Journal of the Economic Science Association (2024) 10:15–30
https://doi.org/10.1007/s40881-023-00145-3

ORIGINAL PAPER

Invariance of equilibrium to the strategy method I: theory

Daniel L. Chen[1] · **Martin Schonger**[2]

Received: 31 March 2023 / Revised: 29 May 2023 / Accepted: 6 August 2023 /
Published online: 5 October 2023

Abstract

This article highlights a potential and significant economic–theoretical bias in the widely used strategy method (SM) technique. Although SM is commonly employed to analyze numerous observations per subject regarding rare or off-equilibrium behaviors unattainable through direct elicitation (DE), researchers often overlook a critical distinction. The strategic equivalence between SM and DE is applicable in the context of monetary payoff games, but not in the actual utility-based games played by participants. This oversight may lead to inaccurate conclusions and demand a reevaluation of existing research in the field. We formalize the mapping from the monetary payoff game to this actual game and delineate necessary and sufficient conditions for strategic equivalence to apply.

Keywords Theory of experiments · Strategy method · Social preferences · Intentions · Deontological motivations

JEL Classification C90 · D64 · A13 · D03

Daniel L. Chen acknowledges IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, Grant ANR-17-EUR-0010. This research has benefited from financial support of the research foundation TSE-Partnership and ANITI funding, and of Alfred P. Sloan Foundation (Grant No. 2018-11245), European Research Council (Grant No. 614708), Swiss National Science Foundation (Grant Nos. 100018-152678 and 106014-150820), and Templeton Foundation (Grant No. 22420).

✉ Daniel L. Chen
daniel.chen@iast.fr

Martin Schonger
mschonger@ethz.ch

[1] Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France

[2] ETH Zurich, Center for Law and Economics, Zurich, Switzerland

1 Introduction

The strategy method (SM), an increasingly popular way to estimate preferences, consists of asking participants to indicate their choices at all information sets rather than only those actually reached. One then compares the differences in decisions at different information sets. For example, to identify the effect of a low offer in an ultimatum game, one might compare the changes in decisions for the low-offer information set with the decisions for the high-offer information set. The appeal of SM comes from its simplicity as well as its potential to elucidate the equilibria that are actually played when theoretical models indicate there are multiple equilibria. SM also has the potential to circumvent many of the endogeneity problems that arise in estimating preferences when making comparisons between heterogeneous individuals.

SM is a straightforward yet powerful tool used in economics research that involves requesting participants to make choices at all possible information sets, rather than exclusively at the ones reached. By comparing the variations in decisions across different information sets, researchers can gain valuable insights. For instance, in an ultimatum game, assessing the effects of a low offer can be achieved by contrasting the decisions made in low-offer and high-offer scenarios. SM's appeal lies in its simplicity and its capacity to reveal the actual equilibria played when theoretical models suggest multiple possibilities. Additionally, SM can help overcome endogeneity issues that emerge when estimating preferences in comparisons between diverse individuals.

However, SM has its limitations and can yield different inferences than data collected using direct elicitation (DE) (Brandts and Charness 2000, 2011; García-Pola et al. 2020; Chen and Schonger 2023). The open question is why and under which conditions the methods are not unambiguously equivalent. We argue that when the payoffs of the game played with SM are an affine transformation of the payoffs at the induced terminal nodes in the game played with DE, the two games are strategically equivalent, and the game played with SM essentially coincides with the strategic form of the game played with DE. Since this condition might not hold, SM is subject to a possibly *severe economic–theoretical bias*. A large body of economic theory renders differences in information sets in SM and DE. The information set for a DE decision node is not the same information set for the same decision node in SM. While economic theory of off-equilibrium motivations is frequently modeled, it is implicitly assumed away by researchers using SM. Three factors make off-equilibrium motivations an especially important issue in the SM context. First, SM usually relies on many decisions at different information sets. Second, the most commonly used dependent variables in SM are typically highly related. Third, and this is an intrinsic aspect of SM, the off-equilibrium decisions can affect the *utility* of decisions at different information sets, even when they do not affect the monetary payoff. These three factors reinforce each other so that, relative to DE, SM for treatment effects could be severely biased.

Motivations that are based on disappointment aversion (Gul, 1991), intentions (Battigalli et al., 2007; Fehr & Schmidt, 2000), self-image (Bénabou & Tirole,

2011), identity (Bullock, 2019), emotions (Elster, 1998; Loewenstein, 2000), or duty (Chen & Schonger, 2022), for instance, can cause equilibrium outcomes to differ between SM and DE. We provide a formal, general framework that embeds prior non-formal (psychological).[1] explanations for differences between SM and DE to show that these explanations only hold under certain conditions.[2]

Another theoretical critique of SM is that the invariance of equilibrium outcomes relies on individuals eliminating weakly dominated strategies. Game theorists may disagree about the actual prevalence in the field of individuals who play weakly dominated strategies, either because eliminating them requires a greater level of cognition or because they may simply be more credible than those eliminated through subgame perfection. Our critique is an independent one. Our theoretical results focus on motivations (preferences) rather than deviations from rationality in decision-making. We provide a model-based elaboration that complements a footnote by Roth (1995, fn 84) that "the notion of subgame perfect equilibrium is lost in the transition from the extensive to the strategic form of the game, since there are no subgames in a game in which players state their strategies simultaneously." We cite a theorem from Moulin (1986, pp. 84–86) and Rochet (1981) that would prima facie invalidate SM.

Dozens of experimental studies have investigated whether SM yields the same responses as DE where participants actually play the extensive form game. Though a recent study concluded in favor of using SM, it reported statistically significant and economically important differences in behavior by elicitation method in a considerable fraction of the studies comparing the two elicitation methods: "We do find, however, that a particular aspect of emotion-related behavior, the use of punishment, is significantly more likely in situations with direct response than with strategy choice." Brandts and Charness (2011, pg. 394) In our reading, the set of games it studied divide into two: in simple games that had moral content,[3] SM and

[1] Much of the debate surrounding the validity of the SM estimate typically revolves around the possibility of emotion or cognitive fatigue associated with making multiple decisions. In psychology, a large body of work is devoted to construal theory, which would prima facie invalidate SM estimates (Liberman & Trope, 1998; Metcalfe & Mischel, 1999; Trope & Liberman, 2003, 2010) Construal theory involves the relation between psychological distance and the extent to which people's thinking (e.g., about objects and events) is abstract or concrete. An example of construal level effects is that planning a summer vacation one year in advance will cause one to focus on broad features of the situation, like fun and relaxation, while the very same vacation planned for next week will cause one to focus on specific features, like what restaurants to make reservations for. Temporal construal is believed to underlie a broad range of temporal changes in evaluation, prediction, and choice.

[2] When behavior does diverge between SM and direct elicitation (DE), researchers have suggested that DE settings involve a different degree of emotions being present, for example, when reacting to an actual violation of a fairness norm than when contemplating a violation (Fehr et al., 2004); or individuals may be induced to think harder in the SM setting (Casari & Cason, 2009a), spend more time making the decision (Rand et al., 2012), or, instead, think less hard in the SM setting, and put less effort at each decision node because they receive less monetary return per decision (Fehr et al., 2004). Those other papers did not present a formal model for the divergence.

[3] These would include ultimatum games (Eckel & Grossman, 2001; Guth et al., 2001; Oxoby & McLeish, 2004; Armantier, 2006; McGee & Constantinides, 2013) punishment games (Brandts & Charness, 2003; Brosig et al., 2003; Falk et al., 2005), trust games (Murphy et al., 2006; Fong et al., 2007;

DE tend to diverge, while in more complex games that were framed as economics games,[4] SM and DE did not diverge or had mixed results. This difference is consistent with the heightened relevance of off-equilibrium considerations in social preference games. García-Pola et al. (2020) test SM vs. DE for four centipede games between two players.[5] They run two centipede games where the incentives of each player are symmetric and two centipede games where the incentives are asymmetric. In the first two centipede games, they find that SM and DE diverges, in particular, SM seems to yield results that are more cooperative (stopping toward the end of the centipede game) whereas DE yields results that are less cooperative (stopping at the beginning). In the second two centipede games, they find that SM and DE yield similar findings. One interpretation of this difference is that the symmetry in payoffs for the two players allowed them to think more about why the player was moving along the centipede in the strategy method. That is, the off-equilibrium outcomes were more salient for the players because of the symmetry. Schotter et al. (1994) presents games in extensive vs. normal form and finds that differences emerge in the simplest games, where subjects are more likely to use and fear incredible threats. This is consistent with our reading of the previous literature and the interpretation whose formalization we present here.

Footnote 3 (continued)

Casari & Cason, 2009b; Solnick, 2007; Meidinger et al., 2001; Cox & Hall, 2010), public goods and cooperation games (Offerman, 2001; Fischbacher & Gächter, 2010; Mengel & Peeters, 2011; Büchner et al., 2007; Muller et al., 2008), and prisoner dilemma/minority games (Schotter et al., 1994; Brandts & Charness, 2000; Linde et al., 2014; Reuben & Suetens, 2012).

[4] These would be games simulating firms (Kübler & Müller, 2002) market entry (Rapoport et al., 1995; Seale & Rapoport, 2000; Sundali et al., 1995), asset pricing (Hommes, 2005), auction (Armantier & Treich, 2009; Goeree et al., 2002; Rapoport et al., 1995), insurance (Bosch-Domenech & Joaquim, 2006), buying and selling games (Cason & Mui, 1998; Sonnemans, 2000), principal-agent games (Falk & Kosfeld, 2006), and negotiation games (Mitzkewitz & Nagel, 1993a; Rapoport et al., 1996; Rapoport & Sundali, 1996).

[5] The centipede game is a sequential, extensive-form game in economics that is often used to explore the concepts of rationality, backward induction, and subgame perfect equilibrium. The game involves two players, typically denoted as Player A and Player B, who take alternating turns in a series of rounds. The game is characterized by a predetermined number of rounds, and at each round, the active player has two options: either "take" the pot of money or " pass" and continue to the next round. When a player chooses to " take," the game ends immediately, and the pot is divided between the players according to the round-specific predetermined split. Generally, the player who " takes" receives a larger share of the pot, while the other player receives a smaller share. If a player chooses to " pass," the game continues to the next round, and the pot grows larger. The centipede game challenges the notion of rational behavior, as the backward induction solution suggests that a fully rational player should " take" the pot in the first round, preventing the game from continuing. However, empirical observations often indicate that players tend to " pass" for several rounds before deciding to " take," thus deviating from the theoretically predicted outcome. In economics, the centipede game serves as an important tool for analyzing decision-making, cooperation, and the discrepancies between theoretical predictions and observed behavior in strategic interactions.

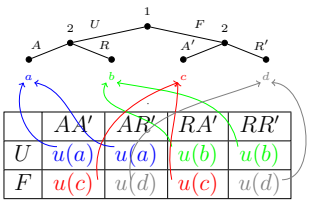
2 Theoretical background: direct elicitation vs. strategy method

In an experiment, the observable vector is that of monetary payoffs and may not capture what utility players might get from feelings, such as revenge, gratitude, kindness, or warm-glow. But even from a purely theoretical perspective, the Kohlberg–Mertens view is not universally accepted. Harsanyi (1988) disagree and argue that in general the solution of a game with a sequential structure simply has to depend on this sequential structure and cannot be made dependent on the normal form only. We show that even if one accepts the Kohlberg and Mertens (1986) view, it cannot be used as a justification for the strategy method of elicitation without further assumptions. The reason, in short, is that researchers neither observe the preferences nor the players' conception of the game, and there are plausible circumstances where use of SM rather than DE can change players' conception of the game.

The upshot, in our view, is not to check theoretically which motivations break invariance in every circumstances, since the number of potential motivations is large. For instance, common theoretical motivations like intentions (Battigalli et al., 2007; Fehr & Schmidt, 2000), disappointment aversion (Gul, 1991), and self-image (Bénabou & Tirole, 2011), to name a few, can cause divergence, but the parameters in the player's utility function are also unobserved. Rather, off-equilibrium considerations accepted by formal theorists and by experimentalists can intuitively break invariance between SM and DE as we illustrate theoretically in the appendix.

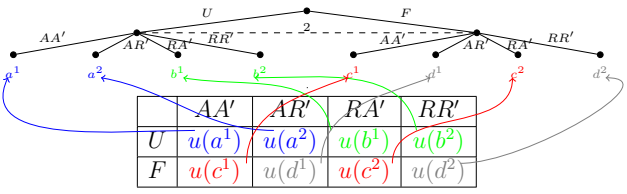
2.1 Linking theory to data

We can visualize the assumption behind experiments that rely on the invariance between SM and DE using the simplified ultimatum game. Under DE:



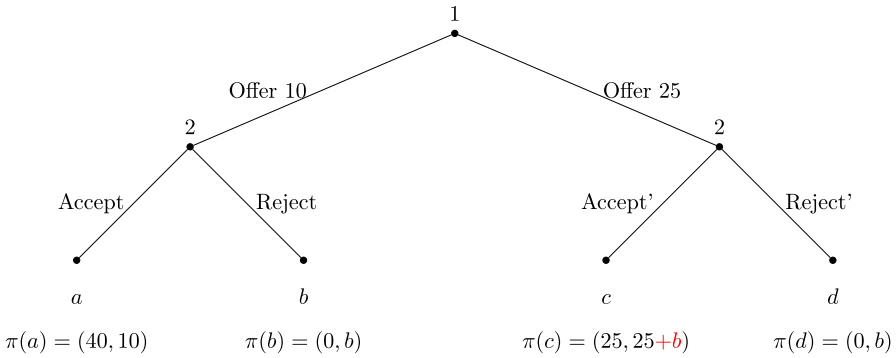
	AA'	AR'	RA'	RR'
U	$u(a)$	$u(a)$	$u(b)$	$u(b)$
F	$u(c)$	$u(d)$	$u(c)$	$u(d)$

Under SM:



	AA'	AR'	RA'	RR'
U	$u(a^1)$	$u(a^2)$	$u(b^1)$	$u(b^2)$
F	$u(c^1)$	$u(d^1)$	$u(c^2)$	$u(d^2)$

Even more concretely, the following simplified 0–50 ultimatum game illustrates how non-consequentialist motivations can breakdown the invariance between SM and DE when collecting data. Suppose player 2 has duty motives: If he did not commit or if he did not, in fact, accept the unfair offer, he gets an additional psychic benefit of $0 < b < 10$. In the DE setting, if player 2 is offered 25 and he accepts, the utilities are $(25, 25 + b)$. If player 2 is offered 10 and he accepts, the utilities are (40, 10).



		AA' $x \geq 10$	RA' $x \geq 25$	AR'	RR'
p	10	$(40, 10)$	$(0, b)$	$(40, 10)$	$(0, b)$
$1 - p$	25	$(25, 25)$	$(25, 25 + b)$	$(0, 0)$	$(0, 0)$

In the SM setting, the strategy, accept $x \geq 10$, yields: $p * 10 + (1 - p) * 25 = 25 - 15p$ (p is the subjective belief of the responder on the choice of the proposer), while the strategy, accept $x \geq 25$, yields: $p * (0 + b) + (1 - p) * (25 + b) = 25 + b - 25p$. Then player 2 picks the strategy, accept $x \geq 25$, if and only if $p < 0.1b$. That is, player 2 only accepts high offers if and only if there is low probability of bearing the adverse consequences of indulging in the psychic benefit of not being a loser. If $p < 0.1b$, then the DE setting yields payoffs (40, 10) while the SM setting yields payoffs (25, 25).[6]

[6] Note that the self-image concern b must not scale with p linearly for this statement to hold. The Kantian categorical imperative would be an example. For a general statement about willingness to act on non-consequentialist motivations when the decision becomes more hypothetical (e.g., in the random lottery incentive), Chen and Schonger (2022) develop a shredding criterion for non-consequentialist motivations.

3 Conclusion

Our study suggests that, because of off-equilibrium motivations, conventional SM estimates may be biased, leading to misleading treatment effects relative to DE. We show that when the payoffs of the game played with SM are an affine transformation of the payoffs at the induced terminal nodes in the game played with DE, the two games are strategically equivalent, and the game played with SM essentially coincides with the strategic form of the game played with DE. However, since a large fraction of SM papers rely on many decisions at different information sets that are typically highly related, the off-equilibrium decisions can affect the *utility* of decisions at different information sets, even when they do not affect the monetary payoff. These factors are mutually reinforcing so that the SM for treatment effects could be biased. Theoretically, SM may be positively or negatively biased away from DE depending on how utility interacts with decisions at other information sets. Chen and Schonger (2023) demonstrate that SM is prone to substantial biases, which may be as significant as other observed treatment effects. Additionally, minor manipulations to salience can intensify these discrepancies to a similar extent. Notably, the direction of treatment impacts can vary greatly between SM and direct effects (DE) and may even reverse in sign.

We have illustrated variance of equilibrium in SM vs. DE with simple models of off-equilibrium motivations. Differences between DE and SM can reveal the importance of motivations beyond strong consequentialist ones. An alternative to the view of natural field experiments (a subset of DE) as the gold standard for causal estimates (Harrison & List, 2004; Levitt & List, 2007) is that differences between SM and DE can be used to understand the general way in which agents' motivations influence behavior (Camerer, 2011). To be sure, another reason to use SM may be if the situation approximates natural decision-making. However, if DE is the gold standard, one possible solution for experiments is to consider a pilot that first tests whether SM and DE diverges before collecting additional data using SM.

The closest economic analog to our argument in the field may be the *drafting* of a contract (Battigalli & Maggi, 2002; Tirole, 1999; Schwartz & Watson, 2004). Contemplation of all possible contingencies involves SM decision-making, while the actual decision when the information set is revealed involves DE decision-making. Differences in decision-making provide another reason, besides incentive compatibility, why agents might not have incentives aligned with principals. Legal doctrine has neglected this dimension of contractual capacity.

Appendix A: theory

A.1. Background and history

The earliest use of a "strategy method" can be found in Selten (1967), where subjects are asked to give a strategy for the entire game instead of being asked only for decisions and information sets that are actually reached. As Roth (1995) points out,

Selten' s strategy method first lets participants gain practice by playing the game several times, only then asking them for strategies. In addition, Selten uses group discussions and individual advising of participants by the experimenter to help subjects formulate strategies in what are rather complex games. In comparison to Selten's games, the games used in more recent studies tend to be much simpler, typically two-player games where each player has only one move. In these recent studies, there is no group discussion or individual advising. Thus, the currently used strategy method is the same as Selten' s except for pre-game practice and the group discussion and advising (for an early example, see Mitzkewitz and Nagel, 1993b). In both Selten' s SM and the modern SM, subjects are made aware of an extensive game, but instead of actually playing it, they are asked for their (hypothetical) decision at every decision node. Typically the game is not represented in strategic (i.e., matrix) form (for an exception see Schotter et al. (1994)). SM contrasts with DE (also referred to as direct response method) where players are only asked for their decisions at information sets that are actually reached. We follow convention and sometimes refer to SM as the cold, and DE as the hot setting.

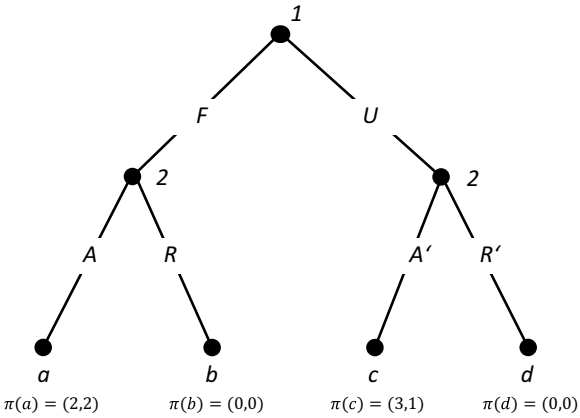
Formally, the games played in SM and in DE can both be represented by an extensive form game. The extensive form games differ, but the corresponding normal form is the same for both methods. In that sense, they are theoretically equivalent. There is the view that for rational players the strategic form captures all relevant information, while different corresponding extensive forms differ only in irrelevant representation. Kohlberg and Mertens (1986, p. 1011) put this view nicely by writing, "In some sense, the fact that the reduced normal form captures all the relevant information for decision purposes results directly from the (almost tautological) fact that what matters for decision purposes in an outcome is only the corresponding utility vector (and not, e.g., the particular history leading to that outcome)." Osborne and Rubinstein (1994, p. 90) echoes Kohlberg and Mertens (1986), "As in the case of a strategic game we often specify the players' preferences over terminal histories by giving payoff functions that represent the preferences."

A.2. Failure of invariance

Consider the game in Fig. 1. It is a kind of mini-ultimatum game. Player 1, the proposer, divides an endowment of \$4 between herself and player 2, the responder. The offer she makes can be either fair (2, 2) or unfair (3, 1). If both players are purely self-interested, the unique subgame perfect equilibrium is (UAA') resulting in terminal node c. In the strategic form, shown in the left of Fig. 2, that is the unique strategy profile to survive iterated elimination of weakly dominated strategies; thus, we have invariance.

Let us vary the example and show how and when something like duty, for example, can break this invariance. First, let us incorporate duty in a way which does *not* break invariance. Assume that whenever the responder has accepted an unfair offer, i.e., responded A' to U, she suffers a psychic loss worth α, where $0 < \alpha < 1$. One can interpret this as damage to her honor (Nisbett, 1996). The unique subgame perfect equilibrium is again (UAA') resulting in terminal node c, which is also the

Fig. 1 A mini-ultimatum game



	AA'	AR'	RA'	RR'
F	$u(a)$	$u(a)$	$u(b)$	$u(b)$
U	$u(c)$	$u(d)$	$u(c)$	$u(d)$

	AA'	AR'	RA'	RR'
F	$(2,2)$	$(2,2)$	$(0,0)$	$(0,0)$
U	$(3,1)$	$(0,0)$	$(3,1)$	$(0,0)$

	AA'	AR'	RA'	RR'
F	$(2,2)$	$(2,2)$	$(0,0)$	$(0,0)$
U	$(3,1-\alpha)$	$(0,0)$	$(3,1-\alpha)$	$(0,0)$

	AA'	AR'	RA'	RR'
F	$(2,2-\alpha)$	$(2,2)$	$(0,-\alpha)$	$(0,0)$
U	$(3,1-\alpha)$	$(0,0)$	$(3,1-\alpha)$	$(0,0)$

Fig. 2 Strategic form: mini-ultimatum games

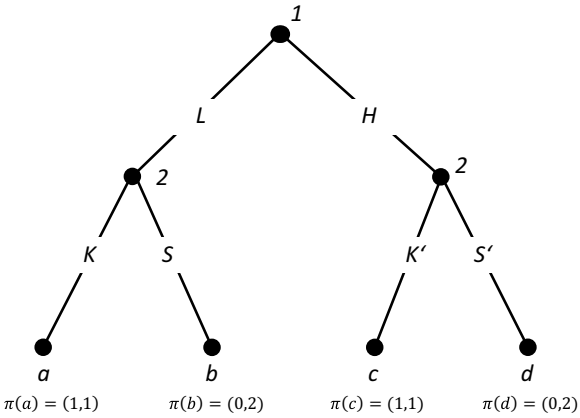
sole surviving profile of iterated elimination of weakly dominated strategies in the strategic form shown in the third matrix of Fig. 2.

Now assume a twist: The responder not only suffers a physic loss α when she has responded A' to U, but also when she has merely bindingly decided to do so. If the game is directly elicited, there is no opportunity to commit, and the unique subgame perfect equilibrium remains (UAA') resulting in terminal node c. If the game is elicited via SM, what is played is shown in the rightmost matrix in Fig. 2: Four strategy profiles survive iterated elimination of weakly dominated strategies, and the Nash equilibria among those are (UAA') as before, but in addition (FAR'). Why does this happen? Note that both (FAA') and (FAR') result in node a. But in the strategic form, they now have different utilities. This means that the reason for failure of invariance is that this strategic from cannot represent a game tree of the form shown in Fig. 1.

Thus, non-consequentialist preferences can generate differential predictions in DE vs. SM settings.[7] One might call these preferences for duty or see them as at least partially rule-based (i.e., to maintain honor). Note that with these parameters, not all concerns that incorporate off-equilibrium information break invariance in this game. For example, a psychic gain when one has committed to accepting an unfair offer (e.g., a turn-the-other-cheek self-image preference) would not break invariance. Such a responder would behave like homo oeconomicus.

[7] A preference is strongly consequentialist if it depends on payoffs (agent' s own and others') only.

Fig. 3 Tribal game



A.3. Invariance example

Next, we provide another example ("tribal game") where emotions affect decision-making, but invariance in DE vs. SM holds.

In the game in Fig. 3, player 1 sends player 2 a message, where L means that she loves ISIS and H means she hates it. Player 2 has an endowment of \$2, and in response can either be kind (K, respectively K') and share equally or selfish and keep all to herself (S, respectively S'). Thus, the payoff function of G^{π} is given by $\pi(a) = \pi(c) = (1, 1)$ and $\pi(b) = \pi(d) = (0, 2)$. If both players are purely self-interested, then the game has two subgame-perfect Nash equilibria (LSS') and (HSS'), which yield the terminal nodes b, respectively. d, and payoff (0, 2). Elimination of weakly dominated strategies in the strategic form gives the same equilibria.

Consider a social preference, specifically Fehr–Schmidt preferences for player 2. In this game, regardless of the choice of parameters for advantageous and disadvantageous inequality, Fehr–Schmidt preferences imply that $u_2(b) = u_2(d) > u_2(a) = u_2(c)$. Thus, the ranking of terminal nodes happens to remain unchanged and the analysis of equilibria is as before.

Now let us construct an example where the social preference changes the equilibria. Consider a very altruistic player 2 with preferences represented by $u_2(a) = u_2(c) > u_2(b) = u_2(d)$. A functional form from payoff vectors into utility that yields such a preference between terminal nodes would be, for example, $u_2(t) = \pi_2(t) + \alpha\pi_1(t)$, where $\alpha > 1$. The game has two subgame-perfect Nash equilibria (LKK') and (HKK'), which yield the terminal nodes a, respectively c, and payoff (1, 1). Elimination of weakly dominated strategies in the strategic form gives the same equilibria. Again, there is an invariance between the extensive and strategic forms.

Now let us change the game by changing the preference of player 2 only. Assume that she is an avid ISIS fan, and thus prefers to be kind to someone who also claims to love ISIS, and unkind to someone who does not. Specifically, assume that $a \succ d \succ b \succ c$, which, moreover, means that she prefers to encounter people who

profess to be fans. Note that these preferences for player 2 are *not* a function of payoffs only; though $\pi(a) = \pi(c)$ she is not indifferent between a and c. Nevertheless, this extensive game is invariant to the method of elicitation: the unique subgame-perfect equilibrium is (Llh') yielding the terminal node a with payoff (not utility) (1, 1). In the strategic form, iterated elimination of weakly dominated strategies gives the same equilibrium. Note that this invariance holds even though emotions play a role in player 2's decisions.

A.4. General proof

In standard game theory, one way to describe an extensive form game with perfect information is by means of a tree Γ, a set of players $\{1, \ldots, I\}$, the set of nodes T, the decision nodes X, and set of terminal nodes Z, who plays at each decision node $\tau : X \to \{1, \ldots, \mathrm{I}\}$, and a complete and transitive preference over the terminal nodes represented by Bernoulli utility functions $u_i(a) : Z \to \mathbb{R}$. Thus let $G = (\Gamma, T_i, u_i, i = 1, \ldots, N)$ describe our extensive form game. Throughout we shall assume rationality and common knowledge.

Whether implemented in a laboratory or field setting, the preferences over terminal nodes are not directly observable by the researcher. One then typically assigns monetary payoffs to each terminal node, thus implementing a "game" $G^{\pi} = (\Gamma, T_i, \pi_i, i = 1, \ldots, N)$, where $\pi_i : Z \to \mathbb{R}$ assigns player i a payoff at every terminal node. G^{π} is a game in the game-theoretic sense with an additional assumption that all players' preferences are purely self-interested and this is common knowledge.

We can denote the DE extensive form game as G^{DE}, with extensive form Γ^{DE} and the corresponding Bernoulli utility functions $u_i^{DE} : Z^{DE} \to \mathbb{R}$. We compare the direct elicitation, $G_{\pi}^{DE} = \left(\Gamma^{DE}, \pi^{DE} : Z^{DE} \to \mathbb{R}\right)$ and $G^{DE} = \left(\Gamma^{DE}, u^{DE} : Z^{DE} \to \mathbb{R}\right)$ with the strategy method, $G_{\pi}^{SM} = \left(\Gamma^{SM}, \pi^{SM} : Z^{SM} \to \mathbb{R}\right)$ and $G^{SM} = \left(\Gamma^{SM}, u^{SM} : Z^{SM} \to \mathbb{R}\right)$.

The design choice of experimenter is Γ^{DE}, π^{DE}. Let $\Gamma^{SM} \equiv \phi\left(\Gamma^{DE}\right)$ (using the natural order of players), where ϕ : ext. forms $\to$ ext. forms and $\zeta : Z^{SM} \to Z^{DE}$ (z^{DE} associated with several strategy profiles). By definition of SM, $\pi^{SM}\left(z^{SM}\right) = \pi^{DE}\left(\zeta\left(z^{SM}\right)\right)$. Note that u^{DE} and u^{SM} are neither a design choice nor directly observable. The following chart summarizes the theorem:

$G_{\pi}^{DE} = \left(\Gamma^{DE}, \pi^{DE} : Z^{DE} \to \mathbb{R}^{\mathrm{I}}\right)$	Strat. iden. $\Leftrightarrow$	$G_{\pi}^{SM} = \left(\Gamma^{SM}, \pi^{SM} : Z^{SM} \to \mathbb{R}^{\mathrm{I}}\right)$
$\uparrow$ equilibrium may change $\downarrow$		$\uparrow$ equilibrium may change $\downarrow$
$G^{DE} = (\Gamma^{DE}, u^{DE} : Z^{DE} \to \mathbb{R}^{\mathrm{I}})$	Thm. $\Leftrightarrow$	$G^{SM} = (\Gamma^{SM}, u^{SM} : Z^{SM} \to \mathbb{R}^{\mathrm{I}})$

Strategic equivalence: G^{DE} and G^{SM} are strategically equivalent if and only if for all players i, there exist real numbers $\alpha_i, \beta_i > 0$ such that for all $z^{SM} \in Z^{SM}$: $u_i^{SM}\left(z^{SM}\right) = \alpha_i + \beta_i u^{DE}\left(\zeta\left(z^{SM}\right)\right)$.

Conventional wisdom: The strategic forms of G_π^{DE} and G_π^{SM} are strategically equivalent.

Outcome-based preferences: If for all players i, there exists a function $f_i : \mathbb{R}^1 \to \mathbb{R}$ such that $u_i^{DE}\left(z^{DE}\right) = f_i\left(\pi\left(z^{DE}\right)\right)$ and $u_i^{SM}\left(z^{SM}\right) = f_i\left(\pi\left(z^{SM}\right)\right)$, then G^{DE} and G^{SM} are strategically equivalent.

These results follow from Axiom 1, as formulated by Moulin (1986, pgs. 84–86):

Axiom 1 (one-to-one) A game $(\Gamma, T_i, u_i, i = 1, \ldots, N)$ satisfies the one-to-one condition if for any terminal nodes $z, z' \epsilon Z(T)$ and any player i:

If $u_i(z) = u_i(z')$ then $u_j(z) = u_j(z')$ for all $j = 1, .., N$.

The theorem below follows the formulation of Moulin (1986, pgs. 84–86) and Rochet (1981):

Theorem 1 *Let $G = (\Gamma, T_i, u_i, i = 1, \ldots, N)$ be an N-player game in extensive form with perfect information satisfying the one-to-one assumption. Then the associated normal form of G is solvable by iterated elimination of weakly dominated strategies, and the equilibrium payoffs are the same as obtained in the extensive form by Kuhn's algorithm.*

Theorem 1 is only applicable if the payoffs given in the game are indeed the Bernoulli utility of the players. But researchers observe the monetary payoffs, but not the Bernoulli utility numbers of the players. Put simply, many motivations commonly modeled and tested in economics research will break invariance and the number of potential motivations is large. We present a few applications to illustrate.

The following two observations extend the applicability of the original theorem. First note that the risk attitude of a player need not be neutral, but can be anything:

Corollary 1 (Risk attitude) *Let $G = (\Gamma, T_i, u_i, i = 1, \ldots, N)$ be an N-player game in extensive form satisfying the one-to-one assumption. Let the domain of preferences be the agent's payoffs. Let preferences be a strict ordering. Then the normal form of G is solvable by iterated elimination of weakly dominated strategies, and the equilibrium payoffs are the same as obtained in the extensive form by Kuhn's algorithm.*

Corollary 2 (Social preferences) extends this result to social preferences:

Corollary 2 (Social preferences) *Let $G = (\Gamma, T_i, u_i, i = 1, \ldots, N)$ be an N-player game in extensive form satisfying the one-to-one assumption. Let the domain of preferences be the vector of payoffs. Let preferences be locally non-satiated. Then, the normal form of G is solvable by iterated elimination of weakly dominated strategies, and the equilibrium payoffs are the same as obtained in the extensive form by Kuhn's algorithm.*

A standard response in behavioral economics to inaccurate predictions of the *homo oeconomicus* model is to assume richer preferences, particularly those that depend not only on the agent's own monetary payoff but also on the payoffs of others. We say that a player is *purely self-interested* (*homo oeconomicus*) if for all terminal nodes a, b, $\pi_i(a) \geqq \pi_i(b)$ if and only if $u_i \geqq u_i(b)$. *Social preferences* is where a player's preference between two nodes is a *function* of their monetary payoffs only.[8] Thus, we say that a player has social preferences if for all terminal nodes a, b, if $\pi(a) = \pi(b)$ then $u(a) = u(b)$.

This subsection discusses whether such preferences can generate differential predictions for DE vs. SM when the standard ones fail to do so. The answer is negative, and it is negative for all strongly consequentialist preferences, which we define as follows:

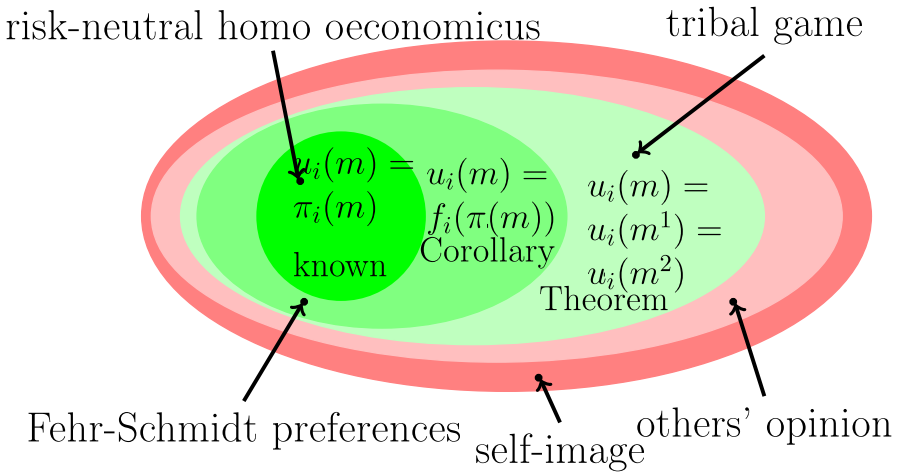
Definition 1 A preference is strongly consequentialist if it depends on payoffs (agent's own and others') only.

Fact *If the equilibrium concept depends on the reduced normal form only, then for all strongly consequentialist preferences the set of equilibria under direct elicitation is identical to the set of equilibria under the strategy method.*

Thus, while social preferences can generate a different prediction than standard preferences about what the equilibrium will be, each social preference creates the same equilibrium prediction for DE and SM as long as one follows the Kohlberg–Mertens view.

[8] In contrast, *homo oeconomicus* preferences are simply the monetary payoffs. Also, by social preferences we refer to preferences like Fehr–Schmidt preferences, but not intention-based preferences, which will be in a separate category.

Venn diagram of theorem:



Green color indicates instances where DE and SM equilibria coincide (see examplesin the appendices). Red color indicates instances where equilibria may differ.

References

Armantier, O. (2006). Do wealth differences affect fairness considerations? *International Economic Review, 47*(2), 391–429.

Armantier, O., & Treich, N. (2009). Subjective probability in games: An application to the overbidding puzzle. *International Economic Review, 50*(4), 1079–1102.

Battigalli, P., & Maggi, G. (2002). Rigidity, discretion, and the costs of writing contracts. *The American Economic Review, 92*(4), 798–817.

Battigalli, P., Maggi, G., & Dufwenberg, M. (2007). Guilt in games. *The American Economic Review, 97*(2), 170–176.

Bénabou, R., & Tirole, J. (2011). Identity, morals, and taboos: beliefs as assets. *The Quarterly Journal of Economics, 126*(2), 805–855.

Bosch-Doménech, A., & Joaquim S., et al. (2006). "Risk Aversion and Embedding Bias," Technical Report .

Brandts, J., & Charness, G. (2000). Hot vs. cold: Sequential responses and preference stability in experimental games. *Experimental Economics, 2*, 227–238.

Brandts, J., & Charness, G. (2003). Truth or consequences: An experiment. *Management Science, 49*(1), 116–130.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics, 14*, 375–398.

Brosig, J., Weimann, J., & Yang, C. L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics, 6*(1), 75–90.

Büchner, S., Coricelli, G., & Greiner, B. (2007). Self-centered and other-regarding behavior in the solidarity game. *Journal of Economic Behavior & Organization, 62*(2), 293–303.

Bullock, J., & Lenz, G. (2019). Partisan bias in surveys. *Annual Review of Political Science, 22*, 325–342.

Camerer, C. (2011). "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List," *Available at SSRN 1977749*.

Casari, M., & Cason, T. N. (2009). The strategy method lowers measured trustworthy behavior. *Economics Letters, 103*(3), 157–159.

Cason, T. N., & Mui, V.-L. (1998). Social influence in the sequential dictator game. *Journal of Mathematical Psychology, 42*(2), 248–265.
Chen, D.L., & Schonger, M. (2022). Social preferences or sacred values? Theory and evidence of deontological motivations. *Science Advances*, *8* (19), eabb3925. https://doi.org/10.1126/sciadv.abb3925.
Chen, D.L., & Schonger, M. (2023). Invariance of equilibrium to the strategy method II: Experimental evidence. *Journal of the Economic Science Association*. https://doi.org/10.1007/s40881-023-00146-2.
Cox, J. C., & Hall, D. T. (2010). Trust with private and common property: Effects of stronger property right entitlements. *Games, 1*(4), 527–550.
Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry, 39*(2), 171–188.
Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature, 36*(1), 47–74.
Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *The American Economic Review*, , pp. 1611–1630.
Falk, A., Kosfeld, M., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica, 73*, 2017–2030.
Fehr, E., & Schmidt, K. M. (2000). Fairness, incentives, and contractual choices. *European Economic Review, 44*(4), 1057–1068.
Fehr, E., Schmidt, K. M., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87.
Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review, 100*(1), 541–556.
Fong, Y.-F., Huang, C.-Y., & Offerman, T. (2007). Guilt driven reciprocity in a psychological signaling game.
García-Pola, B., Iriberri, N., & Kovářík, J. (2020). Hot versus cold behavior in centipede games. *Journal of the Economic Science Association, 6*, 226–238.
Goeree, J. K., Holt, C. A., & Palfrey, T. R. (2002). Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory, 104*(1), 247–272.
Gul, F. (1991). A theory of disappointment aversion. *Econometrica, 59*(3), 667–686.
Guth, W., Huck, S., & Mueller, W. (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior, 37*(1), 161–169.
Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.
Harsanyi, J. C. (1988). *Selten: A general theory of equilibrium selection in games*. MIT Press.
Hommes, C., Sonnemans, J., Tuinstra, J., & van de Velden, H. (2005). A strategy experiment in dynamic asset pricing. *Journal of Economic Dynamics and Control, 29*(4), 823–843.
Kohlberg, E., & Mertens, J.-F. (1986). On the strategic stability of equilibria. *Econometrica, 54*(5), 1003–1037.
Kübler, D., & Müller, W. (2002). Simultaneous and sequential price competition in heterogeneous duopoly markets: experimental evidence. *International Journal of Industrial Organization, 20*(10), 1437–1460.
Levitt, S., & List, J. (2007). What do laboratory experiments tell us about the real world? *The Journal of Economic Perspectives, 21*(2), 153–174.
Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology, 75*(1), 5.
Linde, J., Sonnemans, J., & Tuinstra, J. (2014). Strategies and evolution in the minority game: A multi-round strategy experiment. *Games and Economic Behavior, 86*, 77–95.
Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review, 90*(2), 426–432.
McGee, P., & Constantinides, S. (2013). Repeated play and gender in the ultimatum game. *The Journal of Socio-Economics, 42*, 121–126.
Meidinger, C., Robin, S., & Ruffieux, B. (2001). Jeu de l'investissement et coordination par les intentions. *Revue d'économie politique, 111*(1), 67–93.
Mengel, F., & Peeters, R. (2011). Strategic behavior in repeated voluntary contribution experiments. *Journal of Public Economics, 95*(1), 143–148.
Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3.

Mitzkewitz, M., & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory, 22*, 171–198.
Moulin, H. (1986). *Game Theory for the Social Sciences* Studies in Game Theory and Mathematical Economics, 2 ed., NYU Press.
Muller, L., Sefton, M., Steinberg, R., & Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary contribution experiments. *Journal of Economic Behavior & Organization, 67*(3), 782–793.
Murphy, R. O., Rapoport, A., & Parco, J. E. (2006). The breakdown of cooperation in iterative real-time trust dilemmas. *Experimental Economics, 9*(2), 147–166.
Nisbett, R., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Westview Press.
Offerman, T., Potters, J., & Harrie, A. A. V. (2001). Cooperation in an overlapping generations experiment. *Games and Economic Behavior, 36*(2), 264–275.
Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT Press.
Oxoby, R. J., & McLeish, K. N. (2004). Sequential decision and strategy vector methods in ultimatum bargaining: Evidence on the strength of other-regarding behavior. *Economics Letters, 84*(3), 399–405.
Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430.
Rapoport, A., & Sundali, J. A. (1996). Ultimatums in two-person bargaining with one-sided uncertainty: Offer games. *International Journal of Game Theory, 25*(4), 475–494.
Rapoport, A., Sundali, J. A., & Fuller, M. A. (1995). Bidding strategies in a bilateral monopoly with two-sided incomplete information. *Journal of Mathematical Psychology, 39*(2), 179–196.
Rapoport, A., Sundali, J. A., & Seale, D. A. (1996). Ultimatums in two-person bargaining with one-sided uncertainty: Demand games. *Journal of Economic Behavior & Organization, 30*(2), 173–196.
Reuben, E., & Suetens, S. (2012). Revisiting strategic versus non-strategic cooperation. *Experimental Economics, 15*(1), 24–43.
Rochet, J.C. (1981). *Selection on an Unique Equilibrium Value for Extensive Games with Perfect Information Cahiers de mathématiques de la décision*. Centre de recherche de mathématiques de la décision: Université Paris IX-Dauphine.
Roth, A. E. (1995). Bargaining experiments. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics.* Princeton University Press.
Schotter, A., Weigelt, K., & Wilson, C. (1994). A laboratory investigation of multiperson rationality and presentation effects. *Games and Economic Behavior, 6*, 445–468.
Schwartz, A., & Watson, J. (2004). The law and economics of costly contracting. *Journal of Law, Economics, and Organization, 20*(1), 2–31.
Seale, D. A., & Rapoport, A. (2000). Elicitation of strategy profiles in large group coordination games. *Experimental Economics, 3*(2), 153–179.
Selten, R. (1967). Die Strategiemethode zur Erforschung eingeschraenkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. *Beitraege zur experimentellen Wirtschaftsforschung, 1*, 136–168.
Solnick, S. J. (2007). Cash and alternate methods of accounting in an experimental game. *Journal of Economic Behavior & Organization, 62*(2), 316–321.
Sonnemans, J. (2000). Decisions and strategies in a sequential search experiment. *Journal of Economic Psychology, 21*(1), 91–102.
Sundali, J.A., Rapoport, A., & Seale, D.A. (1995).Coordination in market entry games with symmetric players. *Organizational Behavior and Human Decision Processes*, *64*(2), 203–218.
Tirole, J. (1999). Incomplete contracts: where do we stand? *Econometrica, 67*(4), 741–781.
Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*(3), 403.
Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440.