Discussion Paper No. 1266

Strategic Polarization

by
Adam Kalai
and
Ehud Kalai

July 16, 1999

http://www.kellogg.nwu.edu/research/math

Strategic Polarization

Adam Kalai
Carnegie Mellon University

Ehud Kalai
Northwestern University

July 16, 1999

Abstract

In joint decision making, people with common goals and similar preferences often take drastically opposing positions. In some cases, arbitrarily small discrepancies in preferences result in arbitrarily large losses in utility for all participants. An understanding of the properties of polarization may help game players and mechanism designers avoid its pitfalls.

1 Introduction

In public debates, private organizations, shared living accommodations, and many other types of interaction, participants' behavior is seen to polarize. Even moderate individuals, with similar preferences often take extreme, opposing positions in these settings. To add to the frustration, such polarization can be quite costly: arbitrarily small disagreements in players' preferences may lead to arbitrarily large losses in utility for all participants. However, in many situations, costly polarization can be avoided by changing the rules of interaction.

This note presents a simple game-theoretic analysis that explains some of the reasons for polarization[1]. Despite the apparent game-theoretic simplicity,

[1] One can imagine several other types of polarized behavior. For example, in psychology, "group polarization" refers to a phenomenon in which all members of a group support the *same* decision more emphatically than they would as individuals because of an unconscious desire for unanimity. [1]

seemingly rational individuals design or choose to play games that lead to polarization and utility loss. This suggests that polarization is also subtle and easily overlooked. Hopefully, a better understanding of this phenomenon will enable mechanism designers and game players to avoid its pitfalls.

1.1 Example: Public Debate Polarization

Imagine two people involved in a debate over the severity of an illegal act committed by a politician. At the end of the debate, the average listener has formed an opinion, and each debater would like this average to be close to her own opinion. We view the debate as a two-person simultaneous move game in which each debater advocates a position and the average listener forms an opinion which is at the point midway between the two positions.

Specifically, say the severity of the act can be described by a real number from the interval [0, 1]. Further, suppose the players have opinions represented as real numbers $r_1, r_2 \in [0,1]$, and advocate positions $p_1, p_2 \in [0,1]$. The average listener forms an opinion $a = (p_1 + p_2)/2$. The debaters, having utility functions $u_i(a) = -|a - r_i|$, would like the average to be close to their own opinion.

Case 1. Suppose they agree on the severity $r_1 = r_2 = 0.6$. It is easy to see that when both players advocate their true beliefs $p_1 = p_2 = 0.6$, they are at a Nash Equilibrium.

Case 2. Suppose the players have a small discrepancy in their beliefs, with $r_1 = 0.60$ and $r_2 = 0.61$. In this case, there is a unique Nash Equilibrium at which $p_1 = .20$, $p_2 = 1.00$, and the average listener's opinion is $a = 0.60$. In this configuration, none of the of the players can improve the outcome of the debate relative to her own preferences. Player one has achieved her largest possible payoff, and player two is already pushed to an extreme position and cannot go farther. Furthermore, every other configuration is unstable because one of the debaters will be able to unilaterally change her position to move the average towards her own opinion thereby increasing her payoff.

This type of game models many situations in addition to public political debates. For example, it can model the polarization of parents disciplining a child, where the resulting discipline is the sum not the average of the levels imposed by the parents. Similarly, roommates may polarize over how clean the bathroom should be, with the "cleaner one" cleaning up for both.

Business partners polarize over the conduct of their affairs, and politicians polarize over the conduct of the affairs of a nation.

The next example illustrates that polarization can be quite costly, and the following section shows that polarization extends significantly beyond cases of averaging or summing individual choices.

1.2 Example: Costly Household Polarization

Consider a weekly work-consumption-savings plan of a household consisting of two individuals. Viewed as a strategic game, each of the players, $i = 1, 2$, decides on his own work level w_i and his level of consumption of a frivolous good f_i, both measured in dollar units with $0 \leq f_i$ and $0 \leq w_i \leq 1600$. The income not consumed, $b = (w_1 + w_2) - (f_1 + f_2)$ is deposited as savings in the bank (b could be negative).

Case 1. Suppose both participants have the following identical utility function,

$$u_i = 2\min(b, 500) - (w_1 + w_2) + 0.1(f_1 + f_2).$$

In words, every dollar deposited in the bank yields twice as much utility as the disutility of earning it, up to a maximum of \$500 in the bank. On the other hand, frivolous consumption results in a net loss of \$0.90 per \$1.00 consumed. Each player has equal utility and disutility for his and his partner's consumption and work. In this case, when $w_1 = w_2 = 250$ and $f_1 = f_2 = 0$, the players are at a a Nash Equilibrium with equal positive utilities of 500.

Case 2. Now suppose the players almost agree, with,

$$\begin{aligned} u_1 &= 2\min(b, 500) - (w_1 + w_2) + 0.1(f_1 + f_2) \\ u_2 &= 2\min(b, 500 + \epsilon) - (w_1 + w_2) + 0.1(f_1 + f_2). \end{aligned}$$

In this case, the only Nash Equilibrium has $w_1 = 0$, $w_2 = 1600$, $f_1 = 1100$, $f_2 = 0$, and $u_1 = u_2 = -490$.

Proof. At Equilibrium we cannot have $b > 500$, because player 1 could increase his utility by increasing f_1. We cannot have $b < 500$ because one of the players could reduce f_i or increase w_i. Thus at Equilibrium $b = 500$. But if $w_2 < 1600$ then player 2 can improve by increasing w_2. So at Equilibrium we must have $b = 500$ and $w_2 = 1600$. It is easy to see that $f_2 > 0$ is suboptimal for player 2, then, and that $w_1 = 0$ and $f_1 = 1100$.

2 Polarization in aggregation games

In both of the above examples, an aggregate quantity determines the players' utilities. At all the Nash Equilibria, each player is either completely satisfied, meaning that no outcome could improve her utility, or is polarized, meaning that she is playing a strategy on the boundary of her feasible set of strategies. This is the polarization phenomenon.

We proceed to present simple conditions on the aggregation and utility functions that result in this phenomenon. Informally, the condition on the aggregation function is that any individual player, by changing his own strategies, can move the aggregate value in any direction (specifically in some open set), provided that player is not constrained by his own individual limitations, i.e. playing a strategy on the boundary of his feasible set. Also, each player's utility function must have no local maxima (that aren't also global maxima). Under these two conditions, we argue that, if a player is not completely satisfied, then she is not at a global or local maximum. Furthermore, if she is playing a best response, then she must actually be at a boundary strategy, otherwise she could move the aggregate to increase her utility. We formalize this argument as follows.

Player i in $\{1, 2, ..., n\}$ has a feasible set of strategies S_i. Let $S = \times_i S_i$ denote the set of strategy profiles. An aggregating function AGG: $S \longrightarrow A$ selects an outcome a from a set A for every strategy profile s. Each player i has a utility function u_i: $A \longrightarrow \mathcal{R}$ describing his preferences over the selected outcome. We assume only that the sets S_i and A are subsets of abstract topological spaces, but, in all of our examples, they are subsets of Euclidean spaces.

An aggregation game consists of the simultaneous selection of individual strategies where players' payoffs are computed through the realized outcomes. With an abuse of notation, we denote this by $u_i(s) = u_i(\text{AGG}(s))$.

Next, we would like to capture the notion that every player not limited by his own feasibility constrains, i.e. not playing a boundary strategy, can move the aggregate value within some neighborhood of its current value. Formally, the *range of influence* of player i at $s_{-i} = (s_1, s_2, ..., s_{i-1}, s_{i+1}, ..., s_n)$ is $\text{AGG}(s_1, ..., s_{i-1}, S_i, s_{i+1}, ..., s_n) \subseteq A$. An *individually responsive* aggregating function AGG has the property that, for every player i and strategy profile s, if AGG(s) is on the boundary of i's range of influence at s_{-i}, then s_i is on the boundary of i's feasible set of strategies S_i.

A utility function has *no local maxima* if every local maximum is actually a global maximum. That is, if $u_i(a)$ is a maximum of u_i over a neighborhood of a, then it is a maximum over A.

Finally, player i is *completely satisfied* with a strategy profile s if $u_i(\text{AGG}(s))$ is a maximum of u_i over A. Player i is *polarized* at a strategy profile s if s_i is on the boundary of S_i. Based on these definitions, we have,

Polarization Lemma. *In any Nash Equilibrium of an aggregation game with an individually responsive aggregating function and utility functions with no local maxima, every player who is not completely satisfied is polarized.*

Proof: Suppose not. Say we have a Nash Equilibrium with player i neither fully satisfied nor playing a boundary strategy. Let s be the strategy profile and R_i be i's range of influence at s_{-i}. Since s_i is not on the boundary of i's feasible set of strategies and AGG is individually responsive, $\text{AGG}(s)$ is not on the boundary of R_i. Thus, R_i is a neighborhood of $\text{AGG}(s)$. Furthermore, since i is not completely satisfied and u_i has no local maxima, $u(\text{AGG}(i))$ is neither a global nor local maximum. This means that there must be some a_i' in R_i with $u_i(a_i') > u_i(\text{AGG}(s))$. Since a_i' is in player i's range of influence, s_i is not a best response and we have a contradiction.

Remarks:

1. The scope of the lemma. The sufficient conditions used in the polarization lemma are quite general. Individually responsive aggregation functions, as described by the general topological property above, include many aggregation methods other than those obtained by adding or averaging individual positions. A similar observation is true for the no local maximum condition. Clearly, if the individual utility functions are concave, then the condition holds. But the lemma also holds for the case that the utility functions are convex, because the maxima would be boundary strategies, as well as for a large number of other mixed cases.

2. Mixed strategies. The polarization lemma holds for pure strategy Nash Equilibria but not necessarily for mixed strategies. Consider the public debate example where each player has a target value of 0.6. A mixed strategy Equilibrium exists where each player chooses either 0.5 or 0.7

with equal probability. In this case, neither player is polarized, and neither is completely satisfied. However, if each player's utility is a strictly concave function of his own strategy (keeping the opponents fixed), it is easy to see that there are only pure strategy Nash equilibria, and thus the polarization lemma applies in general.

3. General best response. The polarization lemma can be applied to an individual player in a game. Specifically, if one player's utility function has no local maxima and the aggregation function is individually responsive to her, then any of her best response strategies will result in her full satisfaction or polarization, for any choice of strategies for her opponents. This observation may be useful in studying other best response based notions, such as rationalizability and Cournot best response dynamics.

4. Avoiding polarization. In our examples, polarization can be avoided by slightly modifying the games. For example, in the costly polarization example, either player would be better of by simply delegating his choice of strategy to the other player.

References

[1] Griffin, E. M. (1997) A first look at communication theory New York: McGraw-Hill.