Subjective Utilitarianism: Individual decisions in a social context – Supplementary appendices –

Shiri Alon* and Ehud Lehrer†

March 9, 2020

A Measuring others' influence

In the subjective utilitarianism model, a preference of a decision maker over lotteries may be reversed when the decision maker consumes those lotteries with others. Given such influence of other individuals, a natural comparative question that arises pertains to the degree of influence each of these others has. To answer this question, the notion of 'influence' needs to be precisely defined. Generally speaking, we consider 'influence' to be a case when a significant other causes a reversal of the decision maker's choice. Thus, for a fixed pair of lotteries p and q, a referent other influences the decision maker whenever the decision maker chooses q over p without that referent other, yet reverses this preference to p over q with that same referent. This kind of reversal can occur when a referent other joins only the decision maker, or when the preference of the decision maker in the context of a group is reversed once that referent other joins the group. Instances of influence across different pairs of lotteries are combined by taking the average over all possible pairs (p, q).

To capture reversal of preferences given fixed lotteries p and q, and to facilitate the measurement of others' influence, a simple cooperative game is defined:

For every group $S \subseteq I$,

$$w_{p,q}(S) = \begin{cases} 1 & (p, S) \succsim (q, S) \\ 0 & \text{otherwise} \end{cases} .$$

Let $v_1, \ldots, v_N$ be the subjectively ascribed utilities from Theorem 1. A referent other $i \in I$ may swing a coalition from being a losing coalition in $w_{p,q}$ to being a winning

*Bar-Ilan University, Ramat-Gan 5290002, Israel. e-mail: Shiri.Alon-Eron@biu.ac.il

†The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel and INSEAD, Boulevard de Constance, 77305 Fontainebleau, France. e-mail: lehrer@post.tau.ac.il

coalition, if $v_i(p) > v_i(q)$, and may swing a coalition from winning to losing if a strict inequality in the other direction holds. The Banzhaf value[1] of each player $i \in I$ in the above game may be computed as follows:

$$\beta_i(w_{p,q}) = \frac{1}{2^{|N|-1}} \sum_{S \subset I \setminus \{i\}} [w_{p,q}(S \cup \{i\}) - w_{p,q}(S)] \ .$$

Since a player i with $v_i(q) > v_i(p)$ can swing coalitions from being winning to being losing, the Banzhaf value of players may be negative. However, we are interested in an influence of significant others, no matter in which direction of preference. Moreover, a player who gains a negative Banzhaf value in $w_{p,q}$ will gain a positive Banzhaf value in the symmetric game $w_{q,p}$. In order to measure influence per se, without indicating in which direction of preference it takes place, and since the measure we aim at will eventually average over all pairs of lotteries p and q, we define:

$$B_i(p,q) = \max\left(\beta_i(w_{p,q}), 0\right) \ .$$

Altogether, the influence of referent individual i on the decision maker is the average,

$$B_i = \int_{(p,q)} B_i(p,q) d\lambda,$$

where λ is the Lebesgue measure over Y^2. Note that by taking this average, swings are counted whenever they occur (either for p over q or the other way around).

We lastly show that the influence measure defined is sub-additive in the following sense: consider a decision problem derived from the original one by amalgamating two referent individuals into a single individual, whose utility is the sum of the two referents' utilities. Then the influence of the amalgamated individual on the decision maker can never be more than the sum of influences of the two separate referent individuals. Formally, let $i, j \in I, i \neq j$. For a pair of lotteries p and q, define the game $\bar{w}_{p,q}$ in which i and j are amalgamated into one player $\bar{ij}$ by,

$$\bar{w}_{p,q}(S) = \begin{cases} w_{p,q}(S) & \bar{ij} \notin S \\ w_{p,q}(S \cup \{i,j\} \setminus \{\bar{ij}\}) & \bar{ij} \in S \end{cases}$$

for every $S \subseteq I \setminus \{i,j\} \cup \{\bar{ij}\}$. Denote $\bar{B}_{\bar{ij}}(p,q) = \max\left(\beta_i(\bar{w}_{p,q}), 0\right)$, and $\bar{B}_{\bar{ij}}$ the corresponding average over pairs of lotteries p and q. Then,

Proposition 1. $\bar{B}_{\bar{ij}} \leq B_i + B_j$.

Proof of Proposition 1. It is shown that sub-additivity as in the proposition holds for every pair of lotteries p and q, for every marginal contribution to a group $S \subseteq I \setminus \{i,j\}$.

[1]The Banzhaf value of player i in a cooperative game is the expected contribution of i to a random coalition that does not contain i.

Fix such p, q, and S. Note first that if both $v_i(q) > v_i(p)$ and $v_j(q) > v_j(p)$, then neither i, j nor $\bar{ij}$ can contribute to any group, as only positive contributions are accounted for, and the inequality trivially holds. Now examine the case in which both $v_i(p) > v_i(q)$ and $v_j(p) > v_j(q)$. If $\bar{ij}$ contributes zero to S (namely, $\bar{ij}$ does not swing S) then the inequality for p, q and S is trivially true. Otherwise, if $\bar{ij}$ swings S, then the corresponding marginal contribution to $\bar{B}_{\bar{ij}}(p,q)$ is $\frac{1}{2^{|N|-2}}$. It is shown that the sum of marginal contributions, of i to S and to $S \cup \{j\}$, and of j to S and to $S \cup \{i\}$, is also $\frac{1}{2^{|N|-2}}$. If both $w_{p,q}(S \cup \{j\}) = 0$ and $w_{p,q}(S \cup \{i\}) = 0$, then i swings $S \cup \{j\}$ and j swings $S \cup \{i\}$, and both do not swing S. Therefore each marginally contributes $\frac{1}{2^{|N|-1}}$, adding up to $\frac{1}{2^{|N|-2}}$. Otherwise, if both $w_{p,q}(S \cup \{j\}) = 1$ and $w_{p,q}(S \cup \{i\}) = 1$ then each of i and j swings S, and none swings S with the other, so that their marginal contributions are again $\frac{1}{2^{|N|-1}}$ each and sum to $\frac{1}{2^{|N|-2}}$. If $w_{p,q}(S \cup \{j\}) = 1$ and $w_{p,q}(S \cup \{i\}) = 0$ then i does not swing S nor $S \cup \{j\}$, but j swings both S and $S \cup \{i\}$, twice delivering a contribution of $\frac{1}{2^{|N|-1}}$ hence altogether $\frac{1}{2^{|N|-2}}$.

If $v_i(p) > v_i(q)$ and $v_j(q) > v_j(p)$, then only i can be a swinger. If $\bar{ij}$ contributes zero (namely, $\bar{ij}$ does not swing S) then the inequality for p, q and S holds trivially. If $\bar{ij}$ swings S, yielding a marginal contribution of $\frac{1}{2^{|N|-2}}$, then it must be that i swings both S and $S \cup \{j\}$ (as j only adds to the desirability of q over p, having $v_j(p) - v_j(q) < 0$). Hence i contributes a total of $\frac{1}{2^{|N|-2}}$. The symmetric case, switching i and j, is analogue.

The above is true for every $S \subseteq I \setminus \{i,j\}$, therefore for every pair of lotteries p and q, $\bar{B}_{\bar{ij}}(p,q) \leq B_i(p,q) + B_j(p,q)$, and the proof for the average follows. ■

The proposition states that the influence of two individuals who join forces and always come together, reduces compared to their total influence when they are separated. The intuition for the proposition is quite simple. Recall that a subjective utilitarian decision maker is considerate of the welfare of each significant other individually. Namely, the personal tastes of each such other are always taken under advisement, and to the same extent, regardless of the group to which this referent other joins. As a result, the only non-additive effect on influence of uniting two individuals together is when their tastes are opposite, and so their individual influences cancel out when they are considered together. This is the effect described in the proposition.

B When do subjective utilities equal true utilities?

The model presented in the paper focuses on the preferences of a single individual, asking how these change as a function of the group with which the individual consumes. The expression of others' tastes in the representation is purely subjective, namely, it represents the decision maker's perception of their tastes, rather than their actual tastes. Put differently, observed decisions of an individual may be based on misperceived preferences of others. Nonetheless, in cases where others' actual preferences may be observed, it is interesting to understand when perceived and actual preferences coincide. Perhaps not surprisingly, such coincidence is essentially a result of a Pareto-type condition.

Denote by $\succsim^i$ the (true) preferences of individual $i \in I$ over Y, with symmetric and asymmetric components $\sim^i$ and $\succ^i$, respectively. Suppose that these preferences conform to the vNM axioms, and assume that:

(a) For r^* and r_* the individualistic better and worse outcomes from assumption **B1**, $r^* \sim^i r_*$.

(b) There are $x^0, x_0 \in X$ such that both $x^0 \succ x_0$ and $x^0 \succ^i x_0$.

That is to say, individual i is indeed indifferent between the two individualistic outcomes of Individual Zero that appear in the structural assumption **B1**, and individual i and Individual Zero agree on some strict ranking of outcomes. Under these two assumptions, and supposing that the conditions of Theorem 1 hold, the utility of i as subjectively perceived by Individual Zero coincides with i's true utility, if and only if, whenever individual i and Individual Zero personally agree on a ranking of lotteries, this same ranking holds for the preferences of Individual Zero in the company of i. This is stated in the following proposition.

Proposition 2. *Let $\succsim^i$ be a binary relation over Y, represented by a vNM utility function. Let $\succsim$ be a binary relation over $Y \times 2^I$ that satisfies (ii) of Theorem 1, and denote by v_i the subjective utility ascribed by Individual Zero to referent individual i. Suppose that assumptions (a) and (b) above are satisfied. Then v_i represents $\succsim^i$, if and only if, for every two lotteries p and q, if $p \succ q$ and $p \succ^i q$, then $(p, \{i\}) \succ (q, \{i\})$.*

Proof of Proposition 2. Denote by u_i the vNM utility that represents $\succsim^i$ over Y. Suppose that (ii) of Theorem 1 holds, with v_i the vNM utility subjectively ascribed to individual i by Individual Zero.

Assume first that v_i represents $\succsim^i$. If both $p \succ q$ and $p \succ^i q$, then equivalently, $u_0(p) > u_0(q)$ and $v_i(p) > v_i(q)$, which immediately implies that $u_0(p)+v_i(p) > u_0(q)+v_i(q)$. Hence $(p, \{i\}) \succ (q, \{i\})$.

Now suppose that whenever $p \succ q$ and $p \succ^i q$, it also holds that $(p, \{i\}) \succ (q, \{i\})$. By this and (b) it follows that $u_{\{i\}} = \lambda_0 u_0 + \lambda_i u_i + \tau_i$, for $\lambda_0, \lambda_i \geq 0, \lambda_0 + \lambda_i > 0$ (see De Meyer and Mongin [2]). Normalizing $u_0(r_*) = u_{\{i\}}(r_*) = 0$ yields $\tau_i = -\lambda_i u_i(r_*)$, so that for r^* it holds that $u_{\{i\}}(r^*) = \lambda_0 u_0(r^*) + \lambda_i u_i(r^*) - \lambda_i u_i(r_*)$. Applying (a) and the fact that $u_{\{i\}}(r^*) = u_0(r^*)$ by **B1**, it follows that $\lambda_0 = 1$, so that $u_{\{i\}} = u_0 + \lambda_i u_i + \tau_i$. According to Desirable and Undesirable lotteries (**S3**), there are q^* and q_* such that $u_{\{i\}}(q^*) > u_0(q^*)$ and $u_0(q_*) > u_{\{i\}}(q_*)$, hence $\lambda_i \neq 0$. Finally, (ii) of Theorem 1 means that $u_{\{i\}} = u_0 + v_i$, therefore $u_0 + v_i = u_0 + \lambda_i u_i + \tau_i$, yielding $v_i = \lambda_i u_i + \tau_i$. Namely, v_i represents $\succsim^i$. ■

C Taking into account others' consideration of the decision maker

Bergstrom [1] describes two individuals, Romeo and Juliet, each taking into account the other's wellbeing when evaluating alternatives. In his model, the wellbeing of Romeo is affected by Romeo's own individual tastes over goods, as well as by Juliet's wellbeing. Juliet's wellbeing in turn is similarly determined by her own individual tastes and by Romeo's wellbeing. Thus, Romeo, when evaluating alternatives, takes under advisement Juliet's caring for him (and similarly for Juliet).

In contrast to Bergstrom's model, our model aims to describe a subjective utilitarian decision maker who aims to consider only the genuine individual tastes of others, given through their utilities when they consume alone (these two coincide when the conditions of Proposition 2 are satisfied). Namely, the decision maker we have in mind does not take under advisement others' consideration of her or his own individual tastes. Rather, the decision maker's individual tastes are accounted for only directly, through her or his individualistic utility, the utility from consuming alone. For example, when buying takeout for dinner with friends, a subjective utilitarian decision maker considers his or her own individualistic preferences, as well as the friends' individualistic preferences, over different types of foods. This is as opposed to taking under advisement friends' preference that the decision maker like the food as well.

The subjective utilitarian model offered in this paper can be studied as part of a dynamic feedback system, that allows for considerations as in Bergstrom's paper (see Friedkin and Johnsen [4] for this type of model for opinion change in a group). Suppose a group $I = \{1, \ldots, n\}$ of individuals, each endowed with her or his own individual vNM utility over lotteries in Y. In accordance with our model, each individual $i \in I$ is a subjective utilitarian, and we further assume that each subjective utility coincides with its true counterpart (as in the previous subsection). Let u_i denote i's own individual utility, with u the vector of utilities $(u_i)_{i \in I}$. For each individual $i \in I$, α_i is the weight that i places on others' utilities in general, where the specific weight on the utility of individual $j \neq i$ is w_{ij}, with $w_{ij} > 0$ and $\sum_{j \neq i} w_{ij} = 1$ (α_i is introduced for normalization, in order for the utility to not grow indefinitely over time). Suppose time periods $t = 1, 2, \ldots$, and assume that at time t, utilities are given by,

$$u_i^t = (1 - \alpha_i) u_i + \alpha_i \sum_{j \neq i} w_{ij} u_j^{t-1}, \ t = 1, 2, \ldots \ ,$$

or, equivalently, by,

$$u^t = (I - A) u + A W u^{t-1}, \ t = 1, 2, \ldots \ ,$$

where $A = [a_{ij}]$ is the matrix with $a_{ii} = \alpha_i$ and $a_{ij} = 0$ for every $i \neq j$, $W = [w_{ij}]$ is the matrix with weights w_{ij} in all entries $i \neq j$ and $w_{ii} = 0$, and $u_j^0 = u_j$. The utility of individual i at period t is therefore a weighted average of i's original individual tastes, u_i, and the others' tastes in the previous period, u_j^{t-1}. Individuals thus take under advisement others' caring for them, as individual j's utility u_j^{t-1} depends also on i's utility.

If $(I - AW)$ is nonsingular, there are limiting utilities u_i^∞, given by,

$$u^\infty = (I - AW)^{-1} u \ .$$

These utilities form a steady state, in which each utility u_i^∞ for an individual $i \in I$ is a weighted sum of i's own original utility, u_i, and the others' (steady-state) utilities u_j^∞. Thus, each individual takes under advisement both her or his original tastes, as well as the others' (eventual) tastes, which themselves depend on i's own tastes. For a related treatment see Fershtman and Segal [3].

References

[1] Bergstrom, T. (1989), "Puzzles: Love and spaghetti, the opportunity cost of virtue", *The Journal of Economic Perspectives*, 3, 165-173.

[2] De Meyer, B., and P. Mongin (1995), "A Note on Affine Aggregation", *Econonic Letters* 47, 177-83.

[3] Fershtman, H. and Segal, U. (2017), "Preferences and Social Influence", *American Economic Journal: Microeconomics*, forthcoming.

[4] Friedkin, N. E. and Johnsen, E. C. (1999), "Social influence networks and opinion change", *Advances in Group Processes*, 16, 1-29.