Department of Economics

DISCUSSION PAPER SERIES IN ECONOMICS

DP No. 29

Torn between want and should: Self regulation and behavioral choices

May 2020

Abhinash Borah
Raghvi Garg

https://www.ashoka.edu.in/ecodp

Plot No. 2, Rajiv Gandhi Education City,
National Capital Region
P.O.Rai,Sonepat Haryana-131029

Torn between want and should: Self regulation and behavioral choices

Abhinash Borah and Raghvi Garg*

May 17, 2020

Abstract

We model the behavior of a decision maker (DM) who faces an intrapersonal conflict between what she wants to do (her "want-self") and what she thinks she should do (her "should-self"). In our model, in any choice problem, the DM first eliminates the worst alternative(s) according to the preferences of her should-self, presumably, as a way of managing the guilt that results from making choices she should not. Then, from the remaining alternatives, she chooses the best one according to the preferences of her want-self. Drawing on Freud, we interpret this choice procedure as reflective of the balancing act that preserving one's ego requires. Indeed, this balance is key to a DM's ability to exercise self regulation which our model analyzes in the context of behavioral choices. We characterize the model behaviorally and identify the extent to which the key behavioral parameters can be uniquely identified from choice data.
JEL codes: D01, D91
Keywords: intrapersonal conflict, want and should selves, ego preserving heuristic, self regulation and ego depletion, behavioral choices

1 Introduction

Consider the following two examples:[1]

Example 1: A decision maker (DM) who is on a diet walks into a subs and salad outlet for lunch and has to decide whether to order a cola with her meal or avoid it. If she chooses

*Department of Economics, Ashoka University, Sonepat, Haryana - 131029, India. Emails: abhinash.borah@ashoka.edu.in, raghvi.garg_phd17@ashoka.edu.in. We are grateful to Antonio Penta, Ran Spiegler, Debasis Mishra and Anujit Chakraborty for helpful discussions. We would also like to thank participants at the Behavioral Research in Economics Workshop (2019) at IIM Ahmedabad, Winter School (2019) at Delhi School of Economics, CoRe IGIDR, and Delhi Economic Theory Workshop for comments. Any shortcomings in the paper are, of course, our responsibility.

[1] Examples similar in spirit to these ones have been referenced prominently in the literature on temptation and self control, e.g., in Fudenberg and Levine (2006), Dekel, Lipman, and Rustichini (2009), Noor and Takeoka (2015) and Masatlioglu, Nakajima, and Ozdenoren (2020).

to order it, the only option available on the menu is a small (300 ml) serving of cola. In this case, she decides to forego ordering the cola. In another instance, when she is in an identical outlet for lunch, she is confronted with a similar decision. But, this time around, the menu not only has the option of the small serving of cola but also a large (750 ml) one. In this case, she ends up ordering the small serving of cola with her meal. Such patterns of choice, highlighting what is referred to as the compromise effect, are quite common and have been reported in the experimental literature going at least as far back as Simonson (1989) and Simonson and Tversky (1992). Indeed, exploiting the compromise effect is one of the most common strategies that marketers employ as a way of boosting sales.

Example 2: Experiments based on the dictator game have been widely conducted in labs and fields to study allocation decisions in non-strategic settings. In the standard version of this two player game, one of the players (the "dictator") is made responsible for deciding how a given endowment (of, say, money) is to be divided between the two of them. It has been documented that on average, lab dictators give about 20% of the endowment to the recipient (Camerer, 2011). Despite the widespread replication of altruistic behavior in this standard setting, it has been found that giving is highly sensitive to minor tweaks made in the design [Dana, Weber, and Kuang (2007), Hoffman et al.(1996), Franzen and Pointner (2012), Cherry et al. (2002), Eckel and Grossman (1996)]. One such design variation is in the set of options available to the dictator. A few papers [Bardsley (2008), List (2007)] expand this set to allow for taking, i.e., the dictator not only has the option of sharing the endowment between the recipient and herself in any way she wishes but may also now take part of the recipient's endowment. If our understanding is that giving in the standard version of the game is due to an altruistic motivation, then this giving should remain unaltered in the take version of the game as well. However, when taking is an option, significantly fewer subjects make a positive transfer to the recipient. For instance, in List (2007), in the control setting, both players were allocated \$5, with dictators allocated an additional \$5 that they could divide with recipients in any way they chose (in \$0.50 increments). In this setting, 71% of subjects made positive transfers to the recipients with median and mean transfers being \$1 and \$1.33, respectively. On the other hand, in one of the treatments in which besides the different options available to divide the \$5 like in the control, dictators could additionally take \$1 from the recipient, only 35% of subjects made positive transfers and median and mean transfers were \$0 and \$0.33, respectively.

Although these two examples may appear to be from quite different domains—one of an individual consumer with a self control problem and the other pertaining to distributional preferences—in this paper, we propose a choice procedure that sees them as part of a unified, psychologically informed framework of human behavior. The starting point of the procedure that we propose is the observation that an individual's desires, motivations and objectives operate at different levels. Whereas some of these are in the nature of primitive passions and pleasures, quite often of an instinctual and even impulsive disposition, others are more evolved needs that capture certain ideals or moral judgments. We may call the first, the DM's *wants* and the second, her *shoulds*. In many decision problems both wants and shoulds factor in and an intrapersonal conflict becomes inevitable as ignoring the opposing demands of either is psychologically costly for the DM.[2] If her "want-self" is ignored, then the DM may be subject to anxiety and tension. On the other hand, if her "should-self" is compromised, she may be burdened with feelings of guilt. As such, efficacious decision making may often require the DM to draw a balance between these two opposing demands. The choice procedure that we introduce here speaks to this and proposes one way by which the DM may strike this balance. The procedure assumes that in any given choice problem, the DM is able to rank all the available alternatives according to both her should and want selves. The way she then chooses is the following. Given any set of alternatives, first, she eliminates a subset of these alternatives that are the worst according to the preferences of her should-self. Then, amongst the remaining alternatives, she chooses the one that is the best according to the preferences of her want-self. In our baseline model, the DM follows a simple heuristic of eliminating the worst alternative according to her should-self in any choice problem that she faces. In a generalization of this model, we consider the case where she may eliminate multiple alternatives.

To understand the choice procedure better, consider Example 1 above. In it, given that the DM is on a diet, she knows that she *should* avoid sugarated drinks. At the same time, owing to habit, she may naturally *want* to have a cola with her meal. How does she go about making her choice in this situation? If she chooses according to our simple heuristic of eliminating the worst alternative according to her should-self, then in the first situation, when the choice is between ordering a small serving of cola or no cola, the former which is the worst of the two alternatives according to her should-self is eliminated and she ends up ordering no cola. On the other hand, when the large serving of cola is added to the menu, this becomes the worst alternative according to what she should do and, therefore, gets eliminated. This leaves her with the option of ordering either the small serving of cola or no cola and her want-self makes her choose the former.

A similar reasoning explains choices in Example 2 pertaining to the dictator game. Here, what the DM may *want* to do is act selfishly and pocket as much of the available endowment as possible. On the other hand, her ideals and moral judgments force her to acknowledge that she *should* act more altruistically. Going back to the control setting in List (2007) in which \$5 can be shared, a DM who follows our generalized model, may eliminate, say, the two worst alternatives according to what she should do. These would be the alternatives that give \$0 and \$0.50 to the recipient. Then, amongst the remaining alternatives, she chooses the best one according to what she wants, resulting in a transfer of \$1 and the choice of the allocation $(\$9, \$6)$. On the other hand, in the treatment in which she can also take \$1 from the recipient, given the change in context introduced by the take option, she may

[2]As we discuss below, the idea that such a fracture exists in the human psyche can be traced back in the psychology literature to at least as far as Freud (1921/1961).

eliminate, say, just the worst alternative according to her should-self that involves taking the \$1. Then, amongst the remaining alternatives, she chooses the best one according to what she wants, resulting in the choice of the allocation (\$10, \$5). That is, she ends up making a zero transfer.

Before proceeding further, some discussion on why our choice procedure is psychologically plausible seems appropriate. In this regard, we are particularly influenced by how Sigmund Freud conceptualized the human psyche and personality. In his landmark text, "The Ego and the Id," Freud conceived of the human psyche in terms of three components, the *id*, the *ego* and the *superego* (Freud, 1921/1961). According to Freud, these three components of the psyche each with their unique features interact in complex ways to produce human behavior. The id which is present from birth is the primitive and instinctive component of personality and is animated by our basic urges, needs, and desires. The superego on the other hand, which starts developing around the age of four-five, is the repository of our values, morals and ideals. Finally, the ego is the decision making part of personality that mediates between the primitive demands of the id and the more evolved ideals of the superego. The ego's job, therefore, is a balancing one—it tries to find realistic ways of satisfying the id's desires while at the same time trying not to deviate too far from the ideals of the superego. If it doesn't get this balance right, then it is psychologically costly—ignoring the id's demands results in anxiety and tension whereas straying away from that of the superego's produces guilt. The choice procedure that we propose here has a similar spirit as this balancing job that the ego performs.[3] In this regard, it is particularly noteworthy to focus on the language, albeit in translated form, that Freud uses to explain how the ego, despite very often being compliant to the demands of the id, reconciles with the ideals of the superego. At many places, this explanation draws on such words and ideas as (an internal) "prohibition", "repression" and "moral censorship" of many of the id's desires that run counter to the superego's higher aspirations. Indeed, Freud situates the very origins of the human conscience in the act of a (male) child finally being able to repress and overcome the Oedipus complex around the age of six. Our conception of eliminating the worst alternative(s) according to the DM's should-self is very much in the spirit of this Freudian analysis and speaks to the balancing act that the ego performs. Because of this, we refer to our choice procedure as the *ego preserving heuristic*.

In recent years, one psychological theory that has been inspired by Freud is the *strength model of self regulation (or self control)* (Baumeister, Schmeichel, and Vohs, 2007). The term self regulation refers to an individual's ability to alter her responses or behavior, including overriding one response or behavior and replacing it with a more desired one.[4] A key idea underlying this theory is that acts of effortful self regulation draw on a limited

[3]As should be evident, in terms of the language we adopt here, we may identify the want-self with the id and the should-self with the superego.

[4]At times, the terms self regulation and self control are used interchangeably but strictly speaking self regulation encompasses a broader set of ideas than self control.

energy resource and this resource gets depleted from repeated exertions, akin to how an overused muscle loses strength, making subsequent efforts at self regulation harder—a phenomenon that has been referred to as *ego depletion* in homage to Freud.[5] Our model can be seen in the light of this theory as the effortful and agential act of eliminating the worst alternative(s) according to the should-self is precisely an expression of self regulation. One reason why the theory of self-regulation is useful is because it provides a unified framework to study a wide range of phenomena. For instance, the connection that we drew between the two examples above using our theory is not just a theoretical possibility. There is self regulation based experimental evidence, e.g., Achtziger, Alós-Ferrer, and Wagner (2015), Halali, Bereby-Meyer, and Ockenfels (2013) and Martinsson, Myrseth, and Wollbrant (2012), amongst others, that establishes a connection between self control and social preferences—in general, an ability to exercise greater self control appears to induce greater pro-sociality. Indeed, in some of this work, e.g., Xu, Bègue, and Bushman (2012), it has been shown, like we suggest, that guilt may precisely be the emotion through which this connection runs; specifically, greater ego depletion (i.e., lower self control) decreases guilty feelings and produces more selfish behavior. Therefore, to summarize the discussion: although it is possible to interpret our model without any psychological motivation—as simply a model featuring an ordered pair of rationales, with the DM first eliminating the worst alternative(s) according to the first rationale and, then, choosing the best one according to the second—we believe that the motivation we have provided above adds significantly to the interpretation of the model's behavioral scope, including understanding how its rationale carries over across choice domains.

In this paper, we formalize the ego preserving heuristic (EPH) and its generalization, the generalized ego preserving heuristic (GEPH), in the tradition of behavioral choice theory (BCT) models that employ two-stage, sequential choice procedures [e.g., Manzini and Mariotti (2007), Cherepanov, Feddersen, and Sandroni (2013), Masatlioglu, Nakajima, and Ozbay (2012), amongst others]. The difference between the EPH and GEPH is that in the EPH, in any choice problem, the DM eliminates the worst alternative according to her should-self whereas in the GEPH she may eliminate multiple alternatives. Our key analytical results are the behavioral characterization of these models. That is, we provide testable conditions on behavior (choice data) that an outside observer can use to determine whether choices of a DM agree with the EPH/GEPH. We also determine to what extent the parameters of the model, specifically the two rankings reflecting the DM's want and should selves, and additionally in the case of the GEPH, the set of alternatives eliminated by the should-self in a choice problem, can be uniquely elicited from choice data. Along with the rankings, the identification of the set of eliminated alternatives is meaningful as it allows the analyst to gather information about important psychological drivers of behavior like the level of ego depletion that a DM may be experiencing at the time of choosing—e.g., a smaller number of eliminated alternatives may indicate higher levels of ego depletion.

[5] This limited resource is used not just for self regulation but also for all acts of volition, expressions of agency, effortful choice and active cognitive responses that together constitute the self's executive function.

Overall, this exercise allows us to connect the theme of self regulation to the BCT literature and explore its behavioral foundations in the light of this literature. That is, we answer the question as to what the identification of the psychological phenomenon of exercising self-regulation is in the domain of "non-standard" choices that the BCT literature has endeavored to study. Later in the paper, we compare the empirical content of our model with other related BCT models in the literature. We also use the model to develop some insights about the compromise effect as well as talk about welfare implications in the domain of non-standard choice data.

There is a long tradition in economics of studying decision problems that feature an intrapersonal conflict going at least as far back as Strotz (1955). Strotz and some of the other early papers in this area like Peleg and Yaari (1973) and Blackorby et al. (1973) analyzed such problems from the paradigm of changing tastes. The perspective on intrapersonal conflict that we take in this paper of a DM operating under the influence of two sets of conflicting preferences at the same instance can be traced back in the economics literature to Thaler and Shefrin (1981) and Schelling (1984). Within this perspective of intrapersonal conflict, our work relates most to those models in which the DM is able to exercise self control or self regulation at the time of making choices, very often by exercising costly willpower. In recent years, a major fillip to this line of enquiry has been provided by the pioneering model of temptation and self control introduced in Gul and Pesendorfer (2001) [GP]. Building on the framework of Kreps (1979), GP model a two-stage decision problem confronting a DM who faces temptation while choosing from a menu but is also sophisticated to recognize this at an ex ante stage while choosing between menus to face at the second stage. Our model connects to the second stage of the GP decision problem, where like in our set-up, the DM is able to exercise self-control. Specifically, their model conceives two utility functions u and v over the set of alternatives representing the DM's commitment and temptation perspectives, respectively, such that choice in the second stage from a menu is specified by:[6]

$$c(S) = \operatorname*{argmax}_{x \in S} \left\{ u(x) - \left[\max_{y \in S} v(y) - v(x) \right] \right\} = \operatorname*{argmax}_{x \in S} \left\{ u(x) + v(x) \right\}$$

In other words, the value to the DM of choosing an alternative x from the menu S is given by the commitment utility $u(x)$ adjusted for the cost of self control, $\max_{y \in S} v(y) - v(x)$, involved in choosing x from a menu whose most tempting alternatives are in the set $\operatorname{argmax}_{y \in S} v(y)$. Observe that the GP formulation implies, unlike in our set-up, that choices from menus satisfy the weak axiom of revealed preferences (WARP), which is the benchmark for the standard rational choice approach in economics that models a DM's behavior as resulting from the maximization of a single preference relation.[7] This difference between our paper and the GP framework has been narrowed by subsequent innovations

[6] In the GP framework, alternatives in a menu are lotteries.

[7] It should be noted that when we talk about WARP in the GP framework we do so with respect to a choice correspondence, whereas in our set-up the reference is to a choice function.

introduced to this framework, specifically by Noor and Takeoka (2010) and Fudenberg and Levine (2006, 2011, 2012),[8] who allow self control costs to be convex instead of linear; Noor and Takeoka (2015), who models menu dependent self control costs; and Liang, Grant, and Hsieh (2019) and Masatlioglu, Nakajima, and Ozdenoren (2020), who introduce the ego depletion or limited willpower perspective. In all these papers, choices from menus can violate WARP. For instance, like us, all of these models can accommodate the compromise effect. Another place where the GP framework connects with ours is in the role that the most tempting alternative in a menu plays in determining the difficulty of exercising self-control in that menu. As we will see, a similar idea plays a role in the sharpest behavioral characterization of the GEPH model. Without attempting to be exhaustive, we should point out that the ability to exercise self regulation or self control has also been modeled in Benhabib and Bisin (2005), Ozdenoren, Salant, and Silverman (2012) and Loewenstein, O'Donoghue, and Bhatia (2015), amongst others.

The rest of the paper is organized as follows. In the next section we define and provide a behavioral characterization of the EPH model. In Section 3 we do the same for the GEPH model. Then, in Section 4 we situate our work in the context of some related models in the literature. Section 5 explores some insights about the compromise effect in the context of our model. Finally, Section 6 discusses welfare implications. Proofs of results appear in the Appendix.

2 The Ego Preserving Heuristic

We consider a decision maker (DM) and the choices she makes in different choice problems. Formally, let X be a finite set of alternatives with typical elements denoted by x, y, z etc. $\mathcal{P}(X)$ denotes the set of non-empty subsets of X with typical elements R, S, T etc., which we refer to as choice problems. A choice function $c : \mathcal{P}(X) \to X$ is a mapping that, for any $S \in \mathcal{P}(X)$, specifies the alternative $c(S) \in S$ that the DM chooses in that choice problem.

In the model that we develop, the DM arrives at her choice in any choice problem by a two stage sequential procedure. Specifically, in each of the stages, she makes use of a distinct strict preference ranking on X (By a strict preference ranking, we mean a binary relation that is total, asymmetric and transitive). One may think of the ranking associated with the first stage, denoted by $\succ^* \subseteq X \times X$, as reflecting what she thinks *she should do*; and that associated with the second stage, denoted by $\succ \subseteq X \times X$, as reflecting what *she wants to do*. Faced with any choice problem, in the first stage, the DM eliminates a

[8]Fudenberg and Levine model the interactions between a long-run self and a sequence of myopic short-run selves. The short-run self makes choices but the long-run self can influence these choices at a cost, an interaction which can equivalently be viewed in the light of the GP representation with either a linear or convex cost of self control.

set of alternatives that are worst according to $\succ^*$. Then, in the second stage, amongst the remaining alternatives, she picks the best one according to $\succ$. We refer to the set of alternatives eliminated in the first stage as the DM's *should not set* in the choice problem under consideration. To define this formally, let $\mathcal{P}^*(X)$ denote the set of non-singleton subsets of $\mathcal{P}(X)$.

Definition 2.1. *For any $S \in \mathcal{P}^*(X)$, $W_{\succ^*}(S) \subsetneq S$, $W_{\succ^*}(S) \neq \emptyset$, is a should not set of S if for any $x \in S \setminus W_{\succ^*}(S)$ and $y \in W_{\succ^*}(S)$, $x \succ^* y$. A should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ specifies for each $S \in \mathcal{P}^*(X)$ a should not set $W_{\succ^*}(S)$.*

That is, a should not set $W_{\succ^*}(S)$ of a set $S \in \mathcal{P}^*(X)$ is a non-empty strict subset of S containing its k worst alternatives according to $\succ^*$, where $1 \leq k < |S|$. We adopt the convention that $W_{\succ^*}(S) = \emptyset$ when S is a singleton.

In the baseline model that we develop, we consider a simple heuristic in which the DM's should not set in any choice problem $S \in \mathcal{P}^*(X)$ is a singleton. To formally define this heuristic, denote the set of $\succ$-maximal elements of any set $S \in \mathcal{P}(X)$ by :

$$\overline{M}(S; \succ) = \{x \in S : x \succ y, \ \ \forall y \in S \setminus \{x\}\}$$

and similarly denote the set of $\succ^*$-minimal elements of any set $S \in \mathcal{P}(X)$ by :

$$\underline{M}(S; \succ^*) = \{x \in S : y \succ^* x, \ \ \forall y \in S \setminus \{x\}\}$$

Of course, all such sets are singletons given that $\succ$ and $\succ^*$ are strict preference rankings. Accordingly, in the baseline model, for any choice problem $S \in \mathcal{P}^*(X)$, the DM's should not set is $\underline{M}(S; \succ^*)$.

Definition 2.2. *A choice function $c : \mathcal{P}(X) \to X$ is an ego preserving heuristic (EPH) if there exists an ordered pair of strict preference rankings $(\succ^*, \succ)$ on X such that for any choice problem $S \in \mathcal{P}(X)$,*

$$c(S) = \overline{M}(S \setminus \underline{M}(S; \succ^*); \succ)$$

As suggested in the Introduction, the EPH is a way for the DM to balance the potentially conflicting demands posed by what she thinks she should do and what she wants to do. The way she does this balancing is by always eliminating from her consideration the worst alternative according to the preferences of her should-self. This done, she chooses the best alternative from the remaining ones according to the preferences of her want-self. It is this balancing that allows her to preserve her ego from the two opposing pulls of her ideals and instinctive wishes (i.e., superego and id, respectively, in the language of Freud) and not be overwhelmed by either guilt or anxiety.

2.1 Behavioral Characterization

Suppose we observe the choices of a DM. When can we conclude by observing these choices that they are a result of this DM choosing according to an EPH. To answer this question, we next provide a behavioral characterization of an EPH. That is, we provide conditions on choices (behavior) that allow us to identify an ordered pair $(\succ^*, \succ)$ of strict preference rankings such that with respect to these rankings, choices have an EPH rationalization.

An EPH is characterized by three behavioral properties that are stated below. Note that in the subsequent analysis we abuse notation by often suppressing set delimiters; e.g., we write $c(xy)$ or $c(xyz)$ instead of $c(\{x, y\})$ or $c(\{x, y, z\})$, respectively. Our first condition, referred to as never chosen (NC), says that if an alternative in a set is not chosen in pairwise choice comparisons with any other alternative from that set, then it is not chosen from the set.

Axiom (NC). *For all $S \in \mathcal{P}(X)$ and $x \in S$:*

$$[c(xy) \neq x, \forall y \in S \setminus \{x\}] \Rightarrow c(S) \neq x$$

Our second condition, referred to as No Binary Cycles (NBC), requires that there are no pairwise choice cycles.

Axiom (NBC). *For all $x_1, \ldots, x_{n+1} \in X$:*

$$[c(x_i x_{i+1}) = x_i, i = 1, \ldots, n] \Rightarrow c(x_1 x_{n+1}) = x_1$$

The standard WARP condition fails to hold in our set-up as seen with the compromise effect, which our model can accommodate. Recall that WARP imposes the following consistency on a DM's choices: for all $S, T \in \mathcal{P}(X)$ and $x, y \in S \cap T$, $x \neq y$, if $c(S) = x$ then $c(T) \neq y$. That is, if x is chosen in the presence of y, then y should never be chosen in the presence of x. If choices satisfy WARP, then they can be rationalized by a single strict preference ranking that can be uniquely elicited from these choices. In Example 1 of the Introduction, denoting the alternatives no cola, small cola and large cola by x, y and z, respectively, we see that $c(xy) = x$ and $c(xyz) = y$. That is, introducing z to the menu $\{x, y\}$ changes the choice from x to y. In the context of our model, the reason for this is because in the set $\{x, y\}$, y is the worst alternative according to the DM's should-self and gets eliminated, resulting in the choice of x. On the other hand, in the set $\{x, y, z\}$, the alternative z gets eliminated, and of the two remaining alternatives x and y, the DM's want-self picks y. From a revealed elicitation perspective, this means that when an outside observer sees x being chosen in the presence of y, she has to be cognizant of the fact that this may be due to one of two reasons—either y is the worst alternative in the set according to the DM's should-self and gets eliminated (possibly the DM's want-self may prefer y to x); or neither x nor y is the eliminated alternative, which would mean that the DM's want-self prefers x to y. We now propose an axiom that weakens the WARP condition to account for the distinction between these two possibilities.

The key to the axiom is to identify, *based purely on choice data*, those alternatives in a set that cannot be the worst one according to the DM's should-self and, therefore, do not get eliminated. For any set $S \in \mathcal{P}^*(X)$ the DM's choices reveal that this *set of non-eliminated alternatives* is given by:

$$\psi(S) = \{x \in S : x = c(S'),\, S' \subseteq S,\, |S'| \geq 2\}$$

That is, these are those alternatives in S that are chosen in some non-singleton subset of it. The reasoning for this inference is the following. If the DM's choices are in accordance with the EPH, then none of the alternatives in $\psi(S)$ can be the worst one in S according to the preferences of her should-self because if it were, it would be the worst such alternative in any subset of S that it is present in. Hence, it cannot be the chosen alternative in any such subset. This, in turn, means that the chosen alternative in S must be strictly preferred to these alternatives according to the preferences of her want-self. Our weakening of WARP, which we refer to as ego preserving WARP (WARP-EP), draws on this insight and provides it with a consistent interpretation. It says that if an alternative x is chosen in a set S and y belongs to the set $\psi(S)$, then y is not chosen in any other set T for which x belongs to $\psi(T)$.

Axiom (WARP-EP). *For all $S, T \in \mathcal{P}(X)$ and $x, y \in S \cap T$, $x \neq y$:*

$$[x = c(S), y \in \psi(S), x \in \psi(T)] \Rightarrow c(T) \neq y.$$

The following result establishes that the three axioms stated above constitute a behavioral foundation of the EPH.

Theorem 2.1. *A choice function is an EPH if and only if it satisfies NC, NBC and WARP-EP.*

Proof: Please refer to Section A.1.1.

In Section A.1.2, we also establish that the three axioms underlying the EPH are independent.

2.2 Identification

We next address the issue of how uniquely the parameters underlying an EPH can be identified. This is an important question as one of the challenges that behavioral choice theory models are often confronted with is the inability to uniquely identify the model parameters from choice data. The next result establishes that the EPH is almost immune to this criticism. In the way of notation, note that for any binary relation B on X and any $S \in \mathcal{P}(X)$, $B|_S$ denotes the restriction of B to the set S.

Proposition 2.1. *If $(\succ^*, \succ)$ and $(\tilde{\succ}^*, \tilde{\succ})$ are both EPH representations of a choice function $c : \mathcal{P}(X) \to X$, then $\succ^* = \tilde{\succ}^*$ and $\succ|_{X \setminus \underline{M}(X;\succ^*)} = \tilde{\succ}|_{X \setminus \underline{M}(X;\tilde{\succ}^*)}$.*

Proof: Please refer to Section A.1.3.

In other words, in an EPH representation, the first preference ranking associated with the DM's should-self is uniquely identified; and the second preference ranking associated with her want-self is identified almost uniquely with only the position of the worst alternative under the first ranking being indeterminate under the second.

3 The Generalized Ego Preserving Heuristic

Under the EPH, in the first stage, the DM eliminates the worst alternative according to the preferences of her should-self. However, it is plausible that in many choice problems, just eliminating the worst alternative is not enough to preserve her ego from being overwhelmed by a sense of guilt. In such situations, ego preservation may force her to eliminate more alternatives. For instance, consider once again the control setting in List (2007) discussed in the Introduction in which the dictator can share the additional \$5 in any way she wishes with the recipient in increments of 50 cents. In this choice problem, it may not be enough for the DM to eliminate just the worst alternative according to her should-self of appropriating the entire \$5 and transferring only 50 cents to the recipient. Doing so may not rid her of the guilt associated with behaving selfishly. She may, accordingly, feel compelled to eliminate a few more of the selfish alternatives to avoid this guilt. With these kind of choice situations in mind, we now propose a generalization of the EPH model in which, in any choice problem, the DM's should not set need not be a singleton, i.e., she may eliminate a non-singleton subset of the worst alternatives according to $\succ^*$.

Definition 3.1. *A choice function $c : \mathcal{P}(X) \to X$ is a generalized ego preserving heuristic (GEPH) if there exists an ordered pair of strict preference rankings $(\succ^*, \succ)$ on X and a should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$, such that for any $S \in \mathcal{P}(X)$:*[9]

$$c(S) = \overline{M}(S \setminus W_{\succ^*}(S); \succ)$$

That is, in any choice problem $S \in \mathcal{P}^*(X)$, the DM first eliminates a set of alternatives $W_{\succ^*}(S)$ that are worst according to $\succ^*$ with the eliminated set no longer required to be a

[9] Recall the convention we adopt that $W_{\succ^*}(S) = \emptyset$ if S is singleton.

singleton like in the case of an EPH. Then from the remaining alternatives, she picks the best alternative according to $\succ$.

Remark 3.1. Before we proceed, a word on the interpretation of the should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ in a GEPH representation is in order. Recall the key behavioral tension in the model—that between the DM's feelings of guilt and anxiety. The should not sets speak to this tension. The larger is the set $W_{\succ^*}(S)$ for a choice problem S, the more prominent are the DM's feelings of guilt relative to anxiety in that problem. That is, the psychological cost of guilt associated with choosing an alternative that the DM believes she should not has a relatively bigger sway on her decision making process than that of anxiety associated with not choosing something that she wants to. That is why, in this situation, she is prepared to eliminate more alternatives that are bad according to her should-self. A smaller $W_{\succ^*}(S)$ set implies the opposite. An implication of this interpretation is that the should not sets can be used to model the choice environment and cognitive context in which the DM is facing choice problems featuring an intrapersonal conflict. Specifically, as discussed in the Introduction, the more ego depleted a DM is, the less pronounced are feelings of guilt and more overwhelming are those of anxiety. Therefore, it is reasonable to hypothesize that the should not sets associated with a highly ego depleted cognitive state will be uniformly smaller across choice problems than those associated with a low one. Consequently, ability to exercise self regulation will be lower in the former case than in the latter. As such, the size of the should not sets may be a meaningful indicator of the DM's state of ego depletion, which is one of the key cognitive measures that determine how choice problems featuring want-should type intrapersonal conflicts play out.

Next we deal with the question of the behavioral characterization of the GEPH model. As it turns out, the model is characterized by the two axioms of NBC and NC.

Theorem 3.1. *A choice function is a GEPH if and only if it satisfies NBC and NC.*

Proof: Please refer to Section A.2.1.

The GEPH model does not impose any structure on the should not set mapping. Because of this the model may lack precision when it comes to identifying its underlying parameters. The preference ranking $\succ^*$ reflecting the DM's should-self is still uniquely identified. However, the preference ranking $\succ$ reflecting her want-self is identified less precisely. Specifically, the extent of revealed preferences w.r.t. this ranking is limited to the following binary relation : for $x, y \in X$, $x \neq y$, $x\hat{P}y$ if there exists $S \in \mathcal{P}(X)$ such that $x = c(S)$ and $y = c(xy)$. Finally, as far as the identification of the should not sets are concerned, define for any $S \in \mathcal{P}^*(X)$ and any $x \in S$ the set of alternatives in this set that x dominates on the should criterion, which are revealed to us by choices over binaries:

$$D_S(x) = \{y \in S : x = c(xy)\}$$

Since the chosen alternative in S, $c(S)$, is not eliminated, whatever the actual should not set $W_{\succ^*}(S)$ is, it must be a subset of $D_S(c(S))$. In the GEPH model, the identification of these sets is limited to this observation. As such, the identification of the should not sets in the model may not be very precise. This is a limitation of the model as one of the important features of the decision making process that an analyst may seek information about is precisely the should not sets for the reasons discussed above.

3.1 Monotonic GEPH

The main reason for the limitation of the GEPH model discussed above is that the should not set mapping under it lacks structure. In order to address this limitation, we now raise the question of what restrictions are plausible for this mapping. One such restriction that may be a reasonably accurate description of a DM's thought process is in the nature of a monotonicity condition which entails that if $S \subseteq T$, then $W_{\succ^*}(S) \subseteq W_{\succ^*}(T)$. For instance, consider once again the control setting in List (2007) in which the dictator can share the additional \$5 with the recipient in increments of 50 cents. Call this choice problem S. Suppose $W_{\succ^*}(S) = \{(\$10, \$5), (\$9.50, \$5.50)\}$, i.e., the two most selfish alternatives of giving \$0 and \$0.50, respectively, to the recipient are eliminated. Now consider a modified choice problem T with the same setting but the dictator is given the option of sharing the \$5 with the recipient in increments of 25 cents. Clearly, $S \subseteq T$ and, in this case, it is hard to imagine why $W_{\succ^*}(S)$ won't be a subset of $W_{\succ^*}(T)$. However, it need not be the case that such a monotonicity condition is satisfied in every instance. For instance, consider now the treatment in List (2007) in which, in addition to the different options of sharing the \$5, the dictator is also given the option of taking \$1 from the recipient. When faced with such a choice problem, call it T', one may encounter DMs who in S give the recipient, say, \$1, but in T' make a zero transfer. In the context of our model, this would imply that the DM ends up eliminating only the $\succ^*$-worst alternative in T', i.e., $W_{\succ^*}(T') = \{(\$11, \$4)\}$. This means that although S is a subset of T', $W_{\succ^*}(S)$ is not a subset of $W_{\succ^*}(T')$. What's the difference between T' and T in relation to S? We believe that the difference between the two situations is that in T all the alternatives that are added to S, i.e., alternatives in $T \setminus S$, are better than the worst alternative in S according to $\succ^*$. On the other hand, this is not the case in T'. The sole alternative in $T' \setminus S$ is worse w.r.t. $\succ^*$ than the worst alternative in S. Introducing such alternatives may adversely affect the DM's interpretation, judgment and agency regarding what shouldn't be done and result in a failure of monotonicity. One reason why this may be the case is tied to ego depletion. Specifically, introducing alternatives that are worse on the should dimension may create further desires and temptations for the DM and deplete her ability to exercise self regulation. As such, she may no longer be able to resist the alternatives she was able to in the smaller set. The monotonicity condition that we are seeking should recognize this important feature associated with exercising self regulation. Indeed, this feature is critical to the menu effects and non-standard behavior that we observe under self regulation. As the following result clarifies, if this feature is ruled out, then our framework reduces to the rational choice benchmark with the DM choosing the $\succ^*$-best alternative in any choice problem.

Proposition 3.1. *If the collection $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$ represents a GEPH choice function c, and $W_{\succ^*}$ is such that for any $S \subseteq T$, $W_{\succ^*}(S) \subseteq W_{\succ^*}(T)$, then c satisfies WARP. Moreover, for any $S \in \mathcal{P}(X)$,*

$$c(S) = \{x \in S : x \succ^* y, \forall y \in S \setminus x\}$$

Proof: Please refer to Section A.2.2

We now formally state the appropriate monotonicity condition that follows from our discussion above. To do so, in the way of notation, let $\underline{z}_S$ denote the $\succ^*$-worst alternative in S.

Definition 3.2. *A should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ satisfies mononotonicity if for any $S, T \in \mathcal{P}^*(X)$,*

$$[S \subseteq T \text{ s.t. } x \succ^* \underline{z}_S, \text{ for all } x \in T \setminus S] \Rightarrow W_{\succ^*}(S) \subseteq W_{\succ^*}(T)$$

With this notion of monotonicity, we can introduce a more precise version of the GEPH model.

Definition 3.3. *A GEPH choice function $c : \mathcal{P}(X) \to X$ defined by the collection, $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$ is monotonic if $W_{\succ^*}$ satisfies monotonicity.*

It should be straightforward to verify that an EPH is a special case of a monotonic GEPH.

3.2 Behavioral Characterization of Monotonic GEPH

We next address the question about the behavioral characterization of a monotonic GEPH. The key step, like in the characterization of an EPH, is to identify based purely on choice data, those alternatives in a set that an outside observer can be sure are not eliminated based on the preferences of the DM's should-self. What makes this identification a bit more involved in the case of a monotonic GEPH is the fact that, unlike for an EPH, we do not know the number of alternatives in a choice problem that are eliminated. Even then, drawing on the assumption that the should not set mapping satisfies the monotonicity condition, we can identify a set of alternatives in any choice problem that, based purely on information about the DM's choices, we can guarantee are not eliminated.

To begin with, recall our earlier observation that since the chosen alternative in S, $c(S)$, is not eliminated, we know that whatever the true should not set $W_{\succ^*}(S)$ of the DM in this choice problem is, it must be a subset of $D_S(c(S))$. Now, for any such S, consider the following collection of its supersets:

$$\mathcal{T}_S = \{T \in \mathcal{P}(X) : S \subseteq T \text{ and } x \in T \setminus S \Rightarrow \exists y \in S \text{ s.t. } c(xy) = x\}$$

Note that $S \in \mathcal{T}_S$. Further, the sets in $\mathcal{T}_S$ are precisely those supersets of S to which the content of the monotonicity condition applies. As such drawing on it, we know that for any such $T \in \mathcal{T}_S$, the set $W_{\succ^*}(S)$ is a subset of $W_{\succ^*}(T)$, which in turn is a subset of $D_T(c(T))$. Putting all of this together, we can infer that, for a DM who chooses according to the logic of a monotonic GEPH, $W_{\succ^*}(S)$ must be contained in the set

$$E(S) = \bigcap_{T \in \mathcal{T}_S} D_T(c(T))$$

Using this choice based inference, we can accordingly elicit information about a set of alternatives in S that are guaranteed to not get eliminated from it based on the preferences of the DM's should self. This set is given by:

$$\tilde{\psi}(S) = S \setminus E(S)$$

It is worth reiterating the distinction between what the true set of eliminated alternatives—the DM's should not set $W_{\succ^*}(S)$—in a choice problem S is and the information about this set that an outside observer can elicit from choices. Specifically, $W_{\succ^*}(S) \subseteq E(S)$; and, accordingly, $\tilde{\psi}(S) = S \setminus E(S) \subseteq S \setminus W_{\succ^*}(S)$.

Based on our earlier characterization of an EPH, it may appear that the following axiom, which we refer to as WARP-GEP, is the natural weakening of WARP when it comes to characterizing a monotonic GEPH.

Axiom (WARP-GEP). *For all $S, T \in \mathcal{P}(X)$ and $x, y \in S \cap T$, $x \neq y$:*

$$[x = c(S), y \in \tilde{\psi}(S), x \in \tilde{\psi}(T)] \Rightarrow c(T) \neq y.$$

However, this turns out to be not the case. To characterize a monotonic GEPH, we need a stronger version of this axiom, specifically, a strong axiom of revealed preferences (SARP) type condition specialized to the current context. The axiom, which we refer to as SARP-GEP, states the following.

Axiom (SARP-GEP). *For all $S_1, \ldots, S_n \in \mathcal{P}(X)$ and distinct $x_1, \ldots, x_n \in X$:*

$$\left[x_i = c(S_i), x_{i+1} \in \tilde{\psi}(S_i), i = 1, \ldots, n-1, \text{ and } x_1 \in \tilde{\psi}(S_n)\right] \Rightarrow c(S_n) \neq x_n$$

Before we get to our characterization result, it is instructive to propose an equivalent way of expressing the SARP-GEP axiom. To do so, we define a binary relation P on X as follows: for any $x, y \in X$, $x \neq y$, xPy if there exists $S \in \mathcal{P}(X)$ such that $x = c(S)$ and $y \in \tilde{\psi}(S)$. The following result should be self-evident.

Proposition 3.2. *P is acyclic if and only if c satisfies SARP-GEP.*

We can now introduce our characterization result which establishes that SARP-GEP along with NBC constitutes a behavioral foundation of the monotonic GEPH model.

Theorem 3.2. *A choice function is a monotonic GEPH if and only if it satisfies NBC and SARP-GEP.*

Proof: Please refer to Section A.2.3.

We now present a couple of examples which show how, for a given choice data set, we can verify whether SARP-GEP is satisfied. Note that the choice functions in both the examples below satisfy NBC. Therefore, whether the choice functions are a monotonic GEPH or not boils down to checking whether they satisfy SARP-GEP.

Example 3.1. *(Not a monotonic GEPH).* Let $X = \{x, y, z, w, v\}$ and consider the choice function specified in the tables below. The first row of these tables list all the choice problems; the second, the choices made in these choice problems. The rows after that specify for each choice problem S, the sets $D_S(c(S))$, $E(S)$ and $\tilde{\psi}(S)$ as described above.

S	xy	xz	xw	xv	yz	yw	yv	zw	zv	wv	xyz	xyw	xyv
$c(S)$	x	x	x	x	y	y	y	z	z	w	x	x	x
$D_S(c(S))$	y	z	w	v	z	w	v	w	v	v	yz	yw	yv
$E(S)$	y	z	w	v	z	w	v	w	v	v	yz	w	v
$\tilde{\psi}(S)$	x	x	x	x	y	y	y	z	z	w	x	xy	xy

S	xzw	xzv	xwv	yzw	yzv	ywv	zwv	$xyzw$	$xyzv$	$xywv$	$xzwv$	$yzwv$	$xyzwv$
$c(S)$	x	x	x	y	y	y	z	y	y	y	z	z	z
$D_S(c(S))$	zw	zv	wv	zw	zv	wv	wv	zw	zv	wv	wv	wv	wv
$E(S)$	zw	v	wv	zw	v	wv	wv	zw	v	wv	wv	wv	wv
$\tilde{\psi}(S)$	x	xz	x	y	yz	y	z	xy	xyz	xy	xz	yz	xyz

The set $D_S(c(S))$ for any choice problem S is straightforward to determine by looking at the pairwise choice problems involving $c(S)$ and other alternatives in that set. Next, for any choice problem S, we determine the set $\mathcal{T}_S$. For instance, for the set $\{x, y, v\}$, $\mathcal{T}_{xyv} = \{\{xyv\}, \{xyzv\}, \{xywv\}, \{xyzwv\}\}$. We can then determine the set $E(S) = \bigcap_{T \in \mathcal{T}_S} D_T(c(T))$. For instance, for the set $\{x, y, v\}$, $E(xyv) = \{y, v\} \cap \{z, v\} \cap \{w, v\} \cap \{w, v\} = \{v\}$. Finally, we can determine the sets $\tilde{\psi}(S) = S \backslash E(S)$ for all $S \in \mathcal{P}(X)$. We can now verify that this choice function violates SARP-GEP and hence is not a monotonic GEPH. To do so, we elicit the binary relation P defined above and show it is cyclic, which then based on Proposition 3.2 establishes that c violates SARP-GEP. Consider the set $\{xyw\}$. For this set, $\tilde{\psi}(xyw) = \{x, y\}$ and $c(xyw) = x$, implying xPy. Now, in set $\{xyzw\}$, $\tilde{\psi}(xyzw) = \{x, y\}$ and $c(xyzw) = y$. This gives us yPx. Hence, P has a cycle.

Example 3.2. *(Monotonic GEPH).* Let $X = \{x, y, z, w, v\}$ and consider the choice function c on X specified in the table below.

S	xy	xz	xw	xv	yz	yw	yv	zw	zv	wv	xyz	xyw	xyv
$c(S)$	x	x	x	x	y	y	y	z	z	w	x	x	x
$D_S(c(S))$	y	z	w	v	z	w	v	w	v	v	yz	yw	yv
$E(S)$	y	z	w	v	z	w	v	w	v	v	yz	yw	v
$\tilde{\psi}(S)$	x	x	x	x	y	y	y	z	z	w	x	x	xy

S	xzw	xzv	xwv	yzw	yzv	ywv	zwv	$xyzw$	$xyzv$	$xywv$	$xzwv$	$yzwv$	$xyzwv$
$c(S)$	x	x	w	y	y	w	w	x	x	w	w	w	w
$D_S(c(S))$	zw	zv	v	zw	zv	v	v	yzw	yzv	v	v	v	v
$E(S)$	zw	v	v	zw	v	v	v	yzw	v	v	v	v	v
$\tilde{\psi}(S)$	x	xz	xw	y	yz	yw	zw	x	xyz	xyw	xzw	yzw	$xyzw$

We verify that c satisfies SARP-GEP and hence it is a monotonic GEPH. The sets $D_S(c(S)), E(S)$ and $\tilde{\psi}(S)$ for any choice problem S are determined as discussed in the previous example. We can then infer P as: $wPx, wPy, wPz, xPy, xPz, yPz$. Since P is acyclic, the choice function satisfies SARP-GEP and hence is a monotonic GEPH.

Remark 3.2. The characterization of an EPH and a GEPH appeals to the axiom of NC, but that of a monotonic GEPH does not. The following result clarifies the reason for this.

Proposition 3.3. *If a choice function satisfies SARP-GEP, then it satisfies NC.*

Proof: Please refer to Section A.2.4.

3.3 Identification of a Monotonic GEPH

We next address the question of how uniquely the parameters underlying a monotonic GEPH can be identified. Like in the case of an EPH, the preferences, $\succ^*$, of the DM's should-self can be uniquely identified. However, the identification of the preferences, $\succ$, of her want-self may be less precise than under an EPH. To explain this, we first formally define the notion of revealed preferences of the model w.r.t. the preferences of the DM's want-self.

Definition 3.4. *Let c be a monotonic GEPH. We define the revealed preference relation Q^* on X by: for any $x, y \in X$, $x \neq y$, xQ^*y if for any monotonic GEPH representation $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$ of c, we have $x \succ y$.*

In other words, x is revealed to be preferred to y, if in any monotonic GEPH representation of the DM's choices, it is the case that according to the preferences of the DM's want-self,

x ranks over y. Now, recall the binary relation P that we defined in the last section: for any $x, y \in X$, $x \neq y$, xPy if there exists $S \in \mathcal{P}(X)$ such that $x = c(S)$ and $y \in \tilde{\psi}(S)$. It is fairly straightforward to verify that if the DM's choices are a monotonic GEPH, then $P \subseteq Q^*$. This is because $y \in \tilde{\psi}(S)$ implies there exists $T \in \mathcal{T}_S$ s.t., $y \notin D_T(c(T))$. Now we know that *in any* monotonic GEPH representation, $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$, of c, the should not sets of S and T must satisfy the following inclusion: $W_{\succ^*}(S) \subseteq W_{\succ^*}(T)$. Further, $W_{\succ^*}(T) \subseteq D_T(c(T))$. Putting all this together gives us that $y \notin W_{\succ^*}(S)$. Hence, $x = c(S)$ and $y \in S \setminus W_{\succ^*}(S)$ implies $x \succ y$.

Next, note the following observation. If xPy and yPz, then in any monotonic GEPH representation with the preference ranking $\succ$ representing the DM's want-self, we have $x \succ y$ and $y \succ z$. But, $\succ$ is transitive which implies that $x \succ z$ must be true in any such representation as well. Therefore, xQ^*z even though it may not be the case that xPz is directly elicited from the choices. In other words, denoting by P^* the transitive closure of P, it follows that $P^* \subseteq Q^*$. The following result establishes that the inclusion also goes in the other direction. That is, the binary relations P^* captures the full extent of the revealed preferences w.r.t. the preferences of the DM's want-self in the monotonic GEPH model.

Proposition 3.4. *Let c be a monotonic GEPH. Then $Q^* = P^*$.*

Proof: Please refer to Section A.2.5.

Finally, as far as the identification of the should not sets are concerned, these too may not be identified exactly. But we can provide bounds on these sets that for choice problems with a rich enough set of outcomes can be quite tight.

Proposition 3.5. *Let c be a monotonic GEPH. Then for any any monotonic GEPH representation $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$ of c and any $S \in \mathcal{P}^*(X)$, $W_{\succ^*}(S) \subseteq \bigcap_{T \in \mathcal{T}_S} D_T(c(T))$.*

This conclusion has already been established in the earlier sub-section in which we provided the behavioral characterization of a monotonic GEPH.

Example 3.2 (continued)**.** To better understand the extent of identification under a monotonic GEPH and to contrast this with a GEPH, refer back to Example 3.2. In that example, we identified the binary relation $P = \{(w, x), (w, y), (w, z), (x, y), (x, z), (y, z)\}$. Its transitive closure P^* is the same as P. So, the revealed preferences in this case is the ranking $wQ^*xQ^*yQ^*z$. In other words, we can almost uniquely identify the preferences $\succ$ of the want-self with only the position of the worst alternative under the $\succ^*$ ranking being indeterminate, making identification of the two preference relations in this example identical to what we get under an EPH. Contrast this with the inferences that an outside observer would have made if she had analyzed this choice data as simply a GEPH without the additional restriction of monotonicity on the should not set mapping. In that case, the identification of the preferences of the want-self would have been much less precise. The reason for this is because in a GEPH the should not sets are identified much less precisely—for any choice problem S, all we can say is that the should not set of S is contained in $D_S(c(S))$. On the other hand, for a monotonic GEPH we can ascertain that this set is contained in $\bigcap_{T\in\mathcal{T}_S} D_T(c(T))$. For instance, for the choice problem $S = \{x, y, z, v\}$ in the example, all that we can determine about the should not set for the case of a GEPH is that it is contained in $\{y, z, v\}$, whereas for a monotonic GEPH it is exactly identified as $\{v\}$. Consequently, the revealed preferences w.r.t. the DM's want-self that can be ascertained by analyzing choices in this example within a GEPH framework is a strict subset of that under a monotonic GEPH and is given by the binary relation $\hat{P} = \{(w, x), (w, y), (w, z)\}$:

4 Comparisons with other models

We now compare the EPH/GEPH models with other related behavioral choice theory models in the literature as well as some models featuring costly self control in the Gul and Pesendorfer (2001) tradition. This discussion will help us understand the similarities and differences of our models with these ones, especially when it comes to the question of accommodating non-standard choice data. As mentioned earlier, WARP is a necessary and sufficient condition for observed behavior to be rationalized by a single strict preference ranking. Manzini and Mariotti (2007) provide a very useful typology of the ways WARP can be violated. Consider the following condition referred to as Always Chosen (AC).

Definition 4.1. *A choice function $c : \mathcal{P}(X) \to X$ satisfies AC if for all $S \in \mathcal{P}(X)$ and $x \in S$:*

$$[c(xy) = x, \forall y \in S \setminus \{x\}] \Rightarrow c(S) = x$$

Theorem 4.1 (Manzini and Mariotti (2007))**.** *A choice function satisfies WARP iff it satisfies AC and NBC.*

In other words, all violations of WARP can be categorized as either violations of AC or violations of NBC (or both). As our axiomatization clarifies, both the EPH and GEPH models satisfy NBC. Therefore, in terms of this typology, the only type of violation of classical rationality that they can accommodate is that of AC. Beyond this formal typology, it is worth noting that when it comes to observed violations of rational choice theory seen in experiments and field studies, the prominent classes of violations that have been highlighted are precise those of AC and NBC along with NC. These three form the leading exemplars of non-standard choice data. Different models can accommodate different combinations of these violations. As we have seen, our models accommodate violations of AC but NC and NBC are satisfied by them.

The above observation helps immediately clarify why both the EPH and GEPH models are distinct from the influential Rational Shortlist Method (RSM) or, more generally, the sequentially rationalizable model that Manzini and Mariotti (2007) introduce. A choice function c is an RSM if there exists an ordered pair of asymmetric binary relations (P_1, P_2) on X such that for any choice problem S, the choice is given by:[10]

$$c(S) = \max(\max(S; P_1); P_2)$$

It is fairly straightforward to establish that an RSM satisfies AC and the only violations of classical rationality it can accommodate are those of NBC. Therefore, the behavioral implications of the EPH and GEPH when it comes to accommodating non-standard data can be thought of as orthogonal to that of RSM. However, one feature that we have in common is that both RSM and our models satisfy NC.

RSM is characterized by two axioms: Weak WARP and Expansion. Weak WARP is a key axiom in the behavioral choice theory literature and therefore we formally state it.

Definition 4.2. *A choice function $c : \mathcal{P}(X) \to X$ satisfies Weak WARP if for all $S, T \in \mathcal{P}(X)$ and $x, y \in X$:*

$$\big[\{x, y\} \subseteq S \subseteq T, x = c(xy) = c(T)\big] \Rightarrow y \neq c(S)$$

Several important models in the literature are characterized by it, e.g., the Rationalization model of Cherepanov, Feddersen, and Sandroni (2013) and the Categorize then Choose (CTC) model of Manzini and Mariotti (2012). Therefore, a natural question is whether the EPH/GEPH models satisfy this condition.

Proposition 4.1. *An EPH satisfies Weak WARP but a monotonic GEPH may not.*

Proof: Please refer to Section A.3.1.

Therefore, we can conclude that choice data that fits the EPH model also fit the Rationalization and CTC models, but this need not be true of the monotonic GEPH model. In terms of accommodating non-standard choices, both the Rationalization and CTC model can accommodate violations of NC, AC and NBC. This last observation should also clarify the fact that there exists choice functions that satisfy Weak WARP but are not monotonic GEPH.

There are many models which further explore two-stage sequential choice procedures. Au and Kawai (2011) and Horan (2016) consider the RSM model but with transitive rationales.

[10] For this model, for any asymmetric binary relation P, we define: $\max(S; P) = \{x \in S : \nexists y \in S \text{ s.t. } yPx\}$

Both these models, like RSM, accommodate violations of NBC but not of AC. Another model in this class of models which resonates with our work is the Two-Stage Chooser (TSC) model of Bajraj and Ülkü (2015). They model a DM who sequentially uses two rationales, which are both strict preference rankings, to make choices. In the first stage, the DM shortlists the top two alternatives according to the first ranking and then in the second stage, chooses between them according to the second ranking. Like in our models, the TSC model can accommodate violations of AC but not that of NBC and NC. Given this similarity, it naturally raises the question as to whether any of these models is a special case of the other. We show in Section A.3.2 that this is not the case.

Another class of behavioral choice theory models builds on the observation that in many choice problems a DM may not pay attention to all the available alternatives owing to cognitive limitations or unawareness. The DM thus forms a consideration set in the first stage, i.e., a set of alternatives which receive her attention in a choice problem. She then maximizes in that set like a standard rational agent. A prominent example of this class of models is the Choice with Limited Attention (CLA) model of Masatlioglu, Nakajima, and Ozbay (2012). In the CLA model, the consideration set mapping is required to be an attention filter i.e., if an alternative is not considered by the DM, then her consideration set does not change when this alternative becomes unavailable. Formally, a consideration set mapping Γ is an attention filter if for any S, $\Gamma(S) = \Gamma(S \setminus x)$ whenever $x \notin \Gamma(S)$. Once the consideration set is identified in the first stage, the DM chooses the best alternative from it based on a strict preference ranking. This choice procedure is characterized by the WARP(LA) axiom.

Definition 4.3. *WARP(LA): For any $S \in \mathcal{P}(X)$, there exists $x^* \in S$ such that, for any T including x^*,*

$$\text{if } c(T) \in S \text{ and } c(T) \neq c(T \setminus x^*), \text{ then } c(T) = x^*.$$

The following result establishes the relationship between the EPH/GEPH models and the CLA model.

Proposition 4.2. *An EPH satisfies WARP(LA) but a monotonic GEPH may not.*

Proof: Please refer to Section A.3.3.

Therefore, an EPH can be represented as a CLA but a monotonic GEPH may not be. Another model that relates to the CLA is the Overwhelming Choice (OC) model of Lleras et al. (2017) that studies how the phenomenon of choice overload may cause DMs to consider only a strict subset of the available alternatives. The consideration set mapping that they employ to convey this idea is that of a competition filter, which says that if an alternative is considered in a set, then it must also be considered in any of its subsets where it is present. Formally, Γ is a competition filter if whenever $x \in S \subseteq T$ and $x \in \Gamma(T)$, then $x \in \Gamma(S)$. In terms of characterization, this model can be rationalized by Weak WARP and hence an EPH can be represented as an OC but a monotonic GEPH may not be. We should point out here that the consideration set mapping under an EPH, i.e., for any set S, the set $S \setminus \underline{M}(S; \succ^*)$, is neither an attention filter nor a competition filter. Therefore, when viewed from a cognitive decision making point of view the EPH is quite distinct from either of these models.

Another strand of literature which relates to our paper is that of threshold models. In these models a DM forms her consideration sets by considering all those alternatives in a choice problem which meet certain criteria or are above a certain threshold. She then maximizes on this set like in the models discussed above. The two-stage threshold model of Manzini, Mariotti, and Tyson (2013) is built around such an idea. In their model, a DM has a primary criterion function $f : X \to \mathbb{R}$ on the set of alternatives X and a threshold function θ on the set of menus $\mathcal{P}(X)$. An alternative x in a choice problem S gets consideration if $f(x) \geq \theta(S)$. The maximization in the second stage is according to a secondary criterion function g. One major distinction between their model and ours is that they use a "cardinal" approach as opposed to our ordinal one. Secondly, the behavioral implications of the two models are different—choices in their model need not satisfy NBC like in ours.

Another paper of interest in this line of work is Kimya (2018). In the Choice through Attribute Filters (CAF) model developed in this paper, alternatives have observable attributes and the DM forms consideration sets (attribute filters) based on these attributes. Specifically, a multi-criteria choice data set specifies (i) a finite set of alternatives $X \subseteq \mathbb{R}^k_{++}$, where each alternative $x = (x_1, \ldots, x_k) \in X$ has k attributes and (ii) a choice function on the set of alternatives. A consideration set mapping Γ is an attribute filter if for each $S \in \mathcal{P}(X)$, there exists a threshold $t^S \in \mathbb{R}^k$ such that $\Gamma(S) = \{z \in S : z > t^S\}$.[11] Further in an attribute filter, the thresholds must not "overreact" when an alternative is added to a choice problem. Specifically, when an alternative x is added to S such that x is above the threshold of attribute i, i.e., $x_i > t_i^S$, the threshold on attribute i in $S \cup x$ can only increase, but never to the point that it leads the DM to eliminate some alternative y with $y_i > x_i$. Similarly, when an alternative x that is below the threshold of attribute i is added, the threshold can only decrease, but never to the point that it leads the DM to consider some alternative y with $y_i < x_i$. A choice function c is a CAF if there exists a strict preference ranking $\succ$ over X that is monotonic (w.r.t. attributes) and an attribute filter Γ such that $c(S)$ is the $\succ$-best element in $\Gamma(S)$ for every $S \in \mathcal{P}(X)$. A CAF can accommodate violations of NC.[12] Therefore, CAF is not a special case of our models. We show in Section A.3.4 that EPH/GEPH are also not special cases of CAF.

[11] Note that $z > t^S$ if $z_i > t_i^S$ for each $i = 1, \ldots, k$.

[12] For an example of this, refer to the choices relating to the compromise effect mentioned in Section II.C of Kimya (2018).

Finally, we compare our model with the strand of literature which builds on the Gul and Pesendorfer (2001) framework. Specifically, we focus on models within this framework that accommodate violations of WARP when it comes to choices of DMs who face temptation when choosing from a menu but display an ability to exercise self control albeit at a cost. As discussed in the Introduction, Noor and Takeoka (2010), Noor and Takeoka (2015), Liang, Grant, and Hsieh (2019) and Masatlioglu, Nakajima, and Ozdenoren (2020) are a few papers which have modelled this.[13]

In Noor and Takeoka (2010), the cost of exerting self-control is convex implying that the marginal cost of exerting self control is increasing. Formally, denoting the DM's commitment and temptation rankings by the utility functions u and v, respectively, the DM's choice in any menu S is determined by:

$$c(S) = \underset{x \in S}{\operatorname{argmax}} \left\{ u(x) - \varphi\big(\max_{y \in S} v(y) - v(x) \big) \right\},$$

where φ is a convex function whose argument, $\max_{y \in S} v(y) - v(x)$, can be thought of as the temptation that needs to be overcome through exerting self control when choosing x in a menu S whose most tempting alternative has temptation level $\max_{y \in S} v(y)$.[14] Noor and Takeoka (2015) introduce self-control costs which are menu-dependent. Formally,

$$\begin{aligned} c(S) &= \underset{x \in S}{\operatorname{argmax}} \left\{ u(x) - \psi\big(\max_{y \in S} v(y) \big) \big(\max_{y \in S} v(y) - v(x) \big) \right\} \\ &= \underset{x \in S}{\operatorname{argmax}} \left\{ u(x) + \psi\big(\max_{y \in S} v(y) \big) v(x) \right\}, \end{aligned}$$

where $\psi(.) \geq 0$. That is, the function $\psi(.)$ scales up or down the self control cost associated with a menu depending on the most tempting alternative available in it. Both these models help illustrate a similarity that exists between our model and this framework when it comes to menu effects. To see this, consider adding to a menu S an alternative $\hat{w}$ that is more tempting than any of the alternatives in S—in the language of our model, $w \succ^* \hat{w}$, for all $w \in S$. In our model, this means that $W_{\succ^*}(S)$ need not be a subset of $W_{\succ^*}(S \cup \hat{w})$ and this is the channel through which menu effects show up in the model. A similar observation regarding menu effects is true in these models driven by non-linearities in self control costs. We illustrate this using Example 1 of the Introduction with the three alternatives of no cola (x), small cola (y) and large cola (z), and the convex self control cost model of Noor

[13] It is worth pointing out that these models do have some differences when it comes to their primitives. Noor and Takeoka (2010) and Liang, Grant, and Hsieh (2019) consider only preference over menus as their primitive so that choices from menus is not directly modelled behaviorally but rather implied by the representation of preferences over menus (under the assumption that the DM is sophisticated). Noor and Takeoka (2015) additionally model choices from menus explicitly in terms of its behavioral foundations. The same is the case with Masatlioglu, Nakajima, and Ozdenoren (2020) where the behavioral focus is most prominently on choices from menus.

[14] Noor and Takeoka (2010) also discuss a more general version of this model in which the cost function takes a general form, $\tilde{\varphi}\big(v(x), \max_{y \in S} v(y)\big)$.

and Takeoka (2010). Say, $u(x) = 13, u(y) = 7, u(z) = 2$ and $v(x) = 1, v(y) = 3, v(z) = 5$. Further, $\phi(r) = r^2$. Then in set $S = \{x, y\}$,

$$u(x) - \varphi\big(\max_{w\in S} v(w) - v(x)\big) = 13 - (3-1)^2 = 9 > 7 = 7 - (3-3)^2 = u(y) - \varphi\big(\max_{w\in S} v(w) - v(y)\big)$$

Hence, $c(S) = x$. But, in $S' = S \cup z$, $c(S') = y$ as shown below:

$$\begin{aligned} u(y) - \varphi\big(\max_{w\in S'} v(w) - v(y)\big) &= 7 - (5-3)^2 = 3 > 2 = u(z) - \varphi\big(\max_{w\in S'} v(w) - v(z)\big) \\ &> -3 = 13 - (5-1)^2 = u(x) - \varphi\big(\max_{w\in S'} v(w) - v(x)\big) \end{aligned}$$

Both Masatlioglu, Nakajima, and Ozdenoren (2020) and Liang, Grant, and Hsieh (2019) model a DM who has a limited stock of willpower which determines the extent of self-control she can exert in a menu in terms of the alternatives that are psychologically feasible for her to choose. In Masatlioglu, Nakajima, and Ozdenoren (2020), choice from a menu is determined by:

$$c(S) = \operatorname*{argmax}_{x\in S} u(x) \text{ subject to } \max_{y\in S} v(y) - v(x) \leq w,$$

where $w \geq 0$ measures the DM's stock of will power.[15] On the other hand, choices from menus in Liang, Grant, and Hsieh (2019) are like in Gul and Pesendorfer (2001) but with the willpower constraint. Specifically,

$$c(S) = \operatorname*{argmax}_{x\in S}\{u(x) + v(x)\} \text{ subject to } \max_{y\in S} v(y) - v(x) \leq w$$

As should be evident, the mental process that guides the process of self regulation in our model is quite distinct from these models. For instance, consider the model of Masatlioglu, Nakajima, and Ozdenoren (2020). In terms of the language of our model, their choice procedure, in contrast to ours, can be interpreted as involving an elimination/consideration of alternatives in the first stage based on the preferences of the want self and choice in the second stage based on the preferences of the should self. The distinction between the models show up in terms of behavior as well. Whereas in our model choices satisfy NBC, in all four of these models it can be violated.[16] This observation makes for an interesting

[15] Masatlioglu, Nakajima, and Ozdenoren (2020) also discuss a more general version of this model in which the will power stock varies with the chosen alternative.

[16] E.g., to see why the convex self control model of Noor and Takeoka (2010) may violate NBC, refer back to the cola example with the same values for the u and v functions used above. Then, in the choice problem $\{y, z\}$, "utilities" from y and z are, respectively, 3 and 2; hence $c(yz) = y$. In the choice problem $\{x, z\}$, utilities from x and z are, respectively, -3 and 2; hence, $c(xz) = z$. We have already shown above that $c(xy) = x$. Together these choices violate NBC. In the limited willpower model of Masatlioglu, Nakajima, and Ozdenoren (2020) with, say, a willpower stock of $w = 3$ and the same values of u and v, in the choice problem $\{x, y\}$, since the self control cost of choosing x is $v(y) - v(x) = 3 - 1 = 2 < 3$, both x and y are feasible to choose and, accordingly, $c(xy) = x$. Similarly both y and z are feasible in choice problem $\{y, z\}$ and, accordingly, $c(yz) = y$. But, in choice problem $\{x, z\}$, since $v(z) - v(x) = 4 > 3$, x is not feasible. Hence, $c(xz) = z$.

testable distinction between the two approaches to self regulation. In choice problems like the cola choice one above, both our model and these models would tend to predict that between having no cola and a small cola, the DM will choose no cola; and between a small cola and a large cola, she will choose a small cola. However, when it comes to the choice between no cola and a large cola, whereas out model predicts that she will be able to exercise self regulation and have no cola, these models will often predict that she will have the large cola. Finally, note that a similarity of these models with ours is that all four satisfy NC.

5 Further comments on the compromise effect

As noted above, the compromise effect refers to the often observed behavioral pattern whereby DMs seek to avoid making "extreme" choices and opt for more moderate ones that work as a compromise against extremes. As a starting point, note that the very structure of our model induces a compromise effect. This is because in the type of problems featuring intrapersonal conflicts that we are interested in, the preferences of the want and should selves naturally go in opposite directions. That is why, the process of eliminating the worst alternatives according to the preferences of the should-self that our model features, not only pushes choices away from extreme alternatives according to these preferences but at the same time, given that preferences of the want-self works in the opposite direction, implies that choices are middling ones according to the preferences of the want-self as well.

Given that the compromise effect is a widely documented phenomenon, a large literature in multiple disciplines has sought to identify characteristics of those choice situations in which the compromise effect is stronger and ones in which it is relatively weaker or non-existent. Here we want to consider one such type of situation and use our theory to throw further light on it. The particular situation we have in mind pertains to the effect of ego depletion on the compromise effect. Pocheptsova et al. (2009) report that for more ego depleted experimental subjects, the strength of the compromise effect is systematically lower.

There are two avenues by which such an effect can show up in our model. The first avenue takes us back to Remark 3.1 about the interpretation of the should not set mapping. We mentioned there that a DM's state of ego depletion has an impact on how the underlying guilt-anxiety tradeoff plays out. Specifically, higher levels of ego depletion implies that dealing with anxiety becomes the more salient emotion for the DM and, as such, the size of the should not sets may uniformly shrink in such a state. As we show in the example below this may, amongst other things, reduce the strength of the compromise effect.

Example 5.1. A DM passes by a bakery on a Sunday evening and decides to pick a dessert for herself. The menu has the following options to choose from (calorie intake associated with each dessert is mentioned alongside): fruit yogurt - 90 calories (x), macaroons - 140 calories (y), apple pie - 240 calories (z), cheese cake - 500 calories (u), and Banoffee Nutella waffle - 1000 calories (v). She is conscious about her calorie intake and the preferences of her should-self is given by $x \succ^* y \succ^* z \succ^* u \succ^* v$. Assume that the preferences of her want-self go in the opposite direction: $v \succ u \succ z \succ y \succ x$. Our DM has had a relaxed Sunday and is not ego depleted. After a bit of deliberation, she is able to eliminate without any difficulty the three worst alternatives according to her should-self and settles for the macaroons. Now consider the same DM visiting the bakery on a Thursday evening after a long day at work with a still longer work day on Friday coming up. She is exhausted to put in the effort required to remind herself about the desserts she should not have. The effect of this may show up when she is trying to resist the more tempting desserts. In this ego depleted state, her feelings of anxiety associated with avoiding temptation may far outstrip that of guilt associated with succumbing to it. In this case, she is able to eliminate just the worst alternative according to her should-self and ends up making the more extreme choice of the cheese cake. It is this connection between ego depletion and the size of the should not set that may, in turn, reduce the strength of the compromise effect when our DM is in an ego depleted state.

There is another more subtle avenue through which the strength of the compromise effect may be weakened in our model due to ego depletion, especially in larger menus. This avenue has to do with the nature of alternatives that are introduced as we expand a menu. Consider two such menus S and T, where $S \subsetneq T$. The key consideration in this expansion is the $\succ^*$ standing of the expanded alternatives in $T \setminus S$. Denoting by $\underline{z}_S$ the $\succ^*$-worst alternative in S, if $x \succ^* \underline{z}_S$ for all $x \in T \setminus S$, then we know from the monotonicity condition that $W_{\succ^*}(S) \subseteq W_{\succ^*}(T)$. In this case, the strength of the compromise effect is not diluted as the menu grows as all the alternatives that were eliminated in S are still being eliminated in T. Accordingly, all the alternatives in $T \setminus S$ that are $\succ^*$-worse than the $\succ^*$-best alternative in $W_{\succ^*}(S)$ also are eliminated. However, consider the case when some or all of the alternatives in $T \setminus S$ happen to be $\succ^*$-worse than $\underline{z}_S$. When confronted with such potentially tempting alternatives, the very idea of resisting these may be a source of ego depletion. The cognitive load coming from this direction may be further accentuated by that resulting from having to choose from a larger menu. Confronted with these twin loads, there is no guarantee that our DM is still able to eliminate in T all the alternatives that she was eliminating in S. If she succumbs and her should not set shrinks in T, then this may be another avenue through which the strength of the compromise effect may weaken in our model as a result of ego depletion.

Example 5.1 (continued)**.** Suppose when the DM is in the bakery on Sunday evening, just as she is about to make her choice, the server at the bakery tells her that they have two more delicious options available: brownie sundae cheesecake - 1300 calories and chocolate chip cookie sundae - 1600 calories. These two options are truly decadent and resisting them involves a significant level of ego depletion for the DM. With much difficulty she is able to resist them but nothing further. She ends up ordering the Banoffee Nuetella waffle.

The above discussion also alludes to another question that has received some attention in the literature, e.g., Yoo, Park, and Kim (2018): Is there a weakening of the compromise effect as the size of menus grow? The analysis above suggests that in our model the answer to this question depends on the kind of alternatives that are entering the menu as it is expanded. If these alternatives are of the type that worsen the menu on the should dimension in the sense of the antecedent of the monotonicity condition being violated, then the ego depletion involved in resisting these alternatives may cause the should not set to shrink and make the DM choose more extreme alternatives with a concomitant reduction in the strength of the compromise effect. On the other hand, if the expanded alternatives are such that the antecedent of the monotonicity condition holds, then the compromise effect is not weakened as the menu is expanded.

6 Welfare

We conclude with a few comments about welfare. As is well known, welfare analysis with behavioral DMs is not as clear cut as with rational decision makers. In this regard, there have been arguments made in the literature for both a model free approach [e.g., Bernheim and Rangel (2009)], as well as a model based approach [e.g., Masatlioglu, Nakajima, and Ozbay (2012)] to doing behavioral welfare analysis. The key distinction between the two approaches is that the latter takes a stand on the particular choice procedure that DMs in question employ to arrive at their choices whereas the former does not. Masatlioglu, Nakajima, and Ozbay (2012) make the case that if a DM follows a particular choice procedure, then a policy maker should take this into account when making welfare judgments as the procedure may be informative for the welfare analysis. They point out that by ignoring this information, the model free approach may produce erroneous welfare conclusions.[17] Our work here suggests another reason why some understanding of the procedure by which DMs make choices may matter for welfare, specifically for the question about whether an intervention on the part of the policy maker is at all called for.

To understand the last observation, consider the following pattern of menu dependent non-rational choices over three alternatives x, y and z: $c(xyz) = y$, $c(xy) = x$, $c(xz) = x$, $c(yz) = y$. If the DM making these choices is of, say, the EPH type, then we can conclude that the preferences $x \succ^* y \succ^* z$ and $y \succ x \succ z$ of the should and want selves, respectively, rationalize these choices. Alternatively, consider the possibility that these are the choices of a DM who follows the CLA model of Masatlioglu, Nakajima, and Ozbay

[17] Refer to Example 1 in Section II of Masatlioglu, Nakajima, and Ozbay (2012) for an example illustrating this.

(2012). Specifically, suppose we can elicit that her tastes are specified by the ranking $y \mathrel{\hat{\succ}} x \mathrel{\hat{\succ}} z$ and her consideration set mapping by $\Gamma(xy) = x$, $\Gamma(xz) = xz$, $\Gamma(yz) = yz$, $\Gamma(xyz) = xyz$. That is, the alternative y gets the DM's attention only in the presence of z, otherwise, she considers all alternatives.

Are the welfare implications in terms of the desirability of any kind of intervention different for the two scenarios? We certainly think that they are. If the policy maker happens to know that the DM in case is of the CLA type, then some kind of intervention, albeit of the soft type, may be desirable. Perhaps, she can tinker with the choice architecture so that alternative y draws the DM's attention generally and not just when z is part of the menu. On the other hand, we do not think any such intervention on the part of the policy maker is desirable if the DM is of the EPH type. Such a DM is faced with a non-trivial intrapersonal conflict and makes the best choices she can while trying to balance her emotions of anxiety and guilt. There is not much, if anything, that the policy maker can do to help her make better choices. The policy maker is best advised to leave this DM alone. Hence, our claim that the two situations are different and the ability to draw this distinction comes about precisely from taking cognizance of the choice procedure through which choices are generated.

The fact that the scope for welfare interventions is limited with the type of behavioral DMs we have modelled in this paper is generally true. Perhaps, a case can be made that if the policy maker is sure that some choices of our DM are being made in a severely ego depleted state or under conditions of heavy cognitive load, then some form of intervention that favorably changes the choice context or environment may be in order. Other than this, it is hard to think why welfare interventions may be desirable for EPH/GEPH type DMs. In that sense, our model can stay clear of the tricky debates surrounding paternalistic interventions that the welfare economics of behavioral decision makers often engenders.

A Appendix

A.1 Proofs in Section 2

A.1.1 Proof of Theorem 2.1

Necessity: Let $c : \mathcal{P}(X) \to X$ be an EPH with $(\succ^*, \succ)$ as the ordered pair of strict preference rankings. We show that c satisfies the three axioms of NC, NBC and WARP-EP.

<u>*NC*</u>: Consider $x \in S$, s.t., $x \neq c(xs)$, for all $s \in S \setminus x$. This implies that $s \succ^* x$, for all

$s \in S \setminus x$ and, accordingly, $x = \underline{M}(S; \succ^*)$. Therefore, $x \neq c(S)$.

<u>*NBC*</u>: Consider $X = \{x_1, \ldots, x_{n+1}\}$ where $c(x_i x_{i+1}) = x_i, i = 1, \ldots, n$. This implies that $x_i \succ^* x_{i+1}$, for all $i = 1, \ldots, n$. Since, $\succ^*$ is transitive, it follows that $x_1 \succ^* x_{n+1}$. Accordingly, $c(x_1 x_{n+1}) = x_1$

<u>*WARP-EP*</u>: Let $x = c(S)$, $y \in \psi(S)$ and $x \in \psi(T)$, for $S, T \in \mathcal{P}(X)$. Since $y \in \psi(S)$, i.e, there exists some $S' \subseteq S$, $|S'| \geq 2$, such that $c(S') = y$, it follows that there exists $z \in S'$ such that $y \succ^* z$ and, accordingly, $y \neq \underline{M}(S; \succ^*)$. Hence, $x \succ y$. Similarly, since $x \in \psi(T)$, a similar argument establishes that $x \neq \underline{M}(T; \succ^*)$ and, hence, $x \in T \setminus \underline{M}(T; \succ^*)$. Now, if $y \in T \setminus \underline{M}(T; \succ^*)$, since $x \succ y$, we can conclude that $y \neq c(T)$. Of course, if $y \notin T \setminus \underline{M}(T; \succ^*)$, the conclusion is obvious.

Sufficiency: Let $c : \mathcal{P}(X) \to X$ satisfy NC, NBC and WARP-EP. We show below that we can identify strict preference rankings $\succ^*$ and $\succ$ on X such that with respect to the ordered pair $(\succ^*, \succ)$, c is an EPH.

Define $\succ^* \subseteq X \times X$ as follows: for any $x, y \in X$, $x \neq y$, $x \succ^* y$ if $x = c(xy)$. We establish that $\succ^*$ is a strict preference ranking, i.e., $\succ^*$ is:
<u>Total</u> : $c(xy) \neq \emptyset$, for all $x, y \in X$, $x \neq y$. Thus, either $x \succ^* y$ or $y \succ^* x$.
<u>Asymmetric</u> : Suppose, towards a contradiction, $x \succ^* y$ and $y \succ^* x$. Then by definition, $x = c(xy)$ and $y = c(xy)$!
<u>Transitive</u> : Let $x \succ^* y$ and $y \succ^* z$. This implies $x = c(xy)$ and $y = c(yz)$. Since c satisfies NBC, it follows that $x = c(xz)$. Hence $x \succ^* z$.

Define $\succ \subseteq X \times X$ as follows: for any $x, y \in X$, $x \neq y$, $x \succ y$ if either (i) there exists $S \in \mathcal{P}(X)$, $|S| > 2$ and $x, y \in S$ such that $x = c(S)$ and $y = c(S')$, for some $S' \subseteq S$, $|S'| \geq 2$; or (ii) $y \neq c(S)$ for any $S \in \mathcal{P}(X)$ with $|S| \geq 2$. We establish that $\succ$ is a strict preference ranking, i.e., $\succ$ is:
<u>Total</u> : Since X is a finite set and c satisfies NBC, there exists a unique alternative, call it $\underline{z}$, such that $c(\underline{z}z) \neq \underline{z}$ for all $z \in X \setminus \underline{z}$. Let $x, y \in X$, $x \neq y$. First, consider the case $x, y \neq \underline{z}$ and the set $\{x, y, \underline{z}\}$. Since, $x = c(x\underline{z})$ and $y = c(y\underline{z})$, by NC, we know that $c(xy\underline{z}) \neq \underline{z}$. If $c(xy\underline{z}) = x$, then $x \succ y$, otherwise $y \succ x$. Next, consider the case that one of x or y, wlog say y, is $\underline{z}$. Accordingly, since $c(yz) \neq y$, for any $z \in X \setminus y$, by NC it follows that there exists no $S \in \mathcal{P}(X)$ with $|S| \geq 2$ such that $c(S) = y$. Hence, $x \succ y$. This establishes that $\succ$ is total.
<u>Asymmetric</u> : Suppose $x \succ y$. Clearly, $x \neq \underline{z}$ since there exists no $S \in \mathcal{P}(X)$ with $|S| \geq 2$ such that $c(S) = \underline{z}$. On the other hand, if $y = \underline{z}$, then for the same reason, $\neg[y \succ x]$. Now consider the case $y \neq \underline{z}$. Then, $x \succ y$ implies that there exists $S \in \mathcal{P}(X)$, $|S| > 2$ and $x, y \in S$ such that $x = c(S)$ and $y = c(S')$, for some $S' \subseteq S$, $|S'| \geq 2$. WARP-EP then implies that there does not exist T, $|T| > 2$, and $x, y \in T$, such that $y = c(T), x = c(T')$, for $T' \subseteq T$, $|T'| \geq 2$. Hence, $\neg[y \succ x]$.

Transitive : Let $x \succ y$ and $y \succ z$, for some $x, y, z \in X$. Clearly, by the argument made above, $x, y \neq \underline{z}$. Further, if $z = \underline{z}$, then clearly $x \succ z$ and our desired conclusion is immediate. So, assume $z \neq \underline{z}$. Now, consider the sets $\{x, y, \underline{z}\}$ and $\{y, z, \underline{z}\}$. Since, $x = c(x\underline{z})$ and $y = c(y\underline{z})$, by WARP-EP, it follows that $y \neq c(xy\underline{z})$ and $z \neq c(yz\underline{z})$. Further, by NC, $\underline{z} \neq c(xy\underline{z})$ and $\underline{z} \neq c(yz\underline{z})$. Hence, $x = c(xy\underline{z})$ and $y = c(yz\underline{z})$. Now, consider the set $\{x, y, z, \underline{z}\}$. We know that $a = c(a\underline{z})$, for $a = x, y, z$. Therefore, by NC, $\underline{z} \neq c(xyz\underline{z})$. Further, by WARP-EP, $y, z \neq c(xyz\underline{z})$. Hence, $x = c(xyz\underline{z})$ and it follows that $x \succ z$.

To show: $(\succ^*, \succ)$ is an EPH representation of c.

Pick any set $S \in \mathcal{P}^*(X)$ and let $x = c(S)$. First, consider the case when $|S| = 2$, i.e., $S = \{x, y\}$ for some $y \neq x$. In this case, it follows that $x \succ^* y$ and, therefore, $x = \overline{M}(S \setminus \underline{M}(S, \succ^*); \succ)$. Next, consider the case $|S| > 2$. Since c satisfies NC, there exists $z \in S$, such that, $c(xz) = x$. This implies $x \succ^* z$. Thus, $x \in S \setminus \underline{M}(S; \succ^*)$. Now consider any $y \in S \setminus \underline{M}(S; \succ^*)$, $y \neq x$, i.e, there exists $S' \subseteq S$, $|S'| \geq 2$ such that $c(S') = y$. Hence, $x \succ y$ and $x = \overline{M}(S \setminus \underline{M}(S, \succ^*); \succ)$.

A.1.2 The axioms are independent

(a) The choice function c_1 below satisfies NBC and WARP-EP but violates NC: y is never chosen in any two-element set, but is chosen in $\{x, y, z\}$.

	xy	xz	xw	yz	yw	zw	xyz	xyw	xzw	yzw	$xyzw$
$c_1(.)$	x	x	x	z	w	w	y	w	w	w	w
$\psi(.)$	x	x	x	z	w	w	xyz	xw	xw	zw	$xyzw$

(b) The choice function c_2 below satisfies NBC and NC but violates WARP-EP: $x = c(xyw), w \in \psi(xyw)$ and $x \in \psi(xzw)$, but $w = c(xzw)$.

	xy	xz	xw	yz	yw	zw	xyz	xyw	xzw	yzw	$xyzw$
$c_2(.)$	x	x	x	z	w	w	x	x	w	w	w
$\psi(.)$	x	x	x	z	w	w	xz	xw	xw	zw	xzw

(c) The choice function c_3 below satisfies WARP-EP and NC but violates NBC: there exists a binary cycle in sets $\{x, y, z\}$ and $\{x, z, w\}$.

	xy	xz	xw	yz	yw	zw	xyz	xyw	xzw	yzw	$xyzw$
$c_3(.)$	x	z	x	y	w	w	x	x	x	w	x
$\psi(.)$	x	z	x	y	w	w	xyz	xw	xzw	yw	$xyzw$

A.1.3 Proof of Proposition 2.1

Let $(\succ^*, \succ)$ and $(\tilde{\succ}^*, \tilde{\succ})$ be two EPH representations of a choice function c. Then, for any $x, y \in X$, $x \neq y$,

$$x \succ^* y \Leftrightarrow x = c(xy) \Leftrightarrow x \,\tilde{\succ}^* y$$

Let $\underline{z} = \underline{M}(X, \succ^*) = \underline{M}(X, \tilde{\succ}^*)$. Then for any $x, y \in X$, $x \neq y$, $x, y \neq \underline{z}$,

$$x \succ y \Leftrightarrow x = c(xy\underline{z}) \Leftrightarrow x \,\tilde{\succ}\, y$$

A.2 Proofs in Section 3

A.2.1 Proof of Theorem 3.1

The proof of necessity of NC and NBC for the representation is along similar lines as in Theorem 2.1.

Sufficiency: Let $c : \mathcal{P}(X) \to \mathcal{P}(X)$ satisfy NC and NBC.

Define $\succ^* \subseteq X \times X$ like in the Proof of Theorem 2.1: for any $x, y \in X$, $x \neq y$, $x \succ^* y$ if $x = c(xy)$. Like we showed there, if c satisfies NBC, then $\succ^*$ is a strict preference ranking.

Define $\succ \subseteq X \times X$ as $x \succ y$ if $y \succ^* x$. Clearly $\succ$ is a strict preference ranking.

Define the should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ by:

$$W_{\succ^*}(S) = D_S(c(S)) = \{y \in S : c(c(S)y) = c(S)\}$$

<u>To show</u>: $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$ is a GEPH representation of c.

Consider any $S \in \mathcal{P}^*(X)$ and let $c(S) = x$. By NC, there exists $z \in S$ s.t. $c(xz) = x$. Accordingly, $W_{\succ^*}(S) = D_S(x) \neq \emptyset$. Further, $x \notin D_S(x)$. Hence, $W_{\succ^*}(S) \subsetneq S$. Now, consider any $y \in S \setminus W_{\succ^*}(S)$, $y \neq x$; i.e., $c(xy) = y$. This means $y \succ^* x$ and, accordingly, $x \succ y$.

A.2.2 Proof of Proposition 3.1

Suppose, towards a contradiction, c does not satisfy WARP. That is, there exists $S, T \in \mathcal{P}(X)$, $x, y \in S \cap T$ such that $c(S) = x$ and $c(T) = y$. Wlog, assume that $x \succ^* y$. That is $W_{\succ^*}(xy) = y$. However, since $W_{\succ^*}(xy) \subseteq W_{\succ^*}(T)$, this implies $y \in W_{\succ^*}(T)$ and $c(T) \neq y$!

Next, pick any S and let $\tilde{x} \in S$ be s.t. $\tilde{x} \succ^* z$, for all $z \in S \setminus \tilde{x}$. This implies that $c(\tilde{x}z) = \tilde{x}$, for all $z \in S \setminus \tilde{x}$; or, $W_{\succ^*}(\tilde{x}z) = z$, for all $z \in S \setminus \tilde{x}$. Since, $W_{\succ^*}(\tilde{x}z) \subseteq W_{\succ^*}(S)$, this then implies that $z \in W_{\succ^*}(S)$, for all $z \in S \setminus \tilde{x}$ and, hence, $S \setminus W_{\succ^*}(S) = \tilde{x}$. Accordingly, $c(S) = \tilde{x}$.

A.2.3 Proof of Theorem 3.2

Necessity: Let $c : \mathcal{P}(X) \to X$ be a monotonic GEPH with parameters $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$. We show that c satisfies the NBC and SARP-GEP.

<u>NBC</u>: Let $x_1, \ldots, x_{n+1} \in X$ and $c(x_i x_{i+1}) = x_i$, for all $i = 1, \ldots, n$. This implies that $x_i \succ^* x_{i+1}$, for all $i = 1, \ldots, n$. Since, $\succ^*$ is transitive, it follows that $x_1 \succ^* x_{n+1}$. Accordingly, $c(x_1 x_{n+1}) = x_1$.

<u>*SARP-GEP*</u>: Consider $S_1, \ldots, S_n \in \mathcal{P}(X)$ and $x_1, \ldots, x_n \in X$, all distinct, as in the statement of SARP-GEP. We have shown in the text that for any $S \in \mathcal{P}(X)$, $\tilde{\psi}(S) \subseteq S \backslash W_{\succ^*}(S)$. Accordingly, for $i = 1, \ldots, n-1$, $x_{i+1} \in \tilde{\psi}(S_i)$ implies that $x_{i+1} \in S_i \backslash W_{\succ^*}(S_i)$. Hence $x_i \succ x_{i+1}$, for $i = 1, \ldots, n-1$. Since $\succ$ is transitive, we have $x_1 \succ x_n$. Finally $x_1 \in \tilde{\psi}(S_n)$ implies that $x_1 \in S_n \setminus W_{\succ^*}(S_n)$. Hence $c(S_n) \neq x_n$.

Sufficiency: Let $c : \mathcal{P}(X) \to X$ satisfy NBC and SARP-GEP. We show below that we can identify strict preference rankings $\succ^*$ and $\succ$ on X and a should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ that satisfies monotonicity such that with respect to the ordered pair $(\succ^*, \succ)$ and the mapping $W_{\succ^*}$, c is a monotonic GEPH.

Define $\succ^* \subseteq X \times X$ like we did above: for any $x, y \in X$, $x \neq y$, $x \succ^* y$ if $x = c(xy)$. As established there, $\succ^*$ defined thus is a strict preference ranking.

Next, to define the preference ranking $\succ \subseteq X \times X$, start with the binary relation $P \subseteq X \times X$ defined earlier: for $x, y \in X, x \neq y$, xPy if there exists $S \in \mathcal{P}(X)$ such that $x = c(S)$ and $y \in \tilde{\psi}(S)$. By Proposition 3.2, we know that if a choice function c satisfies SARP-GEP, then P is acyclic. Denote the transitive closure of P by P^*. That is, P^* is a partial order. By Szpilrajn's theorem, we know that this partial order can be extended to a linear order. We define $\succ$ as the asymmetric component of this linear order.

Define the mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ by

$$W_{\succ^*}(S) = \bigcap_{T \in \mathcal{T}_S} D_T(c(T))$$

To establish that this is a well defined should not set mapping, first, note that for any $S \in \mathcal{P}^*(X)$, $c(S) \notin D_S(c(S))$ and, accordingly, $c(S) \notin W_{\succ^*}(S)$. Hence $W_{\succ^*}(S) \subsetneq S$. Next, note that since c satisfies NBC, for any such set S, there exists $\underline{x}_S$ such that $c(y\underline{x}_S) = y$ for all $y \in S \setminus \{\underline{x}_S\}$. Moreover, from Proposition 3.3 we know that since c satisfies SARP-GEP, it satisfies NC as well. Hence, $c(S) \neq \underline{x}_S$. That is, $\underline{x}_S \in D_S(c(S)) \neq \emptyset$. By a similar argument, $D_T(c(T)) \neq \emptyset$ for all $T \in \mathcal{T}_S$, $T \neq S$. Further, for any such T and $z \in T \setminus S$, $c(z\underline{x}_S) = z$. Accordingly, $\underline{x}_S \in D_T(c(T))$. Together, these observations imply that $W_{\succ^*}(S) \neq \emptyset$. Finally, consider any $x \in S \setminus W_{\succ^*}(S)$ and $y \in W_{\succ^*}(S)$. That is, there exists $\hat{T} \in \mathcal{T}_S$ such that $c(c(\hat{T})x) = x$ and $c(c(\hat{T})y) = c(\hat{T})$. By NBC it then follows that $c(xy) = x$ and by our definition of $\succ^*$, $x \succ^* y$. This establishes that $W_{\succ^*}$ is a well defined should not set mapping. To establish that it satisfies monotonicity, let $R \supseteq S$ be such that for all $z \in R \setminus S$, there exists $y \in S$ s.t., $z \succ^* y$; i.e., by the definition of $\succ^*$, $c(yz) = z$. This means that $R \in \mathcal{T}_S$ and $\mathcal{T}_R \subseteq \mathcal{T}_S$. Accordingly, $W_{\succ^*}(S) \subseteq W_{\succ^*}(R)$.

<u>To show</u>: $(\succ^*, \succ, W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X))$ is a monotonic GEPH representation of c.

Pick any set $S \in \mathcal{P}(X)$ and let $x = c(S)$. First, note that since $W_{\succ^*}(S) = \bigcap\limits_{T \in \mathcal{T}_S} D_T(c(T))$ and $x \notin D_S(c(S))$, $x \notin W_{\succ^*}(S)$. Now consider $y \in S$, $y \neq x$ such that $y \notin W_{\succ^*}(S)$. That is, there exists $\hat{T} \in \mathcal{T}_S$ s.t., $y \notin D_{\hat{T}}(c(\hat{T}))$. Thus $y \in \tilde{\psi}(S)$ and accordingly xPy. Finally, since $P \subseteq \succ$, it follows that $x \succ y$.

A.2.4 Proof of Proposition 3.3

Suppose towards a contradiction, that the choice function c satisfies SARP-GEP but violates NC. That is, there exists $S \in \mathcal{P}(X)$, s.t., $c(S) = y$ and $y \neq c(xy)$, for all $x \in S \setminus y$. This implies $D_S(c(S)) = \emptyset$ and hence, $\bigcap_{T \in \mathcal{T}_S} D_T(c(T)) = \emptyset$. Thus $\tilde{\psi}(S) = S$. Next consider the set $\{x', y\}$, for some $x' \in S \setminus y$ and note that $S \in \mathcal{T}_{\{x',y\}}$. Accordingly, $D_S(c(S)) = \emptyset$ implies $\tilde{\psi}(\{x', y\}) = \{x', y\}$. That is, $c(x'y) = x'$ and $y \in \tilde{\psi}(\{x', y\})$. Further $c(S) = y$ and $x' \in \tilde{\psi}(S)$. This violates SARP-GEP!

A.2.5 Proof of Proposition 3.4

Necessity: Suppose $\neg[xP^*y]$. Then, the following two cases are possible: Either yP^*x or $\neg[yP^*x]$. Consider the first case and let $\succ$ be a strict preference ranking of the want-self in a monotonic GEPH representation. Since P^* is defined as the transitive closure of P, yP^*x implies that there exists a sequence $(z_m)_{m=1}^M$ in X such that yPz_1, z_1Pz_2, $\ldots$, z_MPx. Further, for any such $\succ$, since $P \subseteq \succ$ and $\succ$ is transitive, it follows that $y \succ x$. In the second case, where $\neg[yP^*x]$, there exists no sequence $(z_m)_{m=1}^M$ in X such that yPz_1, z_1Pz_2, $\ldots$, z_MPx. In this case it is possible to extend P^* to a linear order under whose asymmetric component $\succ$ we have $y \succ x$. The proof of Theorem 3.1 establishes that any such asymmetric component of a linear order can be part of a monotonic GEPH representation. Therefore, in either case, x is not revealed to be preferred to y, i.e., $\neg[xQ^*y]$.

Sufficiency: We have already shown in Section 3.3 that if xPy, then xQ^*y. Now, consider the case when xP^*y. Since P^* is defined as the transitive closure of P, this implies that there exists a sequence $(z_m)_{m=1}^M$ in X such that xPz_1, z_1Pz_2, $\ldots$, z_MPy. In this case, we know that for any $\succ$ that is part of a monotonic GEPH representation, $P \subseteq \succ$ and, hence, $x \succ z_1$, $z_1 \succ z_2$, $\ldots$, $z_M \succ y$. Further, since $\succ$ is transitive it follows that $x \succ y$ and, hence, xQ^*y.

A.3 Proofs in Section 4

A.3.1 Proof of Proposition 4.1

Let $(\succ^*, \succ)$ be an EPH representation of the choice function c. Further, let $\{x, y\} \subseteq S \subseteq T$ and $x = c(xy) = c(T)$. $x = c(xy)$ implies that $x \succ^* y$. There are two possibilities. (i) If $y = \underline{M}(S, \succ^*)$, then clearly $y \neq c(S)$. (ii) If $y \neq \underline{M}(S, \succ^*)$, then $y \neq \underline{M}(T, \succ^*)$ and $x \succ y$. Therefore, if $y \neq \underline{M}(S, \succ^*)$, then $x \neq \underline{M}(S, \succ^*)$ and $y \neq c(S)$.

The following example shows that a monotonic GEPH may violate Weak WARP. Consider $X = \{x, y, z, w\}$ and the choice function c specified in the table.

	xy	xz	xw	yz	yw	zw	xyz	xyw	xzw	yzw	$xyzw$
$c(.)$	x	x	x	y	y	z	y	y	x	y	x
$W_{\succ^*}(.)$	y	z	w	z	w	w	z	w	zw	zw	yzw

Clearly c violates Weak WARP as $c(xy) = c(xyzw) = x$ and $c(xyz) = y$. It is also straightforward to verify that with strict preference rankings $(\succ^*, \succ)$ given by $x \succ^* y \succ^* z \succ^* w$ and $y \succ x \succ w \succ z$, and should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ specified in the table, c is a monotonic GEPH.

A.3.2 Comparison with Two-stage chooser

Example A.1. (A TSC but not a monotonic GEPH). Let $X = \{x, y, z, w\}$ and consider the choice function c specified in the table.

	xy	xz	xw	yz	yw	zw	xyz	xyw	xzw	yzw	$xyzw$
$c(.)$	y	x	x	y	y	w	y	x	x	y	x

It is straightforward to verify that c is a TSC with the following preference rankings: $x \succ_1 w \succ_1 y \succ_1 z$, $y \succ_2 x \succ_2 w \succ_2 z$. However, c is not a monotonic GEPH. To establish this, assume otherwise. Then choices over the binaries imply that the preferences of the DM's should-self are: $y \succ^* x \succ^* w \succ^* z$. Further, $c(xyw) = x$ implies that $x \succ y$. This is turn implies that, since $c(xyz) = y$, $W_{\succ^*}(xyz) = \{x, z\}$. But then monotonicity of the should not set mapping implies that $\{x, z\} \subseteq W_{\succ^*}(xyzw)$ and, accordingly, $c(xyzw) \neq x$!

Example A.2. (An EPH but not a TSC). Let $X = \{x, y, z, w, v\}$ and consider the choice function specified in the table below.

	xy	xz	xw	xv	yz	yw	yv	zw	zv	wv	xyz	xyw	xyv
$c(.)$	x	x	x	x	y	y	y	z	v	v	y	y	y

	xzw	xzv	xwv	yzw	yzv	ywv	zwv	$xyzw$	$xyzv$	$xywv$	$xzwv$	$yzwv$	$xyzwv$
$c(.)$	x	x	x	y	y	y	v	y	y	y	x	y	y

It is straightforward to verify that c is an EPH with $(\succ^*, \succ)$ given by: $x \succ^* y \succ^* v \succ^* z \succ^* w$ and $y \succ x \succ v \succ w \succ z$. To see that it is not a TSC, suppose otherwise—say it is a TSC with first stage and second stage preference rankings denoted by $\succ_1$ and $\succ_2$, respectively. Then, from choices over binaries, it follows that $\succ_2$ is given by: $x \succ_2 y \succ_2 v \succ_2 z \succ_2 w$. Now consider the sets $\{x, y, z\}$, $\{x, y, w\}$ and $\{x, z, w, v\}$. Since $y = c(xyz) = c(xyw)$ and $x \succ_2 y$, it must be that x gets eliminated in the first round in both these sets, i.e., $y \succ_1 x$, $z \succ_1 x$ and $w \succ_1 x$. But then x gets eliminated in the first round in $\{x, z, w, v\}$ and hence $c(xzwv) \neq x$!

A.3.3 Proof of Proposition 4.2

To show that a choice function c is a CLA, it suffices to show that it satisfies WARP(LA) or, equivalently, that the binary relation $\tilde{P}$ on X, defined from c as specified below, is acyclic:

$$x\tilde{P}y \text{ if there exists } S \in \mathcal{P}(X), s.t., x = c(S) \neq c(S \setminus y)$$

Let c be an EPH and consider $x_1, \ldots, x_n \in X$, s.t., $x_i \tilde{P} x_{i+1}$, for $i = 1, \ldots n-1$. $x_i \tilde{P} x_{i+1}$ implies that there exists $S_i \in \mathcal{P}(X)$, s.t., $x_i = c(S_i) \neq c(S_i \setminus x_{i+1})$. This implies $x_i \succ^* x_{i+1}$, for all $i = 1, \ldots, n-1$. Since $\succ^*$ is transitive, $x_1 \succ^* x_n$. This means there does not exist S s.t., $x_n = c(S)$ and $c(S) \neq c(S \setminus x_1)$ for this would imply that $x_n \succ^* x_1$. Thus, $\neg[x_n \tilde{P} x_1]$ and, hence, $\tilde{P}$ is acyclic. This establishes that c is a CLA.

The following choice function c on $X = \{x, y, z, w, v\}$ in the table below is a monotonic GEPH but not CLA.

	xy	xz	xw	xv	yz	yw	yv	zw	zv	wv	xyz	xyw	xyv
$c(.)$	y	x	w	x	y	w	y	w	z	w	x	y	x
$W_{\succ^*}(.)$	x	z	x	v	z	y	v	z	v	v	z	x	v

	xzw	xzv	xwv	yzw	yzv	ywv	zwv	$xyzw$	$xyzv$	$xywv$	$xzwv$	$yzwv$	$xyzwv$
$c(.)$	x	x	x	y	z	y	z	y	x	y	x	y	w
$W_{\succ^*}(.)$	z	zv	v	z	v	v	v	xz	zv	xv	zv	zv	$xyzv$

For this choice function c, consider the binary relation $\tilde{P}$ defined above. Observe that it has a cycle: $y\tilde{P}w$ as $y = c(xywv)$ and $x = c(xyv)$. Also, $w\tilde{P}y$ as $w = c(xyzwv)$ and $x = c(xzwv)$. Hence, c violates WARP(LA) and is not a CLA. However, c is a monotonic GEPH with the should not set mapping $W_{\succ^*} : \mathcal{P}^*(X) \to \mathcal{P}(X)$ specified in the table and strict preference rankings $(\succ^*, \succ)$ given by: $w \succ^* y \succ^* x \succ^* z \succ^* v$ and $x \succ z \succ y \succ w \succ v$.

A.3.4 Comparison with CAF model

Example A.3. (An EPH but not a CAF). Let $X = \{x, y, z\}$. There are two attributes and the attribute ranking is as follows: $y_1 > z_1 > x_1$ and $x_2 > y_2 > z_2$.

	xy	xz	yz	xyz
$c(.)$	y	x	y	x

To see that the choice function c specified in the table is not a CAF, refer to Kimya (2018) Section II.C. However this choice function is an EPH with preference rankings $\succ^*$ and $\succ$ given by $y \succ^* x \succ^* z$ and $x \succ y \succ z$.

References

Achtziger, Anja, Carlos Alós-Ferrer, and Alexander K Wagner. 2015. "Money, depletion, and prosociality in the dictator game." *Journal of Neuroscience, Psychology, and Economics* 8 (1):1–14.

Au, Pak Hung and Keiichi Kawai. 2011. "Sequentially rationalizable choice with transitive rationales." *Games and Economic Behavior* 73 (2):608–614.

Bajraj, Gent and Levent Ülkü. 2015. "Choosing two finalists and the winner." *Social Choice and Welfare* 45 (4):729–744.

Bardsley, Nicholas. 2008. "Dictator game giving: Altruism or artefact?" *Experimental Economics* 11:122–133.

Baumeister, Roy F, Brandon J Schmeichel, and Kathleen D Vohs. 2007. "Self-regulation and the executive function: The self as controlling agent." In *Social psychology: Handbook of basic principles*, edited by A. W. Kruglanski and E. T. Higgins. The Guilford Press, 516–539.

Benhabib, Jess and Alberto Bisin. 2005. "Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions." *Games and Economic Behavior* 52 (2):460–492.

Bernheim, B Douglas and Antonio Rangel. 2009. "Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics." *Quarterly Journal of Economics* 124 (1):51–104.

Blackorby, Charles, David Nissen, Daniel Primont, and R Robert Russell. 1973. "Consistent intertemporal decision making." *Review of Economic Studies* 40 (2):239–248.

Camerer, Colin F. 2011. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2013. "Rationalization." *Theoretical Economics* 8 (3):775–800.

Cherry, Todd L, Peter Frykblom, and Jason F Shogren. 2002. "Hardnose the dictator." *American Economic Review* 92 (4):1218–1221.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness." *Economic Theory* 33 (1):67–80.

Dekel, Eddie, Barton L Lipman, and Aldo Rustichini. 2009. "Temptation-driven preferences." *The Review of Economic Studies* 76 (3):937–971.

Eckel, Catherine C and Philip J Grossman. 1996. "Altruism in anonymous dictator games." *Games and Economic Behavior* 16 (2):181–191.

Franzen, Axel and Sonja Pointner. 2012. "Anonymity in the dictator game revisited." *Journal of Economic Behavior & Organization* 81 (1):74–81.

Freud, Sigmund. 1921/1961. "The ego and the id." In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIX (1923-1925): The Ego and the Id and Other Works.* 1–66.

Fudenberg, Drew and David K Levine. 2006. "A dual-self model of impulse control." *American Economic Review* 96 (5):1449–1476.

———. 2011. "Risk, delay, and convex self-control costs." *American Economic Journal: Microeconomics* 3 (3):34–68.

———. 2012. "Timing and self-control." *Econometrica* 80 (1):1–42.

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

Halali, Eliran, Yoella Bereby-Meyer, and Axel Ockenfels. 2013. "Is it all about the self? The effect of self-control depletion on ultimatum game proposers." *Frontiers in Human Neuroscience* 7, 240:1–8.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L Smith. 1996. "Social distance and other-regarding behavior in dictator games." *American Economic Review* 86 (3):653–660.

Horan, Sean. 2016. "A simple model of two-stage choice." *Journal of Economic Theory* 162:372–406.

Kimya, Mert. 2018. "Choice, consideration sets, and attribute filters." *American Economic Journal: Microeconomics* 10 (4):223–47.

Kreps, David M. 1979. "A representation theorem for "Preference for Flexibility"." *Econometrica* 47 (3):565–577.

Liang, Meng-Yu, Simon Grant, and Sung-Lin Hsieh. 2019. "Costly self-control and limited willpower." *Economic Theory,* forthcoming. DOI: 10.1007/s00199-019-01231-6.

List, John A. 2007. "On the interpretation of giving in dictator games." *Journal of Political Economy* 115 (3):482–493.

Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay. 2017. "When more is less: Limited consideration." *Journal of Economic Theory* 170:70–85.

Loewenstein, George, Ted O'Donoghue, and Sudeep Bhatia. 2015. "Modeling the interplay between affect and deliberation." *Decision* 2 (2):55–81.

Manzini, Paola and Marco Mariotti. 2007. "Sequentially rationalizable choice." *American Economic Review* 97 (5):1824–1839.

———. 2012. "Categorize then choose: Boundedly rational choice and welfare." *Journal of the European Economic Association* 10 (5):1141–1165.

Manzini, Paola, Marco Mariotti, and Christopher J Tyson. 2013. "Two-stage threshold representations." *Theoretical Economics* 8 (3):875–882.

Martinsson, Peter, Kristian Ove R Myrseth, and Conny Wollbrant. 2012. "Reconciling pro-social vs. selfish behavior: On the role of self-control." *Judgment and Decision Making* 7 (3):304–315.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y Ozbay. 2012. "Revealed attention." *American Economic Review* 102 (5):2183–2205.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Emre Ozdenoren. 2020. "Willpower and compromise effect." *Theoretical Economics* 15 (1):279–317.

Noor, Jawwad and Norio Takeoka. 2010. "Uphill self-control." *Theoretical Economics* 5 (2):127–158.

———. 2015. "Menu-dependent self-control." *Journal of Mathematical Economics* 61:1–20.

Ozdenoren, Emre, Stephen W Salant, and Dan Silverman. 2012. "Willpower and the optimal control of visceral urges." *Journal of the European Economic Association* 10 (2):342–368.

Peleg, Bezalel and Menahem E Yaari. 1973. "On the existence of a consistent course of action when tastes are changing." *Review of Economic Studies* 40 (3):391–401.

Pocheptsova, Anastasiya, On Amir, Ravi Dhar, and Roy F Baumeister. 2009. "Deciding without resources: Resource depletion and choice in context." *Journal of Marketing Research* 46 (3):344–355.

Schelling, Thomas C. 1984. "Self-command in practice, in policy, and in a theory of rational choice." *American Economic Review* 74 (2):1–11.

Simonson, Itamar. 1989. "Choice based on reasons: The case of attraction and compromise effects." *Journal of Consumer Research* 16 (2):158–174.

Simonson, Itamar and Amos Tversky. 1992. "Choice in context: Tradeoff contrast and extremeness aversion." *Journal of Marketing Research* 29 (3):281–295.

Strotz, Robert Henry. 1955. "Myopia and inconsistency in dynamic utility maximization." *Review of Economic Studies* 23 (3):165–180.

Thaler, Richard H and Hersh M Shefrin. 1981. "An economic theory of self-control." *Journal of Political Economy* 89 (2):392–406.

Xu, Hanyi, Laurent Bègue, and Brad J Bushman. 2012. "Too fatigued to care: Ego depletion, guilt, and prosocial behavior." *Journal of Experimental Social Psychology* 48 (5):1183–1186.

Yoo, Jaewon, Hyunsik Park, and Wonjoon Kim. 2018. "Compromise effect and consideration set size in consumer decision-making." *Applied Economics Letters* 25 (8):513–517.