Journal of Economic Behavior & Organization 101 (2014) 100–112

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

How to commit (if you must): Commitment contracts and the dual-self model

Alexander Peysakhovich [*,1]

Program for Evolutionary Dynamics, Harvard University, United States

ARTICLE INFO

Article history:
Received 6 March 2012
Received in revised form 4 February 2014
Accepted 6 February 2014
Available online 26 February 2014

Keywords:
Behavioral economics
Self-control
Dual-self model
Decision-making
Commitment contracts

ABSTRACT

This paper studies how dual-self (Fudenberg and Levine, 2006) decision-makers can use commitment technologies to combat temptation and implement long-run optimal actions. I consider three types of commitment technologies: carrot contracts (rewards for 'good' behavior financed by borrowing from future consumption), stick contracts (self imposed fines for 'bad' behavior) and binding commitment. I compare the welfare implications of these contracts and show that dual-self decision-makers strictly prefer to use carrots instead of either sticks or binding commitments. This is for several reasons: sticks are highly vulnerable to trembles (while carrots are not), sticks and binding commitments create a temptation to cancel them (while carrots do not), and finally carrots allow easy tradeoffs between commitment and flexibility (while sticks and binding commitments do not).

1. Introduction

Many of us would like to exercise, work efficiently and stay away from our bad habits yet we often find ourselves skipping a daily run, looking at funny cat videos on the internet, smoking another cigarette or eating another cupcake. Such mismatches between what we would like to do tomorrow and what we actually end up doing create a demand for technologies to help individuals implement their normative goals. Existing literature[2] shows that this demand for commitment does exist but there is little theoretical work on what forms optimal commitment technologies would take. This paper begins to bridge this gap.

The first step to figuring out optimal commitment technologies is learning what mechanism generates the problem in the first place. Work in psychology and neuroscience tends to focus on decisions as an interplay between an automatic and a controlled process (Kahneman, 2003). Recent work in social psychology makes this model more sharp by positing that the controlled cognitive process uses a costly resource to operate. A standard template in such experiments involves subjects performing a 'cognitive control task' (controlling attention, suppressing emotion, solving math problems) or a task that does

* Tel.: +1 (617) 496 4737.
E-mail address: apeysakh@fas.harvard.edu

[1] I thank Dan Burghart, Adam Brandenburger, Paul Glimcher, Drew Fudenberg, Oliver Hart, Yuichiro Kamada, David Laibson, Aurélie Ouss, David Rand, Al Roth, Tomasz Strazlecki, Ryan Webb, and the participants of the Harvard Economics Theory Lunch for incredibly helpful advice and comments on this project. I thank anonymous referees whose comments have greatly improved both the focus and quality of this paper. All errors can be blamed on the lack of enforceable commitment mechanisms.

[2] Ashraf et al. (2006), Kaur et al. (2010), Ariely and Wertenbroch (2002) consider commitment in the field, Houser et al. (2010) consider commitment in the lab. Bryan et al. (2010) survey the existing research.

http://dx.doi.org/10.1016/j.jebo.2014.02.003
0167-2681/

not require such control followed by a second cognitive control task. Subjects who had to perform the depleting task do worse on the second task than controls (see Muraven and Baumeister, 2000 for a survey). Another set of experiments involve individuals making choices under cognitive load (for example, subjects are asked to remember a 7 digit number), these individuals then act more impulsively than controls (e.g. by choosing more unhealthy foods to eat as in Shiv and Fedorikhin, 1999). In such a model our difficulties at the desert tray come from an interplay of automatic impulses compelling us to eat, and a conscious use of mental resources not to give in.[3]

This paper uses a particular economic model of this process: the dual-self model of Fudenberg and Levine (2006) (henceforth FL). This model is a specific case of a larger set of costly self-control models.[4] In such models decisions are a compromise between a 'temptation' ranking and a 'normative' ranking, with the DM balancing a desire to choose according to his normative preference with a self-control cost of deviating from the temptation ranking. The fully dynamic version of the model (Fudenberg and Levine, 2012) captures the experimental results above by adding a form of convexity to the control costs: the more self-control has been recently used, the higher the marginal cost of resisting new temptations.

The FL model imposes specific restrictions on where the disagreement between the temptation and normative preferences comes from: the DM uses a standard time-separable utility function to evaluate consumption streams (or dynamic plans) but the temptation or automatic process discounts the future at a much sharper rate than the DM would like. Thus the DM is tempted to behave impulsively and must use self-control to choose long-run rewards. Because of this structure FL refer to the automatic impulse as the short-run (SR) self and the cognitive control process as the long-run (LR) self.[5] Note that this is a different type of model than those studied in the literature on hyperbolic/quasi-hyperbolic discounting (e.g. Laibson, 1997; Ainslie and Haslam, 1992). In those models the DM's problems come from the fact that rankings of alternatives change in different periods. In the FL model both the SR and LR self are perfectly time consistent so the tension comes from multiple preferences within a period rather than multiple preferences between periods. The purpose of this paper is not to provide a clean test to differentiate these models, rather it is to look at commitment behavior with the FL model. However, all results that could never hold under a time-inconsistent framework are flagged as such.

The FL model generates a huge demand for commitment and this paper considers two types of technologies that could be available to a DM facing temptation. The first are stick contracts that levy a fine on the DM when he gives in to temptation.[6] The second are carrots that reward a DM who takes normatively good actions. In this paper, carrots are financed by the intertemporal substitution of future consumption to the present conditional on the DM resisting a temptation. The main results show that both of these types of contracts can only be welfare improving for a DM if they change the nature of the temptation he faces – i.e. the SR self's optimal action. The logic is one of revealed preference combined with the economic idea that self-control is treated as a cost. If a DM gives in to a temptation, this is because the self-control required to resist it was too expensive. If a commitment technology does not physically remove the tempting option, then it must remove the temptation associated with that option because if it does not, its ultimate effect is only to make the DM exert the self-control that he didn't find optimal to exert in the first place. This means that if the source of the temptation is sharp discounting, both types of contracts must have one of two features: either their effects must be close in time to the choice or they must be particularly large. In fact, as the paper shows later, their size increases exponentially with the delay between action and contract implementation.

The natural question to ask given these results is whether carrots, sticks or binding commitments are favored by the dual-self DMs. The final set of results shows that carrots are preferred for several reasons. First, sticks will implement a punishment if the DM trembles and executes unintended actions with small probability whereas a tremble in the presence of a carrot makes the DM no worse off than before the contract. Second, sticks and binding commitments require self-control to implement in the first place and if the DM receives opportunities to cancel the contract, the cancellation acts as an additional temptation and source of self-control problems in the case of sticks and binding commitments but not carrots. Finally, the DM may have a desire for flexibility. If the temptation's size stochastic optimal carrots allow the DM to retain the flexibility choose the temptation when it is LR optimal while sticks and binding commitments do not.

The results in this paper are related to those of Ali (2011) who shows that carrots are advantageous for DMs seeking flexibility in a planner-doer model (Thaler and Shefrin, 1981) of self-control in which the planner attempts to learn about the doer's preferences. From a theoretical perspective, this paper together with the results of Ali (2011) indicate that there is much to be learned by adding dynamic behavior into models of self-control. More practically they also indicate that devices beyond binding commitments and self-punishing technologies may be useful ways of dealing with self-control problems.

---

[3] Work in neuroeconomics, eg. Hare et al. (2009), McClure et al. (2004, 2007), points to a possible neural algorithm for this model in which control networks in the dorsolateral prefrontal cortex modulate 'overreactions' by the brain's reward system.

[4] Gul and Pesendorfer (2001, 2004), Dekel et al. (2009), Noor and Takeoka (2010) are highly visible examples in which self-control costs are variable and depend on the amount of adjustment the control process needs to do. Benhabib and Bisin (2005) consider the case where the control process is treated as having a fixed cost to activate.

[5] The analysis of Fudenberg and Levine (2006) considers the case where the SR self is perfectly myopic, Fudenberg and Levine (2012) extends to the case where the SR self can have some degree of patience.

[6] A market version of such contracts is provided by the website StickK.com that allows individuals to set measurable goals, punishments and a referee. If an individual does not accomplish his goal, as reported by the referee, the website will automatically charge the individual's credit card a donation to an 'anti-charity' of his choice (for example, a life-long Democrat may choose to donate to finance the George W. Bush Presidential Library.)