Supplemental materials

Subjective Game Structures: A behavioral game theoretic analysis of hidden perceptions and strategic properties underlying the Israeli-Palestinian conflict

Basic Decision Criteria

(i) Dominant strategy – the choice of an alternative that provides a *better* payoff regardless of which *specific* alternative is chosen by the opponent. A Prisoner's Dilemma (PD) game player who chooses the dominant alternative of *defection* is guaranteed to obtain either the T payoff (if the opponent chooses to cooperate) or the P payoffs (if the opponent chooses to defect). In both cases the player is assured to obtain better payoffs since T > R and P > S.

(ii) MaxiMin strategy – a pessimistic perspective that takes into account the minimal payoffs that may be obtained under each alternative, therefore choosing the alternative that contains the *maximum* of both *minima.* The minimal payoff a PD player may obtain when choosing to cooperate is the S payoff, and the minimal payoff when choosing to defect is the P payoff. Since P > S, defection assures avoiding the minimal possible payoff.

(iii) MaxiMax strategy – an optimistic perspective that takes into account the maximal payoffs that may be obtained under each alternative, therefore choosing the alternative that contains the *maximum* of both *maxima*. The maximal payoff a PD player may obtain if choosing to cooperate is the R payoff, and the maximal payoff if choosing to defect is the T payoff. Since T > R, defection enables obtaining the maximal possible payoff.

(iv) Nash equilibrium - an outcome (i.e., a cell in the matrix) which no party is motivated to leave unilaterally, if assuming the other party does not change its choice (Nash, 1950). Out of the four cells of the PD game the only cell where no player is motivated to change his or her chosen alternative is the intersection of both defective alternatives, where both players obtain the P payoff.

Rapoport and Guyer's Taxonomy of Two-by-Two Games

Rapoport and Guyer (1966) classified all *strictly ordinal* two-by-two games into ten categories based on several strategic properties. These properties include dominance (an alternative that is strictly better than the other alternative for a player, regardless of the other player's choice), MaxiMin (a strategy that considers only the minimal payoff under each alternative and chooses the alternative that contains the maximum of these minima), the Nash equilibrium (an outcome that none of the players is motivated to abandon unilaterally, assuming the other player does not change his/her choice), and the Pareto equilibrium (no other outcome of the game can provide a better payoff for *both* players). Applying these criteria in a specific order constitutes an end-state referred to as the *natural outcome* of the game. To derive the natural outcome, Rapoport and Guyer proposed the following procedure:

1. If both players have a dominant strategy, they both choose it, and the resulting outcome constitutes the natural outcome of the game.
2. Else, if only one player has a dominant strategy, he/she chooses it, while the other player chooses the strategy that maximizes his/her own payoff under the expectation that the dominant strategy is chosen. The resulting outcome constitutes the natural outcome of the game.
3. Else, if the game has a single Pareto equilibrium, each player chooses the strategy containing this equilibrium. The resulting outcome constitutes the natural outcome of the game.
4. Else, if no player has a dominant strategy and there is either no Pareto equilibrium or more than one Pareto equilibrium, each player chooses his/her MaxiMin strategy. The resulting outcome constitutes the natural outcome of the game.

The natural outcome of the game along with other considerations, determine a specific category for each and every *strictly ordinal* game. This classification provides a *theoretical* model that makes it possible to distinguish between ten different interaction types, and may also be considered as a plausible forecast of players' behavior. As mentioned above, this taxonomy leaves out games that are non-strictly ordinal, which are games where at least two cells provide the same payoff for at least one of the players.

Expanding Rapoport and Guyer's taxonomy to include non-strictly ordinal games

Since some natural conflicts cannot be represented by strictly ordinal games, we further expand the natural outcome algorithm to also account for *non-strictly ordinal* games. To this end we modify the algorithm and add a category of games that do not have a natural outcome. The revised algorithm is applicable to all two-by-two games and makes it possible to understand and classify a wide range of ecologically valid conflicts. The modifications of the natural outcome algorithm are as follows:

1. If either one or both of the players obtains identical payoffs under both alternatives (i.e., the same payoffs under each of the opponent's choices) and therefore no strategic decision can be made by one or both of the players, the game is classified as a game with *no natural outcome*.
2. Non-strictly ordinal games may have weak dominance, which is an alternative that provides an identical outcome under one of the opponent's choices yet a better outcome under the opponent's other choice (which is impossible in strictly ordinal games). Therefore, our definition accounts for both weak and strong dominance.
3. Since non-strictly ordinal games may have identical minimal payoffs in both rows or both columns, the players do not necessarily have a MaxiMin strategy. Therefore, if only one player has a MaxiMin strategy, he/she will choose accordingly, and the other player will play under the assumption that the first player is choosing the alternative that contains the MaxiMin payoff. This criterion is in line with Rapoport and Guyer's second criterion, which states that if only one player has a dominant alternative, the other player chooses under the assumption that this alternative is chosen.
4. If neither player has a dominant or MaxiMin strategy, the players choose according to the *MiniMax regret* strategy (a strategy that minimizes the maximal potential loss of a player across all of the opponent's choices; Loomes & Sugden, 1982). The reason for this is that while the player cannot avoid the worst outcome, he/she can still avoid the worst regret. Note that the choice of the MiniMax regret for matrices with no MaxiMin is identical to the choice of the MaxiMax strategy (a strategy that considers only the maximal payoff

under each alternative and chooses the alternative that contains the maximum of these maxima).

5. If only one player has a MiniMax regret strategy, he/she will choose accordingly, and the other player will play under the assumption that this choice is made.
6. If neither player has a dominant strategy, a MaxiMin strategy, and a MiniMax regret strategy, the game is classified as a game with *no natural outcome*.

The Revised and Reduced Taxonomy Applied in the Present Study

Our reduced taxonomy encompasses the following five categories: (1) *Absolutely Stable* games, as defined by Rapoport and Guyer (1966). In these games, both players obtain their maximal payoff and are thus satisfied with the natural outcome. Such games are regarded as no-conflict games. (2) *Stable/Strongly Stable* games, as defined by Rapoport and Guyer. In these games, either one player or both players are not satisfied with the natural outcome, but the unsatisfied player/s are not motivated to try and change the outcome of the game, since shifting or threatening to shift their initial choice neither improves their own expected payoff nor motivates the other player to shift as well. (3) *Non-Stable* games include: unstable, force-vulnerable, threat-vulnerable, two equilibria with equilibrium outcome, two equilibria without equilibrium outcome and cyclic games (also termed games without equilibria), as defined by Rapoport and Guyer. In these games, either one or both players is not satisfied with the natural outcome, but unlike in the Stable/Strongly Stable games, the player is motivated to try to change the outcome of the game by shifting or threatening to shift his/her initial choice. However, any shift or threat motivates the other player to reply with a threat or shift of his/her own. (4) *Prisoners' Dilemma (PD) - like* games are 'strongly stable with deficient equilibrium' games, as defined by Rapoport and Guyer. In these games both players are not satisfied with the natural outcome and are not motivated to try and change the outcome. However, unlike Stable/Strongly Stable games, the natural outcome is the only cell in the matrix which is not a Pareto equilibrium (there is another cell that is more beneficial for both players). Note that while the classical PD game, depicted in Figure S1, requires that T>R>P>S, the extended set of PD-like games allows for T=R for one of the players. (5) *No Natural Outcome* games, as defined above. In these games none of the abovementioned criteria provide strategic guidance for the players.

SERS-Based Taxonomy of Games

The SERS-based taxonomy shifts the focus from the payoff structure per se to the interaction between (i) the game's payoff structure and (ii) the players' strategic perceptions of their opponent. Specifically, their prospects of choosing identical (or opposing) strategies, as driven by their perception of the *strategic similarity* with the opponent. The theory of Subjective Expected Relative Similarity (SERS; Fischer, 2009) computes an Expected Value (EV) that integrates (i) the payoffs expected under each choice, and (ii) the similarity perception as expressed by the probability of the opponent choosing an alternative that is identical to (or different from) the one selected by oneself. Comparing the resulting EVs allows players to choose the alternative that maximizes their expected payoffs when facing a specific opponent. For example, consider two players choosing their alternatives while interacting in a PD game (Figure S1). A player that assumes the other player is likely to choose the same alternative with the probability of p_s (and the other alternative with a probability of $1-p_s$) may compare the EV for the choice of cooperation with the EV for the choice of defection, where EV(cooperation) = $Rp_s + S(1-p_s)$, and EV(defection) = $Pp_s + T(1-p_s)$, and choose the alternative that provides the higher EV. SERS assumes that the strategic similarity between the players is subjectively and individually perceived. Therefore, two players confronting each other may have identical or different perceptions of their similarity to the opponent. In other words, the row player's p_s does not necessarily equal the column player's p_s.

The solution provided by SERS is applicable not only to the PD game, but to many other games. Games in which the SERS-based expected choice *varies* under different perceptions of strategic similarity with the opponent are referred to as Similarity-Sensitive games, whereas games in which the SERS-based expected choice *does not vary* under different perceptions of strategic similarity with the opponent are referred to as Non-Similarity-Sensitive games. Some games can be similarity-sensitive for one of the players, and non-similarity-sensitive for the other. Therefore, the SERS-based taxonomy of games differentiates between *two-player similarity-sensitive games*, *one-player similarity-sensitive games,* and *two-player non-similarity-sensitive games*. Figure S1 depicts examples of a two-player similarity-sensitive game (panel a), a two-player non-similarity-sensitive game (panel b), and one-player similarity-sensitive games (panel c – only similarity-sensitive for the row player; panel d – only similarity-sensitive for the column player).

To easily classify the games according to their sensitivity to similarity, one may compare the EV-maximizing choices under both assumptions of complete strategic similarity ($p_s = 1$) and complete strategic dissimilarity ($p_s = 0$) with the opponent. Each is reflected by one of the matrix's diagonals: the diagonal in which both players choose the same alternative (Aα and Bβ in Figure S1) and the diagonal in which they choose the opposite alternatives (Aβ and Bα in Figure S1). If the row player obtains the higher payoff while choosing a different row under both assumptions (i.e., [Aα > Bβ and Aβ < Bα] or [Aα < Bβ and Aβ > Bα]), the game is similarity-sensitive for the row player. However, if the row player obtains the higher payoff while choosing the same row under both assumptions (i.e., [Aα > Bβ and Aβ > Bα] or [Aα < Bβ and Aβ < Bα]), the game is non-similarity-sensitive for the row player. Clearly, the same considerations apply also for the column player.

a

	α	β
A	10 , 10	5 , 20
B	20 , 5	0 , 0

b

	α	β
A	20 , 20	10 , 5
B	5 , 10	15 , 15

c

	α	β
A	10 , 20	5 , 5
B	20 , 10	0 , 15

d

	α	β
A	20 , 10	10 , 20
B	5 , 5	15, 0

Figure S1. Examples of the three classes of the SERS-based taxonomy of games. Panel **a** depicts a two-player similarity-sensitive game. Panel **b** depicts a two-player non-similarity-sensitive game. Panel **c** depicts a one-player similarity-sensitive game that is similarity sensitive only for the row player. Panel **d** depicts a one-player similarity-sensitive game that is similarity sensitive only for the column player.

References

Loomes, G., & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, *92*(368), 805. https://doi.org/10.2307/2232669

Nash, J. F. (1950). Equilibrium points in n -person games. *Proceedings of the National Academy of Sciences*, *36*(1), 48–49. https://doi.org/10.1073/pnas.36.1.48

Rapoport, A., & Guyer, M. (1966). A taxonomy of 2x2 games. *General Systems*, *11*, 203–214.