Cognitive Equilibrium*

Hannu Vartiainen†
Yrjö Jahnsson Foundation
and
Helsinki School of Economics

September 20, 2006

Abstract

We show that whenever a decision maker *reasons* about an optimal decision he is able to find one, even with non-transitive preferences. The existence of a reasoning process allows him to strategically manipulate *how* he reasons. A reasoning strategy that is robust against (finite) deviations is captured by the notion of cognitive equilibrum. We show that a cognitive equilibrium exists under all complete preferences, and characterize outcomes that can be implemented within it. Cognitive equilibria employ complex cognitive strategies. Simple strategies suffice only under transitive preferences. Robustness of the model is evaluated in the language of von Neumann-Morgenstern stable sets.

JEL: D11, D89.

Keywords: procedural rationality, non-transitive preferences, cognitive equilibrium.

*I am very grateful to Ariel Rubinstein for his comments.

†Address: Yrjö Jahnsson Foundation, Ludviginkatu 3-5, 00130 Helsinki, Finland. E-mail: hannu.vartiainen@yjs.fi.

1 Introduction

All standard economics is based on the assumption of transitive preferences. Indeed, many would equate transitivity with rationality, the key paradigm of the discipline. Without transitivity, it is argued, a rational decision maker (DM) cannot reason the optimal choice.

But what does *reasoning* actually mean? There are many potential ways to describe a reasoning process but, as the word "process" suggests, what is common with the approaches is that the *acts* of reasoning take place in a sequence. This paper studies what constraints does rationality impose on preferences if the ability to commit to a decision is used as a criterion, and reasoning is a process.

To be concrete, let us discuss for a moment the argument of why reasoning would be in conflict with a choice making if preferences are non-transitive. Let the DM's preferences $\succ$ exhibit a cycle $y \succ x$, $x \succ z$, $z \succ y$, over $\{x, y, z\}$. Whenever he is about to choose x, he rather replaces x with y, whenever he is about choose y, he rather replaces y with z, and whenever he is about to choose z, he rather replaces z with x. Hence it seems that the DM is unable to commit to any of the choices.

What this argument says is that the DM cannot identify a rational decision *recursively,* by eliminating options. But this is much more than is usually required in game theory, where the key paradigm is the concept of *equilibrium.* In this paper, we assume that the DM is able to perform equilibrium reasoning.

We model the cognitive process as an internal "reasoning game" where, at each stage, the DM chooses to either implement the outcome that he has proposed to himself in the previous stage (the outcome "on the table"), or he proposes a new outcome (brings a new alternative on the table that replaces the old one).[1] A cognitive strategy specifies a cognitive act for each finite history of cognitive acts. We assume that the DM is free to make finite deviations to his strategy. The question is whether there are deviations that make him better of relative to what would happen if he follows the strategy. Strategies that the DM can commit to are called *cognitive equilibria.*

To see how a cognitive equilibrium works, consider the above 3-cycle. Let

[1] Our model is essentially a one-player version of the Rubinstein bargaining game.

cognitive acts be contingent on the past history of cognitive acts. Partition the set of histories into two *phases,* p_1 and p_2 (start with, say, p_1). The partition is implicitly defined by transition from phase p_i to p_j, $i \neq j$, whenever x is on the table but replaced with y. In phase p_1, let the cognitive strategy implement alternative x or y were one of them on the table, and replace z with x, were z on the table. In phase p_2, implement z were it on the table, and replace x and y with y and z, respectively, were either of them on the table. Figure 1 illustrates the situation. The dashed arrows reflect the potential law of motion, constrained by the phase transitions, and the solid arrows the law of motion suggested by the strategy. Doubly circled alternatives are implemented when reached.

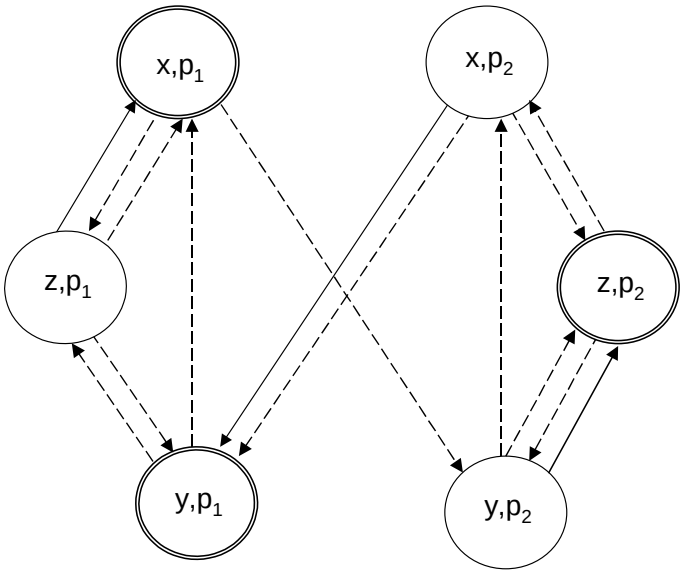

Figure 1.

A one-time (finite) deviation from the proposed strategy cannot lead to implementation of a preferred alternative. For example, if the phase is p_1 and the outcome on the table is x, then the obedient play implements x. A deviation leads either to implementing x in state p_1 via (z, p_1), or z in state p_2 via (y, p_2).

Constructing a cognitive equilibrium in the 3-cycle case is easy because of the symmetry of the situation. However, things become murky in larger choice domains. We characterize outcomes that are implementable in cognitive equilibrium

in any set up. The following combinatorial result makes the characterization possible: Given any complete, asymmetric relation $\succ$ on X, there is a subset C of X such that if an element x in C is dominated by an element y in X, then there is a third element z in C such that x, y and z form a cycle.

We show that a set C, which we call a *consistent choice set*, characterizes the outcomes that the DM can implement in a cognitive equilibrium. Moreover, also the converse holds: a set of outcomes that can be implemented in a cognitive equilibrium coincides with a consistent choice set. If preferences exhibit a maximal element, then the unique consistent choice set contains only this element. Moreover, a consistent choice set exists under any complete preference ordering (asymmetric for convenience). Finally, the *ultimate uncovered set* (see Fishburn, 1977; Miller, 1980; Dutta, 1988) is the unique maximal (in the sense of set inclusion) consistent choice set.[2] Thus the ultimate uncovered set characterizes what can be implemented under *any* cognitive equilibrium. Of central importance is the finding that any *covering set* (Dutta, 1988) is a consistent choice set (the converse is not true).

Cognitive equilibria often employ complex strategies, and may induce choice behavior that is in conflict with the *independence of irrelevant alternatives* (IIA) condition.[3] A simple cognitive strategy can form an equilibrium only if the preferences entertain a maximal element. Thus preferences that allow simple, context-free decision making must be transitive. This conclusion is in line with Rubinstein (1996), Rubinstein and Salant (2005), and Salant (2004), who study preferences that allow (cost) efficient decision making.

An immediate objection against the concept of cognitive strategy is that it assumes a specific correspondence between cognitive actions and physical actions, i.e. that an outcome becomes implemented if (and only if) it the DM decides to implement it when "on the table". However, there is no *a priori* reason to argue that this physical procedure is more plausible than any other. It is therefore important to evaluate how sensitive the model is to the underlying physical structure.

[2]The ultimate uncovered set is derived by an iterative application of the uncovering - operation. Moulin (1986) gives an axiomatic characterization for the uncovered set. Also see Shepsle and Weingast (1984), Banks (1985), and Laslier (1997).

[3]Duggan (2004) is a recent similarly motivated model.

This is done in a separate model that only assumes that a correspondence between cognitive and physical actions exists. Strategies are not restricted by an equilibrium condition but by the external and internal stability criteria á la von Neumann-Morgenstern.[4] Interestingly, the 1-1 correspondence between consistent choice sets and stable behavior remains valid. This suggests that the cognitive equilibrium -model is robust to the details of the physical cognitive structure.[5]

Section 2 defines the model. Section 4 characterizes the equilibria. Section 6 gives the vNM stable set interpretation. Section 7 discusses the properties of consistent choice sets. Section 8 ends with a fuller account of the related literature.

2 The Model

Let preferences of the DM be defined by a complete, asymmetric binary relation $\succ$ (a tournament) over a finite, nonsingleton set $X = \{x, y, z, ...\}$. Denote the associated weak relation by $\succeq$, i.e. $x \succeq y$ if $x \succ y$ or $x = y$, for all $x, y \in X$. We assume that there is a bad outcome $e \in X$ such that $x \succeq e$, for all $x \in X$.

The decision process of the DM is as follows: Nature chooses x_0. At time $t = 0, ...,$ if alternative $x_t \in X$ is on the table, then the DM chooses a cognitive action $a \in X \cup \{\mathtt{stop}\}$. If $a = \mathtt{stop}$, then x_t is implemented and the game ends. If $a = y \in X$, then y becomes the alternative on the table at time $t + 1$, i.e. $x_{t+1} = y$. If the game does not end in finite time, then the implemented outcome is e.

Note that the fact that a particular x has been put on the table puts no restrictions on possible proposals in the consecutive periods. In particular x can be put on the table again.

Denote by H the set of all nonterminal histories $\{x_0\} \times \cup_{k=0}^{\infty} X^k$. A *cognitive strategy* s specifies a cognitive action after each nonterminal history of cognitive actions. That is, s is a function

$$s : H \to X \cup \{\mathtt{stop}\}. \tag{1}$$

[4] As Dutta (1988) shows, a covering set D can be interpreted as the von Neumann - Morgenstern stable set solution, defined with respect to the covering relation over D.

[5] Which may not be that surprising given von Neumann's role in decision theory and in automata theory.

Given any history $h \in H$, a strategy s generates a maximal path $(h, y_0, ...)$ from h onwards such that $y_{k+1} = s(h, y_0, ..., y_k) \in X$, for all $k = 0, ...$ If the length of the generated path is k, i.e. $(h, y_0, ..., y_k, \texttt{stop})$ for some k, then y_k becomes implemented. If the generated path is infinite, then e becomes implemented.

Given strategy s, denote by $\pi_s[h]$ the outcome that becomes implemented if s is followed from history h onwards. Then $\pi_s[(h, x, a)]$ is the outcome that will become implemented if a cognitive action $a \in X \cup \{\texttt{stop}\}$ is chosen at history $(h, x) \in H$, and s is followed thereafter, i.e.

$$\pi_s[(h, x, a)] = \begin{cases} \pi_s[(h, x, y)], & \text{if } a = y \in X, \\ x, & \text{if } a = \texttt{stop}. \end{cases} \tag{2}$$

If, in particular, $a = s(h)$, then $\pi_s[(h, a)] = \pi_s[h]$.

Next we define the solution concept. We identify cognitive strategies to which the DM can commit not to make *finitely* many changes. Since finite deviations improve the DM's position only if the final deviation does, the one-deviation principle applies. Thus we propose the following equilibrium condition.

Definition 1 *A cognitive strategy s forms a* cognitive equilibrium *if*

$$\pi_s[h] \succeq \pi_s[(h, a)], \quad \textit{for all } a \in X \cup \{\texttt{stop}\}, \ \textit{for all } h \in H. \tag{3}$$

We say that the set Y of alternatives is *implementable* in cognitive equilibrium s if

$$Y = \{x \in X : s(h, x) = \texttt{stop}, \text{ for any } h \in H\}.$$

The initial condition x_0 may affect the alternative that will be implemented in Y but not the set Y itself. The sets of implementable outcomes are the main object of our study.

Note that the DM can always guarantee the x on the table by choosing $\texttt{stop}$. Therefore, by the definition of π_s, if s forms a cognitive equilibrium, then

$$\pi_s[(h, x)] \succeq x, \text{ for all } (h, x) \in H. \tag{4}$$

This condition gives a simple rule to characterize sets of alternatives that are implementable in cognitive equilibria.

We say that set X is *bipartitioned* by $\succ$ into X_1 and X_2 if $\{X_1, X_2\}$ is a partition of X and $x_1 \succ x_2$ for all $(x_1, x_2) \in X_1 \times X_2$.

Proposition 2 *Let X be bipartitioned by $\succ$ into X_1 and X_2. Then the set of alternatives that are implementable in a cognitive equilibrium s is a subset of X_1.*

Proof. Let s form a cognitive equilibrium. If it holds, to the contrary of the proposition, that $s(h, x_2) = \texttt{stop}$, for some $x_2 \in X_2$, then, by (3), $\pi_s[(h, x_2)] \succeq \pi_s[(h, x_2, x_1)]$, for any $x_1 \in X_1$. By construction, $\pi_s[(h, x_2)] = x_2$.. By (4), $\pi_s[(h, x_2, x_1)] \succeq x_1$. But then $x_2 \succeq x_1$, which contradicts the assumption that X is bipartitioned by $\succ$ into X_1 and X_2. ■

The following two corollaries are immediate.

Corollary 3 *If there is a (necessarily unique) $\succ$-maximal element x^*, then the set of alternatives that are implementable in a cognitive equilibrium contains only x^*.*

Thus the concept of cognitive equilibrium is consistent with the rational choice -paradigm when preferences are transitive.

Corollary 4 *The bad outcome e cannot ever belong to the set of alternatives that are implementable in a cognitive equilibrium.*

This means that an equilibrium cognitive process always ends in finite time. Since the applies to equilibria from any history onwards, the bad outcome assumption is, in terms of equilibrium behavior, equivalent to *assuming* that all strategies the DM can employ from any history onwards end in finite time. Thus our problem is to answer whether there is a cognitive strategy that is *consistent* with equilibrium reasoning and the assumption that an outcome is always implemented in finite time.

Discussion on the one-deviation assumption and the initial choice x_0

In the standard modelling practice, the one-deviation property is typically derived from a more primitive solution concept that does not put limitations on players' ability to deviate. The same applies to the current framework in the transitive preferences case: the one-deviation property would be implied by the cognitive equilibrium concept alone. However, when the preferences are not transitive, the equivalence no longer holds, and we has to take the one-deviation assumption as the starting point.[6]

One way to motivate the one-deviation property is to appeal on computability constraints. Infinite deviations may require unreasonably large computational resources.

Another way to interpret the one-deviation assumption is to assume instead that the decision process reflects *all* the cognitive activity of the DM. For if an infinite deviation to a strategy was possible for the DM, then he must be able to associate cognitive actions to infinitely many histories. But then there would be no reason for him not to be able to construct a full strategy as well. This would open the question of how to choose a strategy in the first place. Since picking up a strategy should be no different from picking up an alternative that the strategy implements, there should, conceivably, be a deeper decision process for that purpose. But this would be in conflict with the idea that the original decision process reflects all the cognitive activity of the DM.[7] Thus the one-deviation assumption can be seen as a natural way to restrict the DM from making self-referential questions.

Similar considerations do not, however, concern the choice of the initial element x_0. For one thing, the exogenously given initial outcome can be motivated on heuristic grounds. One natural choice for the initial outcome x_0 would be the status quo outcome e, i.e. that no decision is made. Depending the scenario at hand, however, other alternatives could also serve as the initial point: With the grocery shopper's case, the initial good in the basket might be the one that shopper first encounters; with the club's case, the minimally binding rule system, given the general law; etc..

[6] Which does not mean that the one-deviation assumption has not been the starting assumption in the standard modelling practice.

[7] Lipman (1991) develops an alternative approach. He shows that "choosing how to choose..." process has a meaningful solution when the DM is boundedly rational.

For another thing, nothing essential would change in the analysis would the initial element be chosen by the DM himself. Dropping the initial outcome would, however, require one to specify a distinct equilibrium condition for time 1, as the stopping option would no longer be available for the DM. To see why such change would not affect the equilbrium, suppose that $x_0 = e$ in the current model. Then, by Corollary 4, a cognitive equilibrium necessarily entails the DM not stopping the game at time 0. Since the behavior at time 1 would be effectively constrained by the same constraints in the model with $x_0 = e$ and in the one without x_0, the equilibrium strategy of the former could be transferred to the latter without complications. The current modelling choice is mainly for convenience.

3 Simple Cognitive Equilibria

Consider a memoryless strategy where the current cognitive action is dependent only on the alternative on the table. Call such strategy *simple*. A simple cognitive equilibrium may not exist: consider the 3-cycle $x \succ y$, $y \succ z$, $z \succ x$ on $X = \{x, y, z\}$. Suppose that, say, x is implemented in a simple cognitive equilibrium. Then proposing y cannot be profitable when x is on the table. Hence y cannot be implemented which means that z must be, since choosing y is not profitable. But z cannot be an equilibrium since moving from z to x is profitable, and feasible since x is implementable in the equilibrium.

Theorem 5 *An alternative is implemented in a simple cognitive equilibrium if and only if it is the $\succ$ −maximal element.*

Proof. The if-part: Let $x^* \succeq x$ for all $x \in X$. Construct strategy $s : H \to X$ such that $s(h) = x^*$ for all h. Since $\pi_s[h] = x^*$ for all h, a deviation cannot be profitable.

For the only if-part, let s be a simple cognitive equilibrium. Hence a function $s : X \to X$ fully characterizes the cognitive strategy. Let set $Y \subseteq X$ consist of all alternatives y that are implementable in s, i.e. all y's such that $s(y) = \mathsf{stop}$. Suppose that $x, x' \in Y$. Since the deviation from x to x' is not profitable and vice versa, $x' \succeq x$ and $x \succeq x'$. By the asymmetry of $\succ$, $x = x'$ and hence Y is a

singleton set, say $\{x\}$. Thus, by the definition of π_s, $\pi_s[(h, z)] = x$, for all $h \in H$. By (4), $x \succeq z$, for all z. Thus x is the $\succ$ −maximal element. ■

Thus simple procedures do not take us too far as the ability to make a decision requires a maximal element. Together with the condition of *context-freeness,* i.e. that a decision is makeable in every nonempty subset Y of X, this implies that $\succ$ has to be transitive. Conversely, any transitive $\succ$ induces a simple, context-free decision.

Corollary 6 *A simple and context-free cognitive equilibrium exists if and only if* $\succ$ *is transitive.*

For closely related analyses, see Salant (2004) and Rubinstein and Salant (2005).

4 General Characterization

To characterize choices implementable in a cognitive equilibrium, we define the following key concept.

Definition 7 *A nonempty set* $C \subseteq X$ *is a* consistent choice set *if* $x \in C$ *and* $y \succ x$*, for any* $y \in X$*, implies* $z \succ y$ *and* $x \succ z$*, for some* $z \in C$*.*

That is, for any x in a consistent choice set, if y is preferred to x, then there is z *in* the consistent choice set such that (x, y, z) forms a cycle (see Figure 2, where $y \leftarrow x$ reads $y \succ x$, etc.). This implies, by the completeness of $\succ$, that any alternative *not* in a consistent choice set C is preferred by *some* element in C. For if $y \notin C$ would not dominate $x \in C$, then, by the completeness of $\succ$, necessarily $y \succ x$ and, by Definition 7, there is $z \in C$ such that $z \succ y$.

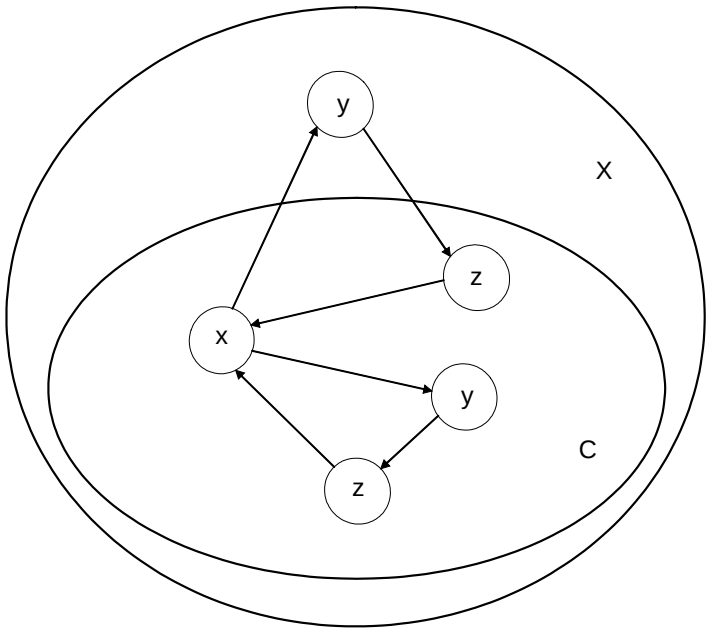

Figure 2.

This means that a consistent choice set coincides with the (unique) maximal element whenever such element exists. Whenever it does *not* exist, a consistent choice set contains at least three elements (apply the definition to any two y z in C), and is a strongly connected component of X.[8]

Now we characterize cognitive equilibria through the concept of consistent choice set. First we construct an equilibrium that implements outcomes in a consistent choice set C. First, define the (weak) *lower contour set* at x :

$$L(x) = \{y \in X : x \succeq y\}.$$

Fix a consistent choice set C. Let us describe the strategy by a quadruple (s, τ, P, p_0), where P is a finite set of "phases" that partitions H, function $\tau : P \times X \to P$ is a transition rule between phases, function $s : P \times X \to X$ is the strategy that is conditional only on the outcome of the table and the current

[8]There is a directed path from any element in X to any other element in X.

phase, and $p_0 \in P$ is the initial phase.[9] [10] Let

$$P = \{p_x : x \in C\}. \tag{5}$$

That is, the elements of P are indexed by the outcomes in C.

By Definition 7, there exists a function $z : X \times X \to X$ such that if $x \in C$ and $y \notin L(x) \cap C$, then[11]

$$z(x, y) \in C \cap [L(x) \setminus L(y)]. \tag{6}$$

Given the function z, let the transition rule τ satisfy

$$\tau(p_x, y) = \begin{cases} p_y, & \text{if } y \in L(x) \cap C, \\ p_{z(x,y)}, & \text{if } y \notin L(x) \cap C. \end{cases}$$

Finally, given the function z, let the strategy s satisfy

$$s(p_x, y) = \begin{cases} \mathsf{stop}, & \text{if } y \in L(x) \cap C, \\ z(x, y), & \text{if } y \notin L(x) \cap C. \end{cases}$$

That is, starting from any $(p_x, y) \in P$ it takes at most one period to implement a decision according to strategy (s, τ, P, p_0). For if $y \in L(x) \cap C$, then y is implemented now. If $y \notin L(x) \cap C$, then $z(x, y)$ becomes the outcome on the table and the phase switches to $p_{z(x,y)}$. Hence $z(x, y) \in L(z(x, y)) \cap C$ is implemented in the next period. Note that in either case, the implemented outcome belongs to $L(x) \cap C$. The proof below relies on this property of (s, τ, P, p_0).

That strategy (s, τ, P, p_0) forms an equilibrium is based on the following intuition. In phase p_x, the strategy implements a certain "self-punishing" outcome, say z, in $L(x) \cap C$. By construction, a deviation from the path that implements z changes the phase to p_z. This would trigger a self-punishment relative to z, by implementing an outcome in $L(z) \cap C$. The fear of self-punishment relative to z provides sufficient incentives to self-punish relative to x by implementing z. This

[9] In other words, the constructed cognitive strategy is implementable by a finite automaton. For more on finite automata in games, see Rubinstein (1986).

[10] To see that P is a partition of H, construct $p_x \subset H$ recursively: Let the empty history belong to p_0. History $(h, y) \in H$ belongs to p_x if $\tau(p, y) = p_x$ and $h \in p$, for some $p \in P$.

[11] If $y \notin L(x)$, then $C \cap L(x) \setminus L(y)$ is nonempty by Def. 7. If $y \in L(x) \setminus C$, then $x \in C \cap L(x) \setminus L(y)$.

circularity in self-punishments eventually makes the strategy self-sustaining. Such construction is feasible due to characteristics of a consistent choice set.

Lemma 8 (s, τ, P, p_0) *forms a cognitive equilibrium.*

Proof. Take any $(p_x, y) \in P \times X$. It suffices to show that a one-time deviation d from $s(p_x, y)$ is not profitable. A notation convention: Since P is a partition of H, and all relevant information of a h is contained in a p such that $h \in p$, we will replace all hs with the corresponding ps in what follows.

First, noting that $s(p_d, d)$ implements d for any $d \in C$ and applying s twice we have

$$\pi_s[(p_x, y)] = \begin{cases} y, & \text{if } y \in L(x) \cap C, \\ z(x, y), & \text{if } y \notin L(x) \cap C. \end{cases}$$

Thus, by (6),

$$x \succeq \pi_s[(p_x, y)], \text{ for all } y \in X. \tag{7}$$

To check that a deviation d from $s(p_x, y)$ is not profitable, it suffices to go through the two cases.

1. Let $y \in L(x) \cap C$. Then $s(p_x, y) = \texttt{stop}$, and

$$\pi_s[(p_x, y)] = y.$$

A deviation $d = w \in X$ changes the phase to $\tau(p_x, y) = p_y$, and implements

$$\pi_s[(p_x, y, w)] = \pi_s[(p_y, w)]. \tag{8}$$

Applying (7) to $\pi_s[(p_y, w)]$,

$$y \succeq \pi_s[(p_y, w)].$$

Thus, by (8), $y \succeq \pi_s[(p_x, y, w)]$, implying that the deviation is not beneficial.

2. Let $y \notin L(x) \cap C$. Then $s(p_x, y) = z(x, y)$ and

$$\pi_s[(p_x, y)] = z(x, y).$$

A deviation $d = \texttt{stop}$ implements y. By the construction of $z(\cdot, \cdot)$, $z(x, y) \succeq y$, thus the deviation is not beneficial. A deviation $d = w \in X \setminus \{z(x, y)\}$ changes

the phase to $\tau(p_x, y) = p_{z(x,y)}$, and implements

$$\pi_s[(p_x, y, w)] = \pi_s\left(p_{z(x,y)}, w\right). \tag{9}$$

Applying (7) to $\pi_s\left(p_{z(x,y)}, w\right)$,

$$z(x, y) \succeq \pi_s\left(p_{z(x,y)}, w\right),$$

Thus, by (9), $z(x, y) \succeq \pi_s[(p_x, y, w)]$ implying that the deviation is not profitable. ■

The next lemma establishes that the converse of Lemma 8 holds, too.

Lemma 9 *Let s be a cognitive equilibrium, and let $Y \subseteq X$ be the set of outcomes that can be implemented with it: $Y = \{x : s(h, x) = x \text{ s.t. } h \in H\}$. Then Y is a consistent choice set.*

Proof. We show that Y satisfies Definition 7. Suppose that $x \in Y$, and $y \succ x$ for any $y \in X$. Then, since $x \in Y$, and since a deviation from s cannot be profitable, $\pi_s[(h, x, y)] \neq y$ for any $h \in H$ such that $s(h, x) = \texttt{stop}$. Since y is not implemented were it on the table, there is $z \in Y$, $z \neq y$, such that $z = \pi_s[(h, x, y)]$. We show that (x, y, z) forms a cycle. By (4), $z \succeq y$. Since $z \neq y$, in fact $z \succ y$. By Definition 3, $\pi_s[(h, x)] \succeq \pi_s[(h, x, y)]$, and by hypothesis $\pi_s[(h, x)] = x$. Thus $x \succeq z$. If $z = x$, then $x \succ y$ contradicting the initial hypothesis. Thus, by completeness of preferences, $x \succ z$ as desired. ■

By Lemmata 8 and 9, there is 1-1 relationship between consistent choice sets and sets of alternatives that are implementable in cognitive equilibria.

Theorem 10 *A set Y of alternatives is implementable in a cognitive equilibrium if and only if Y is a consistent choice set.*

Thus physical features of cognitive equilibria are completely characterized by consistent choice sets.

5 Existence and Uniqueness

Now we discuss the connections of cognitive equilibria to some well-known concepts, and establish the key properties of consistent choice sets. Given $B \subseteq X$, we say that y *covers* x in B, or simply $B-$covers, if $x, y \in B$, $x \neq y$, and $z \succ y$ implies $z \succ x$, for all $z \in B$. Note that if y covers x then $y \succ x$. Denote by $uc(B)$ the *uncovered* set of B, i.e. the set of alternatives in B that are not $B-$covered (Fishburn, 1977; Miller, 1980). That is, for any element x in $uc(B)$ and y in B such that $y \succ x$ there is z in B such that x, y, and z form a 3-cycle.

The uncovered set is uniquely defined and it always exists. The uncovered set plays an important role in voting theory (see e.g. Laslier, 1997).

Consider the following extension of the uncovered set, by Dutta (1988). Set $D \subseteq X$ is a *covering set* if $uc(D) = D$ and $x \notin uc(D \cup \{x\})$, for any $x \in X \setminus D$. That is, any element in D is not D-covered by any element in D, and any element x not in D is $D \cup \{x\}$-covered by some element in D.[12]

Theorem 11 *Any covering set is a consistent choice set.*

Proof. Let D be a Dutta covering set. Let $x \succ y$ and $y \in D$. We prove that there is $z \in D$ such that (x, y, z) forms a 3-cycle.

If $x \in D$, then the result follows since y is not $D-$covered by x.

If $x \notin D$, then, by the definition of D, there is $w \in D$ that $D-$covers x, i.e. $w \succ x$, and there is *no* $z' \in D$ such that $x \succ z'$ and $z' \succ w$. In particular, because $x \succ y$, it cannot be the case that $y \succ w$. Since $x \succ y$, $w \succ x$, and $\succ$ is asymmetric, $y \neq w$. Since $\succ$ is complete, $w \succ y$. By this, and since y is not $D-$covered, there exists *some* $z \in D$ such that (w, y, z) forms a cycle (see Fig. 3 where $x \leftarrow y$ reads $x \succ y$, etc.). Hence, $y \succ z$ for such z. Since $z \succ w$ it follows, by the first sentence of this paragraph, that $z \succ x$. By construction, $x \succ y$, and hence (x, y, z) forms a cycle. ■

[12] In other words, if x is in D and $y \succ x$ for some y in D, then there is z in D such that x, y, and z form a 3-cycle, and if x is *not* in D, then there is y in Z such that $y \succ x$ and such that then there is *no* z in Z such that x, y, and z form a 3-cycle.

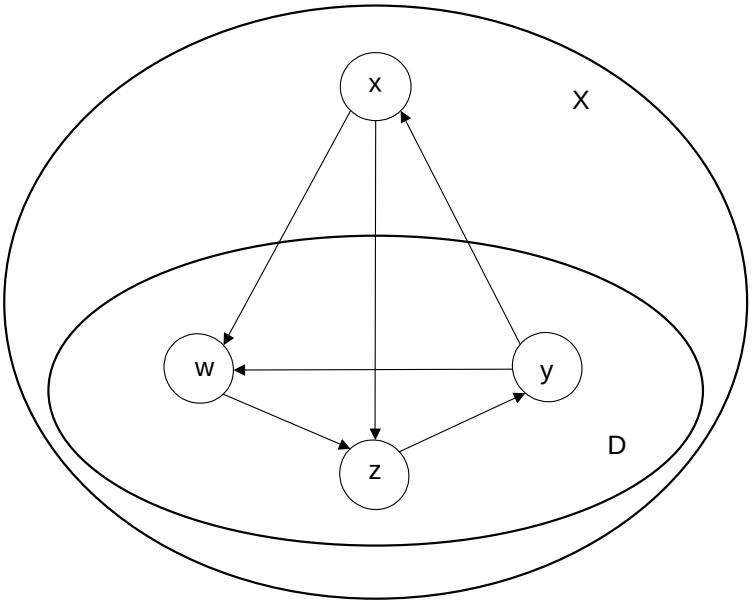

Figure 3.

The Dutta covering set D requires that any element x *outside* D is $D-$covered while the consistent choice set C only requires that an element y *inside* C is *not* $C \cup \{x\}-$covered for any x. Together with the assumption that no two elements in the set cover each other, the former property implies the latter whenever $x \succ y$. However, the converse is not true (for an example see Section 7).

Dutta (1988) shows that the notion of covering set is linked to the iterated version of the uncovered set, the *ultimate* uncovered set (UUC) (originally discussed by Miller, 1980; see also Laslier, 1998). The UUC is defined recursively as follows. Let $uc^{k+1}(X) = uc(uc^k(X))$, for all $k = 0, ...$, and $uc^0(X) = X$. Then $uc^\infty(X)$ is the UUC. Since X is finite, no element in $uc^\infty(X)$ is $uc^\infty(X)-$covered, i.e. every arc spanned by $\succ$ in $uc^\infty(X)$ is an edge a 3-cycle *in* this set.

Due to its recursive structure, the UUC is uniquely defined. Importantly, Dutta (1988, Theorem 1) shows that the *UUC is a covering set.* Thus, by Theorem 11, we have the existence result.[13]

[13] An independent existence proof is available from this author.

Theorem 12 *A consistent choice set exists.*

This implies, by Theorem 10, that a cognitive equilibrium exists. We state this as a result.

Corollary 13 *A cognitive equilibrium exists.*

In fact, Dutta (1988) shows that the UUC is the maximal covering set. We now establish the corresponding result with consistent choice sets. Since a consistent choice set need not be a covering set, this is *not* implied by Dutta (1988).

It is clear that a union of two consistent choice sets is also a consistent choice set. Hence, by finiteness of X the *existence* of the maximal consistent choice set is clear. To prove that the UUC is the maximal set, we need to show that no element removed in an iterative round can be contained by a consistent choice set.

Theorem 14 *The UUC is the maximal consistent choice set.*

Proof. Let C be a consistent choice set. Since $uc^{\infty}(X)$ is a Dutta covering set, and any such set is a consistent choice set by Theorem 11, we only need to show that C does not contain an element not in $uc^{\infty}(X)$. Recall that if $x \in C$ and $y \succ x$ for some $y \in X$, then there is $z \in X$ such that (x, y, z) form a cycle.

If $C = \{x\}$, then Definition 7 implies that x is the maximal element. But then it is also uncovered and hence $uc^{\infty}(X) = \{x\}$.

If C contains more than one element, then, by Definition 7, any element $x \in C$ cannot be $X-$covered by $y \in X$ since then there would not be $z \in C$ that forms a cycle with x and y. Denote by D_0 the $X-uncovered$ elements. Then $C \subseteq X \setminus D_0$. Any element $x \in C$ cannot be $X \setminus D_0-$covered by $y \in X \setminus D_0$ since then there would not be $z \in C \subseteq X \setminus D_0$ that forms a cycle with x and y. Then $C \subseteq X \setminus \cup_{j=0,1} D_j$. Continuing this way, $C \subseteq X \setminus \cup_{j=0,1,...} D_j$. Since $X \setminus \cup_{j=0}^{\infty} D_j = uc^{\infty}(X)$, the result follows. ■

By this and Theorem 10 implies that, in the language of the implementation theory, the UUC is implemented in cognitive equilibrium.

Corollary 15 *An alternative is implementable in a cognitive equilibrium if and only if it belongs to the UUC.*

6 Stable Decisions

An immediate objection against the cognitive procedure analyzed in the previous sections is that the procedure assumes a specific correspondence between cognitive actions and physical actions, captured by the function π_s. That is, the outcome x is implemented after history (h, x) if the chosen cognitive action is `stop`. Needless to say this is only one possible way cognitive actions could transform into physical consequences. An important question is then how sensitive is the model to the specific assumptions of the underlying physical structure?

As such, the concept of a cognitive equilibrium does not require any specific form of π_s but it *does* require that some such function exists (any non-cooperative equilibrium concept relies on a game form). Without a cognitive structure, one would not be able to compare the consequences of cognitive acts. In this section, we present a model that suggests that the predictions are robust to the details of π_s.

We now do not assume a specific structure between cognitive and physical actions, but we do assume that such correspondence exists. The relation is not modelled. However, we now assume that the hardware of the DM consists not only of preferences but also of a framework where cognitive activity takes place. The hardware is called a *cognitive machine* (the "brain"). What is new is that a cognitive machine specifies a set of "real" states of mind, and a transition rule from one state to another as a function of cognitive choices. A cognitive choice represents an idea or a tentative plan of which outcome will *eventually* be implemented, without taking any stand on *how* it will be implemented (not even whether this happens in finite time). The states are part of the physical information processing machinery but they do not have any payoff relevance.

While the DM is no longer programmed to a particular cognitive strategy s, he is not completely free to choose his cognitive actions: he is bounded to be *consistent*. The consistency conditions, which are inspired by the stability criteria á la von Neumann-Morgenstern (see also Dutta, 1988), are defined with respect to the DM's hardware.[14] We ask (i) does there exist a machine that induces vNM stable cognitive behavior (to be defined below), and (ii) which physical actions are

[14] Which has, perhaps, emerged through evolution.

inducible with such cognitive machine.

Cognitive Machine Denote a *cognitive machine* by $\mu = (\lambda, Q)$. A machine consists of a set of states Q and a transition function $\lambda : X \times Q \to X$. The role of a machine is to read finite strings of the vocabulary X. The way μ processes information is simple. Given a state $q_0 \in Q$, a string $x_0, x_1, ...$ induces a string $q_0, q_1, ...$ of states such that $\lambda(x_k, q_k) = q_{k+1}$ for all $k = 0, 1, ...$ The machine does not implement an outcome, it just specifies a transition rule over the "states of mind", which may guide the cognitive process.

Preferences $\succ$ span an *indirect* dominance relation, denoted by $\succeq_\mu$, over $X \times Q$ as follows: $(x', q') \succeq_\mu (x', q)$ if there are sequences $x_0, ..., x_K \in X$ and $q_0, ..., q_K \in Q$ such that $\lambda(x_k, q_k) = q_{k+1}$ and $x_K \succ x_k$ for all $k = 0, ..., K-1$, and such that $(x', q') = (x_0, q_0)$ and $(x, q) = (x_K, q_K)$.

That is, (x, q) indirectly dominates (x', q') if one can move from (x', q') to (x, q) by manipulating the cognitive states via outcome choices in a way that x is always preferred to the outcomes in the middle. If such conditions materialize, then the DM is willing to move along the cognitive path from (x', q') to (x, q).

Indirect dominance reflects farsightedness in domination: The deviant looks to the end of the domination chain to see whether the deviation is profitable. The notion of indirect dominance was introduced by Harsanyi (1974), and analyzed in a general framework by Chwe (1992).

If the length of the dominance chain is $K = 1$, then the dominance is said to be *direct.* Since $\succ$ may not exhibit a maximal element, there need not be an undominated (and hence indirectly undominated) element. A natural weakening of the undominance criterion is the following.

Definition 16 (Stability) *A cognitive machine $\mu = (\lambda, Q)$ induces an (indirectly)* stable set $V \subset X \times Q$ *if the following hold:*

1. *(External Stability) If $(x, q) \notin V$, then there is $(x', q') \in V$ s.t. $(x, q) \succeq_\mu (x', q')$.*

2. *(Internal Stability) If $(x, q) \in V$, then there is* no *$(x', q') \in V$ s.t. $(x, q) \succeq_\mu (x', q')$.*

Now we establish a result that is analogous to Lemma 9..

Lemma 17 *V is a stable set is induced by a cognitive machine only if $\{y : (y,q) \in V\}$ is a consistent choice set.*

Proof. Let V be induced by $\mu = (\lambda, Q)$. We show that $\{y : (y,q) \in V\}$ meets Definition 7. Take $x \in \{y : (y,q) \in V\}$. Identify $q \in Q$ such that $(x,q) \in V$. Suppose that there is x' such that $x' \succ x$.By internal stability, $(x',q') \notin V$, for $q' = \lambda(x,q)$. By external stability, there is $(x'',q'') \in V$ such that $(x',q') \ \beta_\mu \ (x'',q'')$. By the definition of indirect dominance, $x'' \succ x'$. By internal stability, $(x,q) \not\beta_\mu (x'',q'')$. Since $\lambda(x,q) = q'$, it must be, by completeness of $\succ$, that $x \succ x''$, as desired. ■

Indirect dominance can be defined for any inducement correspondence á la β. In general, the existence of an indirectly stable set is not guaranteed (see Chwe, 1992). However, we now show that in our context the existence is not a problem.

Let us construct a cognitive machine that induces a stable set. Fix a consistent choice set C. Let

$$Q = \{q_x : x \in C\}, \tag{10}$$

and let the transition function λ satisfy

$$\lambda(q_x, y) = \begin{cases} q_y, & \text{if } y \in L(x) \cap C, \\ q_x, & \text{if } y \notin L(x) \cap C. \end{cases} \tag{11}$$

The working of cognitive machine $\mu = (\lambda, Q)$ is based on the by now familiar logic. The idea is to get "trapped" to a state, say q_x, that leads to implementation of an outcome in $L(x) \cap C$. Implementation an outcome y in $L(x) \cap C$ is self-sustaining since not doing so would only lead to implementation an outcome in $L(y) \cap C$.

Lemma 18 *The cognitive machine $\mu = (\lambda, Q)$ as defined in (10) and (11) induces a stable set V such that $\{y : (y,q_x) \in V\} = C$.*

Proof. Construct $V = \{(y,q_x) : y \in C \cap L(x), x \in C\}$. Since $y \in C \cap L(x)$ for some $x \in C$ only if $y \in C \cap \{y\}$, we have $\{y : (y,q_x) \in V\} = \{y : y \in C \cap L(x), x \in C\} = \{y : y \in C \cap \{y\}\} = C$. It suffices to show that V is a stable set.

External stability: Take any (y, q_x) such that $y \notin C \cap L(x)$. Then there is $z(x, y)$ as defined in (6) such that $z(x, y) \in C \cap L(x) \setminus L(y)$. By the construction of V, $(z(x, y), q_x) \in V$. Since $\lambda(y, q_x) = q_x$, we have $(y, q_x) \ \succ_\mu \ (z(x, y), q_x)$.

Internal stability: Take any (y, q_x) such that $y \in C \cap L(x)$. Then $(y, q_x) \in V$. Suppose that $(z, q) \ \succ_\mu \ (y, q_x)$, for any $(q, z) \in Q \times X$. By the definition of indirect dominance, there are sequences $x^0, ..., x^K \in X$ and $q^0, ..., q^K \in Q$ such that $\lambda(x^k, q^k) = q^{k+1}$ and $y \succ x^k$ for all $k = 0, ..., K-1$, and such that $(z, q) = (x^0, q^0)$ and $(y, q_x) = (x^K, q^K)$. There are two cases. (i) Suppose that $q^0 = ... = q^K = q_x$. Then, since there is no transition away from state q_x, it follows by the construction of λ that $x^k \notin L(x) \cap C$ for all $k = 0, ..., K-1$. In particular, $z \notin L(x) \cap C$. Since $q_x = q$, this implies that $(z, q) \notin V$. (ii) Let $k \leq K$ be the highest integer such that $q^k \neq q_x$. Since there is no transition away from state q_x after k, it must be, by the construction of λ, that $x = x^k$. Since $y \succ x^k$ also $y \succ x$. But this contradicts the initial assumption $y \in C \cap L(x)$. ■

As in the previous section, let us say that the set of outcomes B of alternatives is implementable in a stable set of a cognitive machine if there is a cognitive machine that induces a stable set V such that $B = \{y : (y, q) \in V\}$.

By Lemmata 17 and 18 we can now state the analogues of Theorem 10 and Corollary 15.

Theorem 19 *A set B of alternatives is implementable in a stable set of a cognitive machine if and only if B is a consistent choice set.*

Corollary 20 *An alternative can be implemented within a stable set of a cognitive machine if and only if it belongs to the UUC.*

Thus cognitive machines implement the same set of physical decisions than cognitive strategies. Since the concept of cognitive machine does not require one to specify the underlying physical structure, the results suggest that outcomes that are implementable in cognitive equilibrium are not sensitive to the details of the model. This can also be verified by considering the cognitive process that implements an outcome on the table only if it has remained there for $n \leq \infty$ stages. It is not difficult to prove that the results of Section 4 would remain unchanged. Changing the assumptions concerning the cognitive process does not seem to have

physical consequences as long as the cognitive process is rich enough. What seems to be relevant is that before implementing an outcome, the DM always has an option to veto the alternative on the table; when considering to implement an outcome, he should be committed to doing that. We conjecture that any decision making procedure having this feature implements the same cognitive equilibria.

While the concepts of cognitive equilibrium and stable cognitive machines look different at the outset, they have, of course, common underpinnings. The fact that the vNM stability concept is more parsimonious suggests that this concept is more "primitive". The position that external and internal stability capture in reduced form essential features of strategic thinking is powerfully exposed in Greenberg (1990).

7 Properties and Examples

A *choice function* f associates a nonempty subset of outcomes to each decision problem $B \subseteq X$. The natural interpretation of a choice function is that it chooses outcomes that are induced by cognitive equilibrium. What properties do such choice functions satisfy?

The *independence of irrelevant alternatives* (IIA) is of specific interest:[15] if an element $x \in f(X)$ is chosen under X, and $x \in B \subset X$, then $x \in f(B)$.

It is clear that no single valued choice function satisfies IIA (try cycle on $X = \{x, y, z\}$). Thus the choice function has to be multivalued. A natural candidate is to associate the choice function to a specific consistent choice set.

There many ways to extend the IIA to multivalued choice functions. Consider Sen's property β: If $B \subseteq X$, then $f(B) \subseteq f(X)$ or $f(X) \cap f(B)$ is nonempty. That is, if both x and y are chosen in B, then only one of them cannot be chosen in X that contains B. If this condition holds for f associating to a consistent choice set, then removing an element in a consistent choice set cannot force an element outside the set to become a member. The next example demonstrates that this cannot hold (in all figures, $x \leftarrow y$ reads $x \succ y$, etc.).

Example 21 *Let $X = \{x, y, z, u\}$ and preferences as depicted in Figure 4. Then*

[15] The weak axiom of revealed preference or Chernoff condition (Sen's property α) (cf. Arrow, 1959).

u is covered by x and the unique consistent choice set is $\{x, y, z\}$.

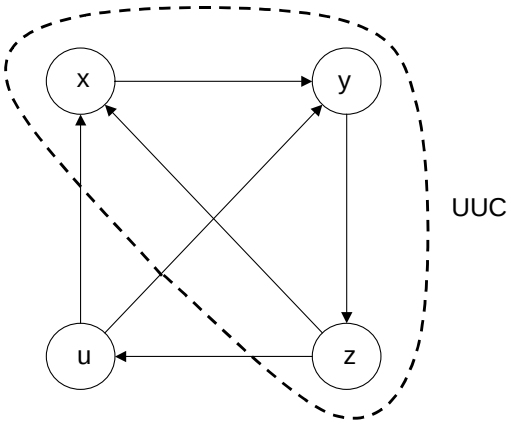

Figure 4.

However, the unique consistent choice set of $X \setminus \{x\}$ is $\{y, z, u\}$, not a subset of $\{x, y, z\}$.

However, there are choice functions that meet a weaker IIA condition, the *strong superset property* (cf. Laslier, 1997):[16] If $B \subseteq X$ and $f(X) \subseteq B \subseteq X$, then $f(X) = f(B)$. To see this, note that if C is a consistent choice set in X and $C \subseteq B \subseteq X$, then C is a consistent choice set in B.

More specifically, which choice functions meet the strong superset property? A natural candidate is the union of the consistent choice sets, i.e. the UUC. Unfortunately, the UUC does not meet the strong superset property. When moving to a smaller set, a new consistent choice set may emerge that enlarges the UUC, as is demonstrated in the next example.

Example 22 *Let $X = \{x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3\}$. Preferences are described in Figure 5a (non-depicted arrows are down) Now y_i covers x_i. After removing all x_i's, y_i is covered by z_i. Thus the UUC is $\{z_1, z_2, z_3\}$. Consider now $X \setminus \{y_1, y_2, y_3\} = \{x_1, x_2, x_3, z_1, z_2, z_3\}$ (see Figure 5b). No element is covered and hence the UUC is $\{x_1, x_2, x_3, z_1, z_2, z_3\}$.*

[16] This is a strengthening of Sen's property α.

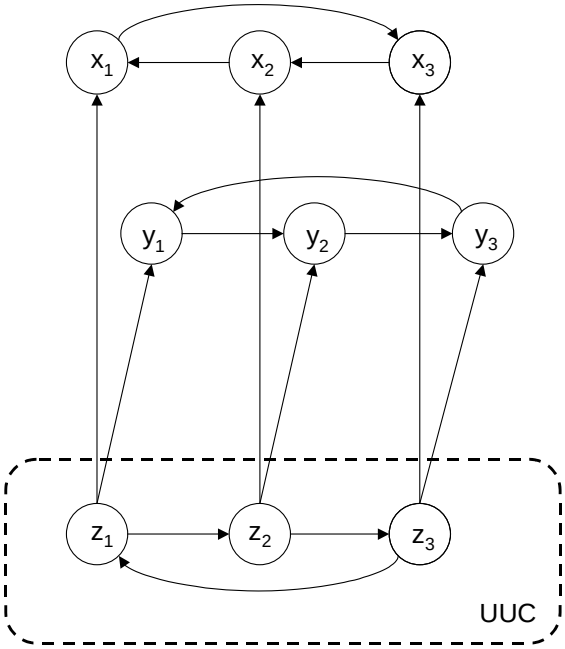

Figure 5a.

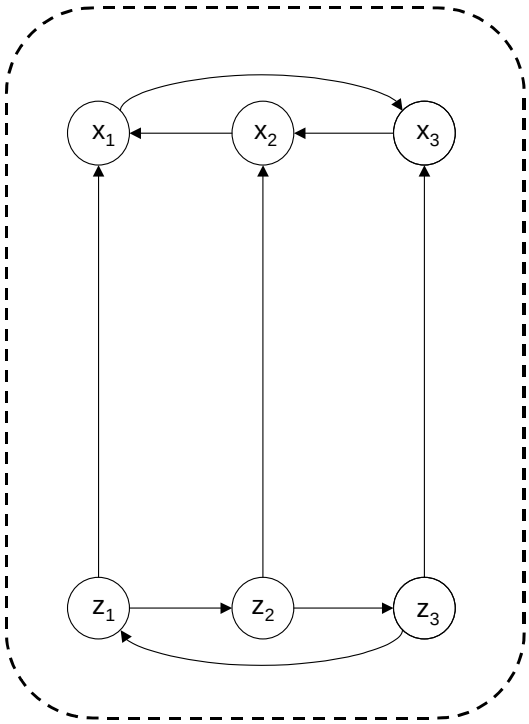

Figure 5b.

The fact that the UUC correspondence does not satisfy the IIA is surprising given that the uncovered set itself *does* meet the condition (see e.g. Laslier, 1997). However, as discussed above, there are correspondences that pick from the class of consistent choice sets that meet the strong superset property. The *minimal covering set* (MC), whose eixstence nd uniqueness is shown by Dutta (1988), satisfies the strong superset property. Laslier (1997) demonstrates that MC is also monotonic and independent of the losers, unlike the UUC. Since the MC (as well as the UUC) is also Condorcet consistent, it is a natural candidate choice correspondence. The question is whether MC is also a minimal consistent choice set. Unfortunately, it is not.

Example 23 *Consider regular*[17] *preferences on set* $Y = \{y_1, ..., y_5\}$*, as depicted in Figure 6. Then* $uc(Y) = Y$*. Adding* z *with preferences as depicted by the shaded arrows, we have* $uc(Y \cup \{z\}) = Y \cup \{z\}$*. Thus* Y *is not a Dutta covering set in*

[17] Whose nodes have equal outdegree.

$Y \cup \{z\}$. *However,* Y *is a consistent choice set in* $Y \cup \{z\}$.

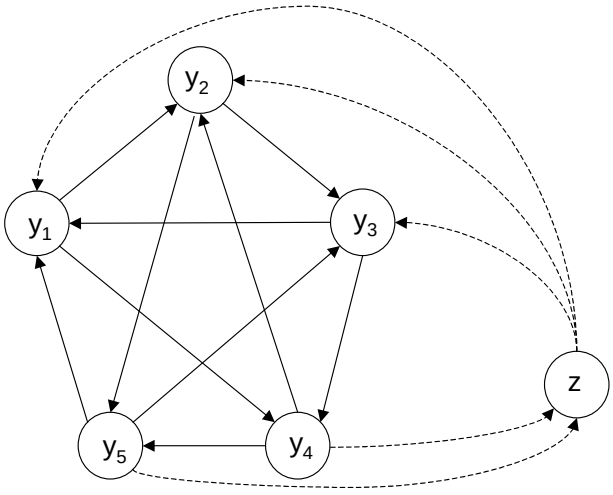

Figure 6.

Since MC may not be a minimal consistent choice set, a natural question is whether a unique minimal consistent choice set exists. It does not, as the next example demonstrates. The example exhibits two consistent choice sets with *no* common elements.

Example 24 *Consider choice set* $\{x_1, ..., x_6, y_1, ..., y_2\}$. *Assume* $uc(\{x_1, ..., x_6\}) = \{x_1, ..., x_6\}$ *and* $uc(\{y_1, ..., y_6\}) = \{y_1, ..., y_6\}$ *(cf. Fig 4b). Let preference between* x_i *and* y_j, $i, j = 1, ..., 6$, *be as in Figure 7 (all non-depicted arrows are down).*

Then both $C_x = \{x_1, ..., x_6\}$ *and* $C_y = \{y_1, ..., y_6\}$ *are consistent choice sets.*

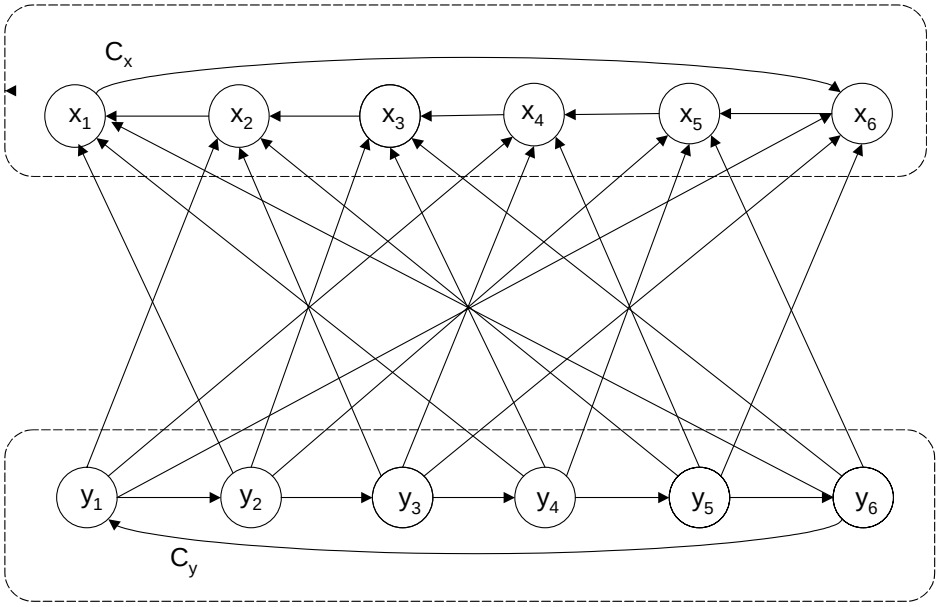

Figure 7.

Whenever a consistent choice set does not consist of a single element, it contains at least three. How to choose from them? An appealing method would be to choose from *all* consistent choice sets at the same time. However, since the intersection of consistent choice sets may be empty, as the previous example demonstrates, this method is groundless.

Changing the direction of a preference between two alternatives can have a peculiar impact on the unique consistent choice set.

Example 25 *(Moulin, 1986) Let a complete and transitive* $\succ$ *impose an order* $x_0, ..., x_n$ *on the set* X*. The unique consistent choice set* C *under* $\succ$ *is* $\{x_0\}$*. Switch the direction of preferences between* x_0 *and* x_n*, and let* $\succ'$ *differ from* $\succ$ *only in how pair* (x_0, x_n) *is ordered, i.e.* $x_n \succ' x_0$ *and* $x_k \succ' x_l$ *for all* $k < l$ *such that* $(k, l) \neq (0, n)$*. Now the unique consistent choice set* C' *under* $\succ'$ *is* $\{x_0, x_1, x_n\}$*.*

This is depicted in Figure 10, for $n = 4$.

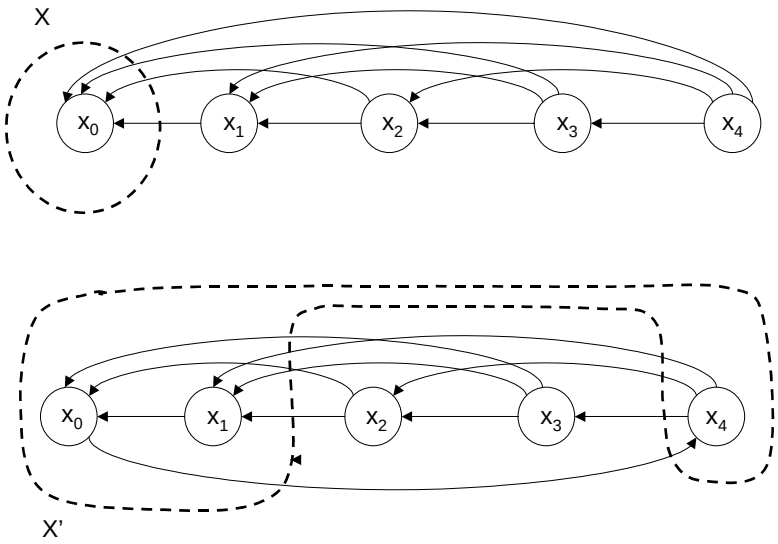

Figure 8.

Thus an alternative with a lowest score (indegree) may belong to the unique consistent choice set. Ranking the alternatives according to their scores is not a useful method to identify a consistent choice set; one has to focus on the topology of the graph.

An Extension The results assume that the binary relation under consideration is a complete and asymmetric. Asymmetry can be relaxed in an obvious way. Consider a complete binary relation $\succeq$ on X. Since $\succeq$ is complete (but not necessarily asymmetric) it contains a complete and symmetric subrelation $\succ$. Now the notion of cognitive equilibrium can be defined with respect to $\succ$. However, interpreting the symmetric part of $\succeq$ as an indifference relation also changes the interpretation of the solution. If the solution is defined with respect to $\succ$, then certain actions may be needed solely for the equilibrium purposes a proposed equilibrium path may be needed even when DM is indifferent in following it or implementing the outcome in the table. Moreover, since there are many ways to derive $\succ$ from $\succeq$, one can, in general, associate many ultimate uncovered sets to $\succeq$. Thus the choice correspondence is no longer uniquely defined.

8 Literature

The cognitive psychology framework roots from the metaphor that the brain is a computer and the mind its software.[18] Partly due to the recent methodological controversy around behavioral economics, the the interplay between preferences and information processing has been analyzed also in the economics literature. The recurrent theme there is that transitivity of preferences allows *economical* decision making procedures (see Rubinstein, 1996, 2000). The theme of this paper is complementary: We ask *how* should the DM reason to be able to make a decision under? We have argued that non-transitivity as such does not prevent decision making. However, complex strategies cannot be avoided outside the transitive preferences -paradigm.

Kalai et al. (2002) study rationalization of choice functions by multiple rationales. They analyze context dependent preference structures that are economical in a sense that they explain the observed data with the least number of orderings. Any behavior can be rationalized with multiple rationales. However, to the contrast, not all choice functions are consistent with behavior in cognitive equilibrium. Consider choices in the set $\{x, y, z\}$. Then choice x under $\{x, y, z\}$, y under $\{x, y\}$, and z under $\{x, z\}$ cannot be generated in any cognitive equilibrium since the latter two imply the existence of a maximal element in $\{x, y, z\}$ that is distinct from x, and the first means that there cannot be a maximal element distinct from x (Theorem 5). Hence such data cannot be explained by non-transitive preferences whereas it can be explained by multiple rationales.

Rubinstein and Salant (2005) analyze rationalization of choices that are made from an *exogenously* given list of alternatives. They characterize reasonable choice procedures, and show that a version of independence axiom *alone* implies maximization (or minimization) of an ordering. Thus independence, sequential choice, and the avoidance of the worst outcome imply that DM is a maximizer. Salant (2005) studies the closely related question of what kind of preferences permit computationally economical decision making. His main result is that the most economical choice functions, i.e. ones which require the least amount of memory, are rationalizable by transitive preferences. Otherwise, more memory is needed.

[18] For philosophical underpinnings, see Dennet (1991) or Binmore (1994). Speigler (2002, 2004) are recent applications to economics.

There is a wide literature on closely related issue of money-pump: a non-transitive DM can be exploited (and hence presumably leave the market) if he cannot commit to a choice (see e.g. Machina, 1989). The key observation of this paper is that a DM *can* effectively commit to a choice. Hence it cognitive equilibrium is a shield against the money-pump. In equilibrium, reversing an offer in favor of a better candidate only leads to the implementation of an outcome that is even worse than the original outcome. Thus the DM can commit to not reversing his decision.

Mandler (2005) makes a connected point by allowing an agent to condition his current physical choice on his past physical choices. In Mandler's model (psychological) preferences over a set of outcomes are transitive but not necessarily complete. He defines an indirect dominance relation based on these preferences, and shows that they may induce indirectly undominated choice behavior that is observationally non-transitive.

Preference reversal phenomenon in the context of risky choices is perhaps the best known manifestation of non-transitive choice behavior. Regretful sensation is a common explanation (cf. Loomes and Sudgen 1982; Fishburn, 1982). Regret theory answers successfully *why* non-transitive choice behavior can be observed in pairwise comparisons, but it puts less emphasis into the question of *how* decisions should be made in a three-or-more alternatives set up. Answer to the latter question, which is the theme of this paper, requires extra assumptions. The standard procedure to achieve a decision is to guarantee the existence of a maximal element. Loomes and Sudgen (1987) do this by assuming a context dependent behavioral rule and Fishburn (1985) through a domain assumption. Quiggin (1994) gives a sufficient condition for non-cyclic choice.

Non-transitivity of preferences can be thought to mirror multi-dimensional decision making criteria. The analogue to the social choice literature is clear. From the perspective of this paper, the most relevant part of this vast literature is the analysis of agenda formation and strategic voting. An agenda can be thought as a list of alternatives of which the society votes in a sequential order. Shepsle and Weingast (1984) show that any sophisticated equilibrium of such voting game implements an outcome in the uncovered set.[19] The exact form of the agenda affects

[19] See also Banks (1985).

the equilibrium outcome. Dutta et al. (2001) analyze endogenous formation of finite agendas. Finiteness of the agenda guarantees that the solution is well defined. In a companion paper Vartiainen (2005) shows that endogenous *unbounded* agenda formation leads to the implementation of an outcome in a consistent choice set (defined with respect to the tournament that is spanned by the underlying voting game).

9 Conclusion

We show that a cognitive equilibrium exists under all complete preferences, and characterize outcomes that can be implemented within it. This suggests that (procedural) rationality does not imply any restrictions on (complete) preferences, if the decision making ability of the DM is viewed as the criterion. In this sense, the many axioms imposed on behavior in decision theory are genuinely independent from the concept of rationality. From this viewpoint, the axioms should be judged on different grounds, e.g. on how useful they are in modeling exercises or, possibly, how well they fit to our introspective view of a good axiom.

It is tempting to interprete a "cognitive phase", used to construct equilibrium stratgies, as an emotion. While this is far-fetched in the current simple model, they do bear some nice emotion-like features. On the one hand, the role of phases in our cognitive equilibrium construction is to filter out the relevant information of the massive amount of data that is hidden in all possible histories. This is line with the position that emotion's role is to serve as an information processing device (as a "frame") (see e.g. Cohen, 2006). One the other hand, if the strategies are assumed to be implemeted via automata like constructions, as in Section 4, then the cognitive phase story would *necessitate* multiple emotions: complex decisions could not be made without more than one cognitive phases (Section 3).[20] Moreover, from the phase = emotions -point of view one could not dictate emotions; they would emerge in equilibrium. These observations should not be inconsistent with everyday life.

There are interesting directions for future research. As discussed in the previous section, not all choice functions can be rationalized with non-transitive pref-

[20] For a similarly grounded motivation, see e.g. Frank (1988).

erences and cognitive equilibria. Exactly which choice functions can be generated by a single preference profile in a cognitive equilibria is an apt question. A choice function that is based on transitive preferences is easy to compute (see Salant, 2005). How difficult it is to compute a choice function (correspondence) that associates to a consistent choice set in all subproblems? As pointed out by Dutta (1988), identifying the minimal covering set can be computationally hard. Perhaps this is the genuine source of our tendency to focus on transitive preferences. But since the whole revealed preference methodology relies on the assumption of transitive preferences, understanding exactly what it means is important.

.

References

[1] Arrow, K. (1959), Rational choice functions and orderings, *Econometrica* 26, 121-27.

[2] Banks, J (1985), Sophisticated voting outcomes and agenda control, *Social Choice and Welfare* 1, 295-306

[3] Binmore, K. (1994), *Game theory and social contract, vol. 1: Playing fair*, MIT Press, Cambridge MA.

[4] Chwe, M. (1994), Farsighted stability, *Journal of Economic Theory* 63, 299-325.

[5] Cohen, J. (2005), The vulcanization of the human brain, A neural perspective of interactions between cognition and emotion, *Journal of Economic Perspectives* 19, 3-25.

[6] Dennet, D. (1991), Consciousness Explained, *Penguin Press*

[7] Dutta. B., Jackson, M., and M. LeBreton (2001), Equilibrium Agenda Formation, *Social Choice and Welfare* 23, 21-37.

[8] Dutta. B (1988), Covering sets and a new condorcet correpondence, *Journal of Economic Theory* 44, 63-80.

[9] Duggan, J. (2004), Systematic approach to the construction of non-empty choice set, manuscript, University of Rochester.

[10] Duggan, J. and LeBreton, M. (1996), Dutta's minimal covering set and shapley's Saddles, *Journal of Economic Theory* 70, 257-65.

[11] Fishburn, P. (1977), Condorcet social choice function, *SIAM Journal of Applied Math* 33, 295-306.

[12] Fishburn, P. (1982), Non-transitive measurable utility, J*ournal of Mathematical Psychology* 21, 191-218.

[13] Frank, R. (1988), *Passions within reason: the strategic role of emotions*. New York: Norton.

[14] Greenberg, J. (1990), *The theory of social situations*, Cambridge University Press, UK

[15] Harsanyi, J. (1974), An equilibrium point interpretation of stable sets and a proposed alternative definition, *Management Science* 20, 1427-95.

[16] Kalai, G., Rubinstein, A., and R. Spiegler (2002), Rationalizing choice functions by multiple rationales, *Econometrica* 70, 2481-8.

[17] Laslier, J.-F. (1997), *Tournament solutions and majority voting*, Springer-Verlag, Berlin.

[18] Lipman, B. (1991), How to decide how to decide how to ... : Modeling limited rationality, *Econometrica* 59, 1105-25

[19] Loomes, G. and Sugden, R. (1982), Regret theory: an alternative theory of rational choice under uncertainty, *Economic Journal* 92, 805-24.

[20] Loomes, G. and Sugden, R. (1983), A rationale for preference reversal, *American Economic Review* 73, 428-432.

[21] Miller, N. (1980), A new solution set to for tournaments and majority voting, *American Journal of Political Science* 24, 68-96.

[22] Moulin, H. (1986), Choosing from a tournament, *Social Choice and Welfare* 3, 271-91.

[23] Rubinstein, A (1996), Finite Automata Play the Repeated Prisoner's Dilemma, *Journal of Economic Theory* 39, 83-96.

[24] Rubinstein, A (1996), Why are certain properties of binary relations relatively more common in natural language, *Econometrica* 64, 343-55.

[25] Rubinstein, A. (2000), *Economics and Language*, Cambridge UP.

[26] Rubinstein, A. and Salant Y (2006), A model of choice from lists, *Theoretical Economics 1*, 1-17.

[27] Salant, Y. (2004), Limited computational resources favor rationality, manuscript, Stanford University.

[28] Shepsle, K. and B. Weingast (1984), Uncovered sets and sophisticated outcomes with implications for agenda institutions, *American Journal of Political Science* 28, 49-74.

[29] Spiegler, R. (2002), Equilibrium in Justifiable Strategies, *Review of Economic Studies* **69**, 691-706.

[30] Spiegler, R. (2004), Simplicity of Beliefs and Delay Tactics in a Concession Game, *Games and Economic Behavior* **47**, 200-20.

[31] Quiggin, J. (1994), Regret theory with general choice sets, *Journal of Risk and Uncertainty* 8, 153-65.