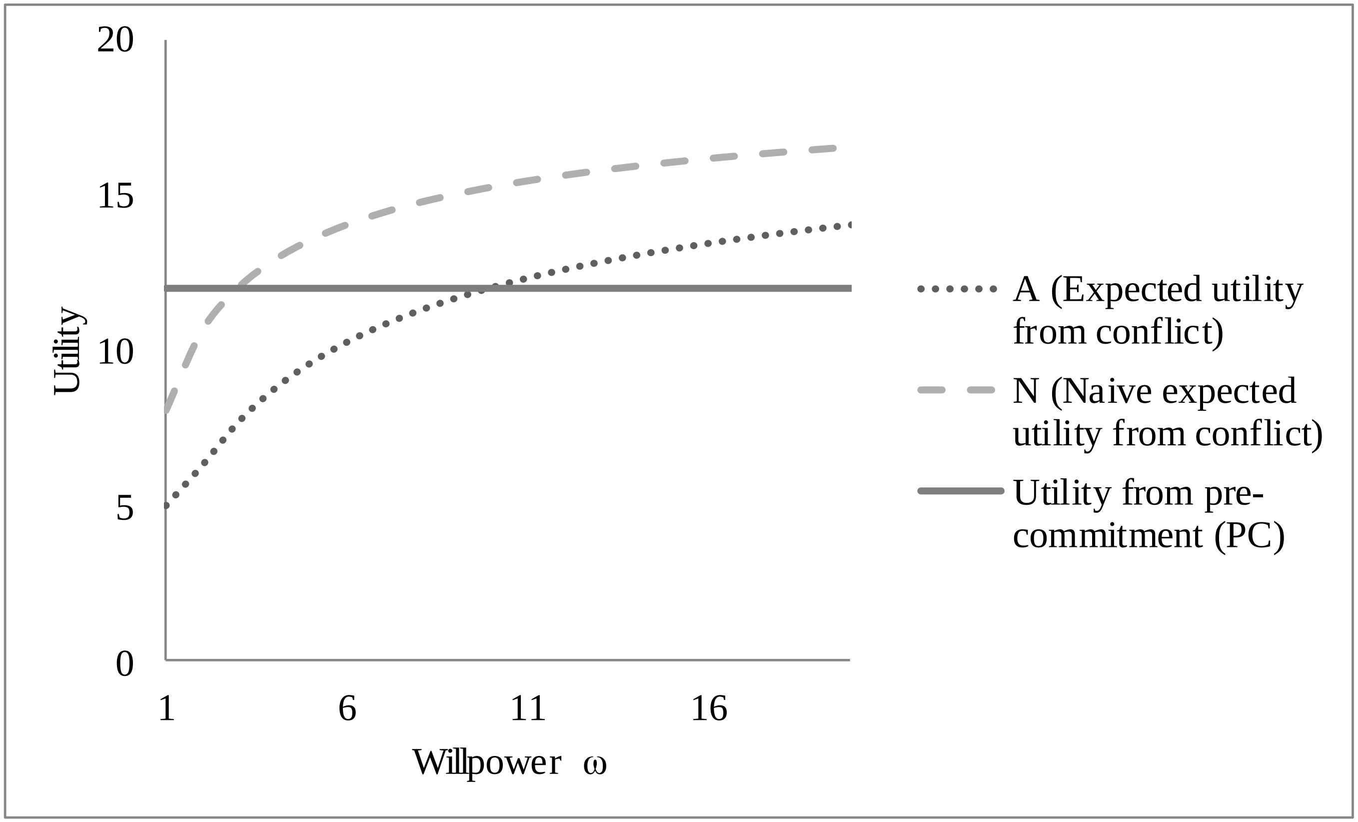


FIGURE 4: EXPECTED UTILITY FROM CONFLICT, NAÏVE EXPECTED UTILITY FROM CONFLICT, AND UTILITY FROM PRE-COMMITMENT AS FUNCTIONS OF WILLPOWER

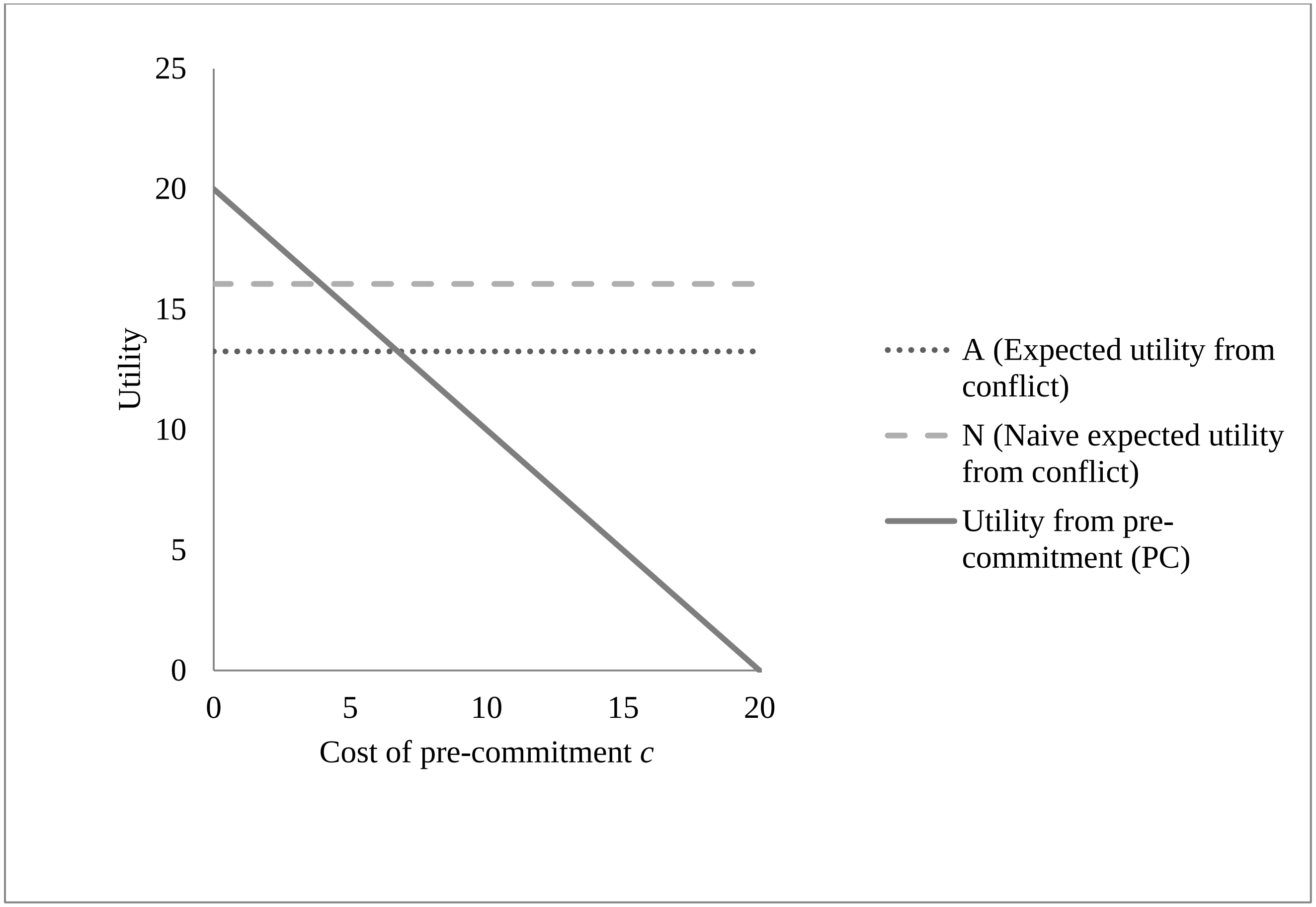


FIGURE 5: EXPECTED UTILITY OF CONFLICT, NAÏVE EXPECTED UTILITY OF CONFLICT, AND UTILITY OF PRE-COMMITMENT AS FUNCTIONS OF COST OF PRE-COMMITMENT

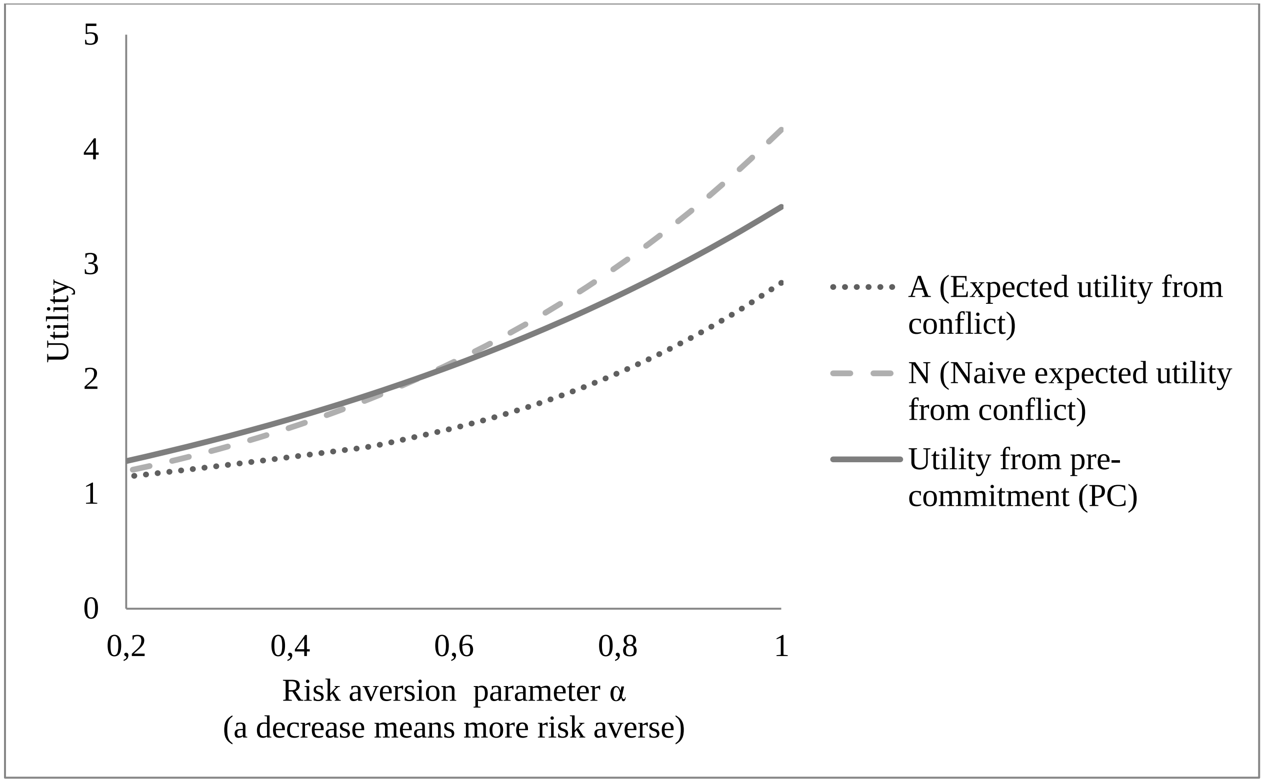


FIGURE 6: EXPECTED UTILITY OF CONFLICT, NAÏVE EXPECTED UTILITY OF CONFLICT,

AND UTILITY OF PRE-COMMITMENT AS FUNCTIONS OF RISK AVERSION

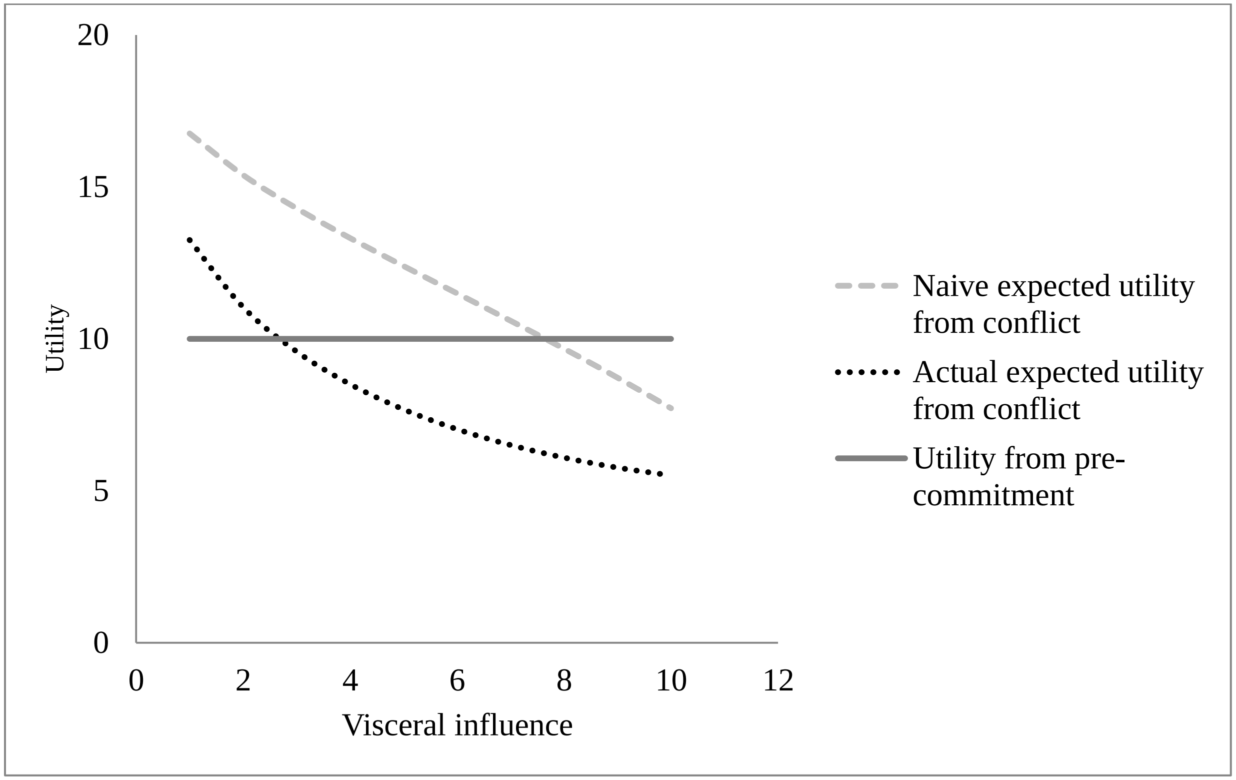


FIGURE 7: EXPECTED UTILITY OF CONFLICT, NAÏVE EXPECTED UTILITY OF CONFLICT, AND UTILITY FROM PRE-COMMITMENT AS FUNCTIONS OF VISCERAL INFLUENCE