Incorporating Risk in Choice Theory: Some Observations

Geoffroy de Clippel

May 2020

Abstract

This paper provides a warning: a property in the spirit of the sure-thing principle that may sound intuitive at first, and indeed standard in classic models, is systematically violated when considering choices that cannot be obtained through the maximization of a preference ordering. Beyond choice theory, this observation also has relevant implications for game theory, social choice, and mechanism design.

1 Introduction

This paper provides a warning: a property in the spirit of the sure-thing principle that may sound intuitive at first, and indeed standard in classic models, is systematically violated when considering choices that cannot be obtained through the maximization of a preference ordering. This matters for at least two reasons. First, advances in behavioral economics highlight how individual choices can systematically violate rationality. Second, even if individuals are rational, decisions are oftentimes made by groups (e.g., households, committees, countries, etc.), and rationality need not be preserved. The warning is not only of concern to choice theorists, but also has relevant implications for game theory, social choice, and mechanism design.

We start with a couple of examples to provide some preliminary insight.

Example 1. *A card will be drawn from a deck of black and red cards. A couple won a contest, and can pick one of three bets to determine their prize. Under bet 1, they win a wireless phone charger if the card is black, but nothing*

otherwise. Under bet 2, they win the wireless phone charger if the card is red, but nothing otherwise. Under bet 3, they win a nice bottle of wine whatever the card color. Tables below summarize the situation, and present the two spouses' utility functions.

	Black	*Red*
Bet 1	*Charger*	$\emptyset$
Bet 2	$\emptyset$	*Charger*
Bet 3	*Wine Bottle*	*Wine Bottle*

	u_1	u_2
Charger	1	3
Wine Bottle	3	1
$\emptyset$	0	0

Which bet will the couple choose? They agree to proceed as follows: spouse 1 *eliminates one bet, and spouse* 2 *selects her favorite among the two surviving options. Notice that the couple selects bet* 3 *when they know the card is black, as well as when they know the card is red. In other words, the couple would select bet* 3 *ex-post whatever the color that comes up. In the spirit of the sure-thing principle, one might then expect that the couple would pick bet* 3 *whatever the relative proportion of black cards in the deck. But this is wrong. For instance, the couple selects bet 1 when the proportion of black cards is strictly between* 1/2 *and* 2/3.

While the previous example pertained to group choices, the next example looks at a shortlisting method in the spirit of Manzini and Mariotti (2007) to capture individual choices that need not be rational.

Example 2. *As in the previous example, a black or red card will be selected at random from a deck. This time an individual decides between two bets with prizes in a set* $\{x, y, z\}$*. The decision-maker maximizes her preference over the set of bets that are Pareto efficient according to two preliminary selection criteria. Tables below depict the bets as well as the individual's preference* (u) *and selection criteria* (s_1 *and* s_2)*.*

	Black	*Red*
Bet 1	x	y
Bet 2	z	z

	s_1	s_2	u
x	0	3	2
y	3	0	2
z	1	1	3

He picks bet 2 *if he knows the state is black (*x *and* z *are not Pareto comparable according to criteria* s_1 *and* s_2*, and he prefers* z *over* x*). Similarly, he picks bet* 2 *if he knows the state is red. Since he'd pick bet* 2 *whatever the card color, one may expect him, in the spirit of the sure-thing principle, to*

pick that same bet whatever the proportion p of black cards in the deck. But this is wrong again. Applying expected utility, we see that only bet 1 *will be in the individual's shortlist when p is strictly between* $1/3$ *and* $2/3$.

The sure-thing principle is quite standard. Indeed, it does hold quite pervasively under the rational benchmark model (e.g., for any preference ordering consistent with first-order stochastic dominance). We formalize a related, choice-based property in the spirit of the above examples, see Property P below. In view of the multitude of collective choice rules (e.g., Borda, egalitarianism, plurality, tournaments, (relative) utilitarianism, etc.), and the multitude of individual choice functions capturing a variety of behavioral biases, is behavior in Examples 1 and 2 a peculiarity of the rules studied there? Contributions at the intersection of behavioral economics and choice theory often focus on problems involving deterministic outcomes. Thus another, related question arises in that context: are there extensions of these individual choice functions to the larger domain of choice over risky prospects that respect a choice-based version of the sure-thing principle?

These questions are addressed in Section 2. We will see in particular that Examples 1 and 2 are the rule rather than the exception when it comes to the variety of choice functions one might consider. Indeed, similar issues arise for any 'welfarist' social choice rule (Theorem 2), and any choice function violating of rationality on deterministic outcomes (Theorem 1). But our choice-based variant of the sure-thing principle can be compatible with violations of rationality provided that rationality is preserved over deterministic outcomes. Considering choices over monetary lotteries, for instance, any choice function (rational or not) that never selects a lottery that is strictly first-order stochastically dominated by a feasible alternative satisfies Property P (Theorem 3).

Beyond choice theory, our observations have implications in game theory and mechanism design. Under rationality, a player's strategy is dominant if it is a best response whatever its belief about its opponents' actions. But optimality is typically checked only against opponents' *pure* strategies. This is fine when restricting attention to expected utility (or any preference ordering satisfying first-order stochastic dominance). Overlooking probabilistic beliefs is invalid when accommodating violations of rationality (Theorem 4). In mechanism design, we will encouter social choice rules that may seem strategy-proof because participants would select truth-telling whatever their opponents' reports, but are not implementable in dominant strategy because

players are uncertain about others' reports. Looking at the case of serial dictatorship, we will point out that a mechanism designer should favor the dynamic allocation procedure over its static variant because it eliminates uncertainty regarding available items when players make their choices. Finally, we will see that the notion of ex-post equilibrium can fail to produce a robust solution to behavioral games of incomplete information: while a strategy profile may be a Nash equilibrium in choices (as in de Clippel (2014)) in each ex-post game of complete information, uncertainty about others' types (and hence actions) may lead a player to unilateraly deviate.

To summarize, one should exercise caution, and avoid misleading intuition gleaned from experience with more standard frameworks when incorporating bounded rationality or group choices in models involving risky prospects.

2 Framework and Main Results

Let Ω be a (finite) set of *states of the world*, and O be a (finite) set of relevant *outcomes*. An *act* is a map that associates an outcome to each state of the world. A *lottery* is a probability distribution over O. A *choice function* c associates to each (finite) set L of lotteries a subset $c(L)$. Since outcomes are degenerate lotteries, a choice function also defines choices over subsets of O. Though arguments hold more generally, the restriction of c on subsets of O is assumed to be single-valued. This restriction is *rational* if there exists a preference ordering $\succ$ on O such that $c(S) = \arg\max_{\succ} S$, for each subset S of O. The states' relative likelihoods are captured by a probability distribution $p \in \Delta(\Omega)$. Assuming state-independence, as we do throughout the paper, means that only lotteries associated to acts matter to the decision-maker. Given any finite set A of acts, let $L^p(A)$ be the set of lotteries $\ell^p(a)$ – "$a(\omega)$ obtains with probability $p(\omega)$" – obtained by varying $a \in A$.

In the spirit of the sure-thing principle, we investigate the following property on choice functions over risky prospects:

Property P *Let A be a set of acts, and let $a \in A$. If $a(\omega) = c(A(\omega))$ for each $\omega \in \Omega$, then $c(L^p(A)) = \{\ell^p(a)\}$ for all $p \in \Delta(\Omega)$.*

If the decision-maker knows that the state is ω, then picking an act amounts to choosing an outcome within $A(\omega)$. Now suppose that the act a has the unique property of delivering her chosen outcome in *each* state ω. This is a very stringent property that often does not apply as an act may provide the

chosen outcome is some state but rarely in all states. Also, notice that at most one act can have that property since c is single-valued over subsets of outcomes. Given state-independence, picking an act from A given p amounts to picking a lottery from $L^p(A)$. Property P requires that lottery to be $\ell^p(a)$, indeed the one associated to act a. At first this seems reasonable because a delivers the outcome he wants to pick whatever the state realization. Yet this property is violated as soon as the restriction of c to deterministic outcomes is not rational. In particular, none of the many choice functions over deterministic outcomes discussed in the recent literature at the intersection of behavioral economics and choice theory extend to problems involving risk while satisfying our choice-based analogue of the sure-thing principle.

Theorem 1. *If c satisfies Property* 1*, then it is rational over deterministic outcomes.*

Proof. If c is not rational, then there exist a choice problem $T \subseteq X$ and $x \in T$ distinct from $c(T)$ such that $c(T) \neq c(T \setminus \{x\})$. Let Ω' be a nonempty strict subset of Ω, and consider the following acts:

	$\omega \in \Omega'$	$\omega \in \Omega \setminus \Omega'$
a	$c(T)$	$c(T \setminus \{x\})$
a'	$c(T \setminus \{c(T)\})$	$c(T \setminus \{x\})$
a''	$c(T \setminus \{x\})$	$c(T)$
a'''	x	$c(T)$
a_y	y	y

for each $y \in T \setminus \{x, c(T), c(T \setminus \{x\})\}$ (if any). Let A be the set of all acts appearing on the table. Then $A(\omega) = T$ for each $\omega \in \Omega'$, and $A(\omega') = T \setminus \{x\}$ for each $\omega \in \Omega \setminus \Omega'$. Let $p \in \Delta_{++}(\Omega)$ be such that $p(\Omega) = p(\Omega \setminus \Omega') = 1/2$. By Property P, $c(L^p(A)) = \ell^p(a) = \frac{1}{2}c(T) \oplus \frac{1}{2}c(T \setminus \{x\})$. Let $\hat{A} = A \setminus \{a\}$. Then $\hat{A}(\omega) = T \setminus \{c(T)\}$ for each $\omega \in \Omega'$, and $\hat{A}(\omega) = T \setminus \{x\}$ for each $\omega \in \Omega \setminus \Omega'$. By Property P, $c(L^p(\hat{A})) = \ell^p(a') = \frac{1}{2}c(T \setminus \{c(T)\}) \oplus \frac{1}{2}c(T \setminus \{x\})$. A contradiction arises then from the fact that $L^p(A) = L^p(\hat{A})$ and $c(T) \neq c(T \setminus \{c(T)\})$. □

Comparing Property P with Savage's Sure-Thing Principle The sure-thing principle was defined by Savage in the context of choice under *uncertainty*. A first point of departure is that we take beliefs (either objective or subjective) as given. A second point of departures is that Savage's condition applies to all partitions of Ω, while no restriction is imposed under Property

P when that partition is not fully revealing the state. More importantly, Savage assumes at the outset that individuals are rational, while we explore a choice-based version of the property that applies to any choice function. Interestingly, Savage (1972, page 39)'s justification for the sure-thing principle is in fact phrased in terms of choices, not preferences.

> *A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant to the attractiveness of the purchase. So, to clarify the matter for himself, he asks whether he would buy if he knew that the Republican candidate were going to win, and decides that he would do so. Similarly, he considers whether he would buy if he knew that the Democratic candidate were going to win, and again finds that he would do so. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains, or will obtain, as we would ordinarily say.*

Since preference maximization and choices are one and the same in Savage's framework, he could have written this justification either way. But it is perhaps telling that the story does sound reasonable expressed in choices and seemingly independently of how these choices are made. Yet the present paper shows that, contrary to intuition, the property is most often violated without preference maximization.

Necessity of Rationality for Utility-Based Social Choice Having established that there is no way to extend irrational choice functions over deterministic outcomes to risky pospects while satisfying Property P, we investigate in this subsection and the next whether there are irrational choice functions over risky prospects satisfying Property P (which thus would have to be rational over deterministic outcomes). The answer is negative when performing utility-based social choice.

Most of social choice theory is welfarist, meaning that judgements are conducted by relying on realized utilities at available options. Say the universal set of options in a given problem is O, and i's utility function over lotteries is given by $u_i : \Delta(O) \to \mathbb{R}$.[1] Let $u = (u_1, \ldots, u_n)$ be the profile of such utility functions. The social choice rule $r_{(O,u)}$, associating a nonempty subset to any

[1]It could be expected utility, but it does not have to be; any function will do.

set $L \subseteq \Delta(O)$ of lotteries,[2] could depend on these primitives, but a further simplification is usually made. One starts from a choice function c defined over finite subsets of utility profiles, compute the set

$$u(L) = \{(u_1(\ell), \ldots, u_n(\ell)) | \ell \in L\}$$

of utility profiles achievable via lotteries in L, and then select from L all the lotteries that generates utility vectors in $c(u(L))$. The rule in Example 1, and the many rules mentioned at the bottom of page 2 proceed this way. Indeed, most of social choice theory is welfarist this way. To formalize this notion of welfarism, we think of a social choice rule as being first defined for any problem (O, u) where O is a finite set of outcomes, and u is a profile of utility functions (one for each individual in N) defined on O.

Definition 1. *The social choice rule r is* utility-based *if there exists a choice function c associating to each finite set S of utility profile a nonempty subset $c(S) \subseteq S$ such that for all problems (O, u) and all finite sets of lotteries $L \subseteq \Delta(O)$, $r_{(O,u)}(L) = \{\ell \in L | u(\ell) \in c(u(L))\}$.*

Theorem 2. *Suppose that r is utility-based and satisfies Property P, that is, $r_{(O,u)}$ satisfies Property P for all problems (O, u). Then r is rational, that is, $r_{(O,u)}$ is rational for all problems (O, u).*

Proof. Let c be the choice function associated to r in Definition 1. For any two utility vector x, y in $\mathbb{R}^N$, say

$$x \succ y \text{ if } c(\{x, y\}) = \{x\}.$$

We start by proving that $\succ$ is an ordering and that

$$c(\mathcal{U}) = \arg\max_{\succ} U, \tag{1}$$

for each finite subset $\mathcal{U}$ of $\mathbb{R}^N$. It is easy to check that $\succ$ is complete, since $r_{(O,u)}$ is single-valued over sets of deterministic outcomes, whatever the combination (O, u). Next, for transitivity, consider three utility vectors x, y, z such that $x \succ y$ and $y \succ z$. Let then $O = \{a, b, d\}$, $u(a) = x$, $u(b) = y$

[2] As before, $r_{(O,u)}$ is assumed to be single-valued when L contains only deterministic outcomes.

and $u(d) = z$. Since $r_{(O,u)}$ satisfies Property 1, there exists by Theorem 1 an ordering $\succ_{(O,u)}$ on $\{a, b, c\}$ such that

$$r_{(O,u)}(S) = \arg\max_{\succ_{(O,u)}} S,$$

for all subset S of $\{a, b, d\}$. By definition of c, $r_{(O,u)}(\{a, b\}) = c(\{x, y\})$. Hence, $a \succ_{(O,u)} b$ if, and only if, $x \succ y$. Similarly, $a \succ_{(O,u)} d$ if, and only if, $x \succ z$, and $b \succ_{(O,u)} d$ if, and only if, $y \succ z$. Since $x \succ y$ and $y \succ z$, we have that $a \succ_{(O,u)} b$ and $b \succ_{(O,u)} d$. By transitivity of $\succ_{(O,u)}$, it follows that $a \succ_{(O,u)} d$, which in turn implies that $x \succ z$, as desired for transitivity of $\succ$. As for (1), let $\mathcal{U} = \{x_1, \ldots, x_{|\mathcal{U}|}\}$ be a finite subset of $\mathbb{R}^N$, let now $O = \{a_1, \ldots, a_{|\mathcal{U}|}\}$ and $(u_1, \ldots, u_n)$ be utility functions such that $u(a_k) = x_k$, for each $k = 1, \ldots, |\mathcal{U}|$. By a similar reasoning as above, we know by Theorem 1 that there exists an ordering $\succ_{(O,u)}$ on O such that

$$r_{(O,u)}(S) = \arg\max_{\succ_{(O,u)}} S, \tag{2}$$

for all $S \subseteq O$, and

$$a \succ_{(O,u)} a' \Leftrightarrow u(a) \succ u(a'), \tag{3}$$

for all $a, a' \in O$. We have:

$$c(\mathcal{U}) = u(r_{(O,u)}(O)) = u(\arg\max_{\succ_{(O,u)}} O) = \arg\max_{\succ}\{u(a)|a \in O\} = \arg\max_{\succ} U,$$

where he first equality corresponds to Definition 1, the second equality follows from (2), and the third one from (3).

Now fix any problem (O, u) and any two lotteries $\ell, \ell' \in \Delta(O)$. Say that $\ell \sim^* \ell'$ if $u(\ell) = u(\ell')$, and $\ell \succ^* \ell'$ if $u(\ell) \succ u(\ell')$. It is easy to check that $\succeq^*$ is complete and transitive, since $\succ$ is an ordering on $\mathbb{R}^N$. Finally, let $L \subseteq \Delta(O)$ be any finite set of lotteries. Then

$$r_{(O,u)}(L) = \{\ell \in L|u(\ell) \in c(u(L))\} = \{\ell \in L|u(\ell) \in \arg\max_{\succ} u(L)\} = \arg\max_{\succeq^*} L,$$

where the first equality corresponds to Definition 1, the second equality follows from the first part of the proof, and the last equality follows from the definition of $\succeq^*$. □

Notice, for instance, how the social choice rule defined in Example 1 is in fact rational over the deterministic outcomes $\emptyset$, Charger, Wine Bottle. But

the procedure applied to other utility profiles is irrational, which can be seen for instance when considering choice over lotteries in that example: when choosing between the three bets with 3/5 of black cards, the couple selects the first bet; but they pick the third bet instead when dropping bet 2, an irrelevant alternative.

Satisfying Property 1 without Rationality Theorem 1 establishes that c must be rational over deterministic outcomes. Are there irrational choice functions satisfying Property 1 while their restriction over deterministic outcomes is rational? For concreteness, suppose that outcomes are monetary payoffs and that choices are obtained by payoff maximization in the absence of risk.

Even then, Property 1 need not be satisfied. Consider, for instance, a "*cautious investor comparing alternative portfolios first eliminates those that are too risky relative to others available, and then ranks the surviving ones on the basis of expected returns*" (Manzini and Mariotti (2007, page 1825)). To formalize this, suppose that the investor computes each lottery's coefficient of variation (standard deviation divided by the mean), and eliminates those with a coefficient that is strictly above average (within the set of available lotteries). Consider then a first investment paying $\$1,500$ whatever the state, and a second investment paying $\$1,510$ if the state belongs to $\emptyset \neq \Omega' \subset \Omega$ and $\$1,520$ otherwise. While the investor is rational over deterministic outcomes, Property 1 is violated since the investor overlooks the second investment whenever he places strictly positive probability on both Ω' and its complement. As the next result shows, this violation of Property 1 is attributable to a violation of first-order stochastic dominance, or more precisely its extension to choice functions that need not be rational.

Definition 2. *The choice function c is* consistent with first-order stochastic dominance *if, for each set of lotteries L over monetary amounts, there is no lottery $\ell \in L$ that first-order stochastically strictly dominates $c(L)$.*

Theorem 3. *If c is consistent with first-order stochastic dominance, then c satisfies Property* 1*.*

Proof. Let A be a set of acts, and let $a \in A$. Suppose that $a(\omega)$ is the maximal monetary payoff in $A(\omega)$, for each $\omega \in \Omega$. Then, for all $p \in \Delta(\Omega)$, $\ell^p(a)$ first-order stochastically strictly dominates all lotteries in $L^p(A)$ distinct from $\ell^p(a)$, and hence $c(L^p(A)) = \{\ell^p(a)\}$, as desired. □

The above theorem means that we can find many examples of irrational choice functions over monetary lotteries satisfying Property 1. For any choice function c, let $\hat{c}$ be the modified choice function that selects, from any set L of monetary lotteries, the lottery $c(\hat{L})$, where $\hat{L}$ is the subset of lotteries that are not first-order strictly dominated by an alternative in L. It follows from the above result that $\hat{c}$ satisfies Property 1.

3 Beyond Choice Theory

Dominant Strategies Switching to interactive decision-making, a *rational strategic-form game* specifies for each player $i \in N$ a finite set S_i of strategies, a preference ordering $\succ_i$ over outcomes in O, and an outcome function $f : S \to O$ where $S = \times_{i \in N} S_i$ is the set of strategy profiles. A standard notion is that of a dominant strategy, oftentimes defined as follows: strategy s_i^* is *dominant* for player i if $f(s_i^*, s_{-i}) \succ_i f(s)$ for all $s \in S$ such that $f(s_i^*, s_{-i}) \neq f(s)$.

As discussed throughout the paper, players may use choice functions that are not compatible with preference maximization, either because of behavioral biases or because players are groups instead of individuals. The notion of a game easily extends: a *behavioral strategic-form game* is obtained simply by changing each i's preference ordering $\succ_i$ by her choice function c_i in the definition. It is tempting then to consider the following definition of dominant strategy. As it will appear inadequate (too weak), strategies satisfying it will be called "seemingly dominant."

Definition 3. *Strategy s_i^* is* seemingly dominant *for i if*

$$c_i(\{f(s)|s_i \in S_i\}) = \{f(s_i^*, s_{-i})\},$$

for all $s_{-i} \in S_{-i}$.

If i expects others to pick s_{-i}, then the opportunity set of outcomes she faces when picking her own strategy is $\{f(s)|s_i \in S_i\}$. Suppose she'd pick o from that set, which happens to be precisely the outcome she gets when picking s_i^*. For s_i^* to be seemingly dominant, this property must hold whatever s_{-i}. This is the natural extension of the property of dominance one checks in rational games.

But players' beliefs about others' strategies need not be deterministic. For a strategy to be dominant, the desired property is really that it would be

selected whatever the player's belief about her opponents' strategies. This is inconsequential in standard rational games (that is, with expected utility or any preference consistent with first-order stochastic dominance), as optimality against each pure strategy guarantees optimality againt opponents' correlated strategies. Given Theorem 1, however, one should be suspicious that seemingly dominant strategies need not be dominant when taking into account that players' beliefs need not be deterministic. We start by introducing some notations: given a belief $p \in \Delta(S_{-i})$, let $\ell^p(s_i)$ be the lottery that selects $f(s)$ with probability $p(s_{-i})$ and let $L_i^p = \{\ell^p(s_i) | s_i \in S_i\}$ be the opportunity set of lotteries that i faces when picking her strategy.

Definition 4. *Strategy s_i^* is* dominant *for i if $c_i(L_i^p) = \{\ell^p(s_i^*)\}$ for all $p \in \Delta(S_{-i})$.*

In games, one can think of opponents' strategy combinations as states of the world. Then, the proof of Theorem 1 can be adapted to show that a lack of rationality over deterministic outcomes necessarily breaks down the justification for paying attention only to opponents' pure strategies. Instead, one must use the full definition of a dominant strategy, considering any belief over opponents' correlated strategies.

Theorem 4. *Let i be a player. If c is not rational over deterministic outcomes, then there exists a behavioral strategic-form game where i's choice function is c and i has a seemingly dominant strategy which is not dominant.*

Proof. Suppose, by contradiction, that in all games, a seemingly dominant strategy for player i must be dominant. Consider then a first two-player game G derived from the table in the proof of Theorem 1, where $S_i = \{a, a', a'', a''', a_y\}$, the opponent's strategy set is $\{\Omega', \Omega \setminus \Omega'\}$, and the outcome function is defined as in the table (replacing c by c_i). Following the argument in the proof of Theorem 1, a is seemingly dominant for player i in G. By our hypothesis, it is dominant as well, and hence $c_i(L^p) = \{\ell^p(a)\}$ for all mixed-strategy p of i's opponent. Consider now the game $\hat{G}$ derived from G by eliminating a from S_i. Following the argument in the proof of Theorem 1, a' is seemingly dominant for player i in $\hat{G}$. By our hypothesis, it is dominant as well, and hence $c_i(\hat{L}^p) = \{\hat{\ell}^p(a')\}$ for all mixed-strategy p of i's opponent. As in the proof of Theorem 1, $L^p = \hat{L}^p$ and $\ell^p(a) \neq \hat{\ell}^p(a')$, hence the contradiction. □

Mechanism Design Though typically defined for rational individuals, *serial dictatorship* extends naturally to any allocation problem where players are endowed with choice functions. Player 1 moves first, and is free to pick any one item in a set X; she picks $c_1(X)$. Player 2 moves next, and is free to pick any one item from among those that remain; she picks $c_2(X \setminus \{c_1(X)\})$. The procedure goes on like that until everyone has received an item, or all items have been allocated.

This dynamic allocation method performs well even in more realistic circumstances where players' choice functions are their private information. Formally, it implements by backward induction the *serial dictatorship social choice rule*, r_{SD}, which associates to any profile of choice functions the allocation one would obtain by running serial dictatorship with reported choice functions. In fact, players are not even required to form expectations about, let alone correctly anticipate, future actions since the item a player gets is independent of subsequent players' choices. Having observed past moves – the choices of players with lower indices – a player faces no uncertainty about her opportunity set of items when making a choice.

In the special case of rational players, this naturally implies that reporting one's true choice functions is a dominant strategy in the associated strategic-form of the dynamic serial dictatorship game. Put it differently, r_{SD} is strategy-proof and truth-telling is a dominant strategy in the direct mechanism defined by r_{SD}. Suppose now players may have more complex choice functions. Truth-telling remains an adequate choice whatever others' reports. Indeed, a player's opportunity set, obtained by varying its report, is the set of items not allocated to players with smaller indices based on their fixed reports. One could think that r_{SD} is thus implementable in dominant strategy over any domain of choice functions.

But this is wrong because one has not considered situations where players are unsure about others' reports. As should be clear by now, a willingness to report the truth against deterministic reports need not survive when considering probabilistic beliefs. The next example illustrates the issue. With a single player corresponding to a group whose choices might be irrational, serial dictatorship ceases to be implementable in dominant strategies (not using the direct mechanism, nor any other mechanism!).

Example 3. *There are four items,* $X = \{a, b, d, e\}$*, and four players,* $N = \{1, 2, 3, 4\}$ *(one item each). Each player other than* 2 *is rational. Her type encodes her strict Bernoulli function over* X*; all orderings are possible. Player*

2, *on the other hand, is the couple we encountered in Example 1. Its type encodes their choice function over lotteries of items. For extreme clarity, let's make a very small departure from rationality. The two spouses may fully agree on how to rank items and lotteries, in which case their joint choices following Example 1's procedure is rational. Allowing for any common ranking, we get a set of types comparable to those for players* 1, 3 *and* 4. *There is, however, one additional type to consider where the spouses' preferences disagree. For that type* t_2^*, *Bernoulli utilities are*

	u_1	u_2
a	1	1
b	2	2
d	5	3
e	3	6

A mechanism simply specifies for each individual i *a finite set* M_i *of possible* messages *and an outcome function* $g : M \to O$ *where* $M = \times_{i \in N} M_i$ *is the set of message profiles and* O *is the set of allocations of items to individuals. Suppose there is such a mechanism implementing* r_{SD} *in dominant strategies, with* i*'s dominant strategy* m_i^* *associating to each type* t_i *a message* $m_i^*(t_i)$. *Let's consider in particular a type* t_1^a *which ranks* a *top, and a type* t_1^b *which ranks* b *top. For implementation,* g *has to allocate* d *to player* 2 *for the message combination* $(m_1^*(t_1^a), m_2^*(t_2^*), \cdot, \cdot)$ *(whatever players* 3 *and* 4*'s messages). Indeed, player* 1 *must get item* a *since she reported her dominant-strategy message for a type that ranks* a *top, player* 2 *can get any of the three remaining items by reporting its dominant strategy-message associated to a rational preference that ranks it top, and player* 2 *chooses* d *out of that set (with spouse* 1 *first eliminating* e*, and spouse* 2 *then choosing* d *out of* $\{b, d\}$*). Similarly,* g *has to allocate* d *to player* 2 *for the message combination* $(m_1^*(t_1^b), m_2^*(t_2^*), \cdot, \cdot)$. *But suppose now the spouses believe it is equally likely that player* 1 *reports* $m_1^*(t_1^a)$ *or* $m_1^*(t_1^b)$. *Let* p *be any belief over* M_{-2} *with that feature. By the previous argument,* $\ell^p(m_2^*(t_2^*))$ *is simply* d *for sure. Other lotteries in* $L^p(m_2^*(t_2^*))$ *include* $\frac{1}{2}a \oplus \frac{1}{2}e$ *(using for* 2 *a dominant strategy message for a rational preference that ranks* a *top and* e *next) and* $\frac{1}{2}b \oplus \frac{1}{2}e$ *(using for* 2 *a dominant strategy message for a rational preference that ranks* b *top and* e *next). But notice that, at* t_2^**, spouse* 2 *ranks both* $\frac{1}{2}a \oplus \frac{1}{2}e$ *and* $\frac{1}{2}b \oplus \frac{1}{2}e$ *above* d. *Given that spouse* 1 *can eliminate only one option from* $L^p(m_2^*(t_2^*))$*, player* 2 *with type* t_2^* *won't pick* $\ell^p(m_2^*(t_2^*))$ *from* $L^p(m_2^*(t_2^*)))$*, and hence* r_{SD} *is not dominant-strategy implementable.*

Thus, to implement r_{SD} the mechanism designer should favor the dynamic mechanism over its static reduction (and over any static mechanism) if she doubts the players' rationality, because the former has the virtue of eliminating uncertainty players face when making a choice. While the added uncertainty is strategically inconsequential when players are expected utility maximizers, we understand from Theorem 1 that this ceases to be the case with more complex choice functions. Li (2017) also suggest that the dynamic mechanism is preferable to implement r_{SD}, but for a very different reason. In his setting, players are rational and only deterministic outcomes must be considered. The added information provided by the dynamic mechanism is then preferable in his setting – giving rise to a mechanism that is not just strategy-proof, but in fact 'obviously strategy-proof' – because players may be confused when assessing the consequences of their strategies.

Given the prevalence of behavioral biases, and of the oft-overlooked fact that players may really be groups (e.g. households, admission boards, etc.) instead of individuals, it seems important to revisit under this light the mechanism design literature (e.g., in the matching literature) that often relies on strategy-proofness static mechanisms. Even the very premise of participants reporting preferences in a direct mechanism may be wrong since their choices may be incompatible with preference maximization. We hope that future work will explore this question theoretically and empirically.

Ex-Post Equilibria Aside from mechanism design, Example 3 also highlights an issue when generalizing the notion of ex-post equilibrium to boundedly rational choices. Consider the incomplete-information behavioral game associated to r_{SD}. A strategy s_i^* for player i specifies which type to report as a function of its true type.

Assume first that players maximize expected utility. The strategy profile s^* forms an *ex-post equilibrium* if, for each type profile t, $(s_i^*(t_i))_{i\in N}$ forms a Nash equilibrium in the complete-information, ex-post game associated to t. This may seem strange at first: why focus on ex-post games when the problem is to solve games of incomplete information? Its raison d'être is to provide a robust solution to games of incomplete information, as ex-post equilibria are Bayesian Nash equilibria independently of the players' belief hierarchies consistent with their types.

But establishing this robustness relies on the fact that expected utility satisfies the sure-thing principle. Without it, ex-post equilibria may fail to

be robust, as is the case in Example 3. To see this, first consider a reasonable extension of Nash equilibrium to general choice functions. Keeping rational expectations, a strategy profile forms a Nash equilibrium in choices (as used in de Clippel (2014)) if, for each player, the equilibrium outcome belongs to his choice set when considering all outcomes he can generate through unilateral deviations. Given any type profile, reporting one's true type in the corresponding ex-post behavioral game induced by r_{SD} is clearly a Nash equilibrium in choices. But player 2 would not pick truth-telling within its opportunity set when others are truth-telling, and it believes player 1 is equally likely to be of type t_1^a or t_2^b. The reasoning is similar to the one developed in Example 3. Even though truth-telling is selected in each ex-post game, in the absence of uncertainty about player 1's action, player 2 opts against it when uncertain about player 1's type (player 1's choices violate Property P).

References

de Clippel, G., 2014. Behavioral Implementation. *American Economic Review* **104**, 2975-3002.

Li, S, 2017. Obviously Strategy-Proof Mechanisms. *American Economic Review* **107**, 3257-87.

Manzini, P., and M. Mariotti, 2007. Sequentially Rationalizable Choice. *American Economic Review* **97**, 1824-1839.

Savage, L., 1972. The Foundations ot Statistics (2nd edition).