THE LIMITS OF EMPATHY

LOS LÍMITES DE LA EMPATÍA

Marco Monge*

RESUMEN

Este trabajo teórico busca, sobre un fundamento formal y como su título indica, establecer los límites de las necesidades humanas cuando "nos ponemos en los zapatos de otros". La forma de hacerlo es analizar la lógica detrás de la función de utilidad social. El artículo tiene como meta investigar la empatía, en contextos de incertidumbre y de certidumbre. Para el primer caso, proporciona una alternativa y prueba del teorema de agregación de Harsanyi y para lo segundo, demuestra la imposibilidad de incluir todas las transformaciones de una función de utilidad individual en una función de utilidad social (de ahí la imposibilidad de conocer las preferencias ciudadanas). Con el impacto de ambos contextos, explica –desde los sesgos encontrados en la economía conductual– la imposibilidad de crear una función de utilidad social en cualquier caso. Lógicamente, si el argumento se aplica a las necesidades humanas, quien quiera incorporar la felicidad de los otros en el bienestar propio puede extenderse a la habilidad de los políticos, para satisfacer las necesidades populares con la figura del dictador benevolente.

ABSTRACT

This theoretical work seeks, based on a formal foundation and as stated in the title, the human limitations when it comes to "putting ourselves in other's shoes." The way to do it is analyzing the logic of the social utility function. The goal is to research empathy both in contexts with uncertainty and in those without it. For the former, it provides an alternative and much simpler proof of Harsanyi's Aggregation Theorem and for the latter, it demonstrates the impossibility of including all the transformations of an individual utility function in a social utility function (hence, the impossibility of really knowing the preferences of citizens). With an impact on both, it argues - from the biases found by behavioral economics - the impossibility of creating a social utility function. Logically, if the reasoning applies to any human being who wants to incorporate the happiness of others into his own well-being, it can extend to the ability of politicians to satisfy popular needs with the figure of the benevolent dictator.

KEY WORDS: social preferences, micro-foundations, aggregate utility

JEL Code : I31

Fecha de recepción: 10 de marzo de 2021

Fecha de aceptación: 02 de abril de 2021

---

*Marco Vinicio Monge Mora is a student of economics at Universidad de Costa Rica. He has contributed to the microeconomic foundations of the quantity theory of money and has extensive work in opinion journalism. Resident of San Jose, Costa Rica. His email address is marcomonge10@gmail.com and his ORCID is https://orcid.org/0000-0003-1830-6174

This paper has been deeply enriched by Danilo Monge Acuña's feedback and Saúl Sánchez Arroyo's translation advices.

The capacity of being empathic could be summarized in the existence of a cardinal welfare ("social" utility) function that incorporates the welfare of our fellow men according to our ethical perspectives.

Proposition 1 (*Harsanyi's aggregation theorem*): If individual and "social" preferences over lotteries can be represented by a correspondent cardinal utility function, the representation of the social utility function, that satisfies Pareto indifference principle, is a weighted sum of individual utilities.

Proof: It is known that a Von Neumann–Morgenstern *-E[u(·)]-* preserves the preference order after a transformation *-v(·)-* if and only if it is a positive linear one. That is

$$v(\cdot) = a + bE[u(\cdot)] \text{ where } a, b \in \mathbb{R} \text{ and } b > 0 \tag{1}$$

Suppose that there is a population of a natural finite positive number *"N"* of individuals and each *"n"* of them is supposed to have a Von Neumann–Morgenstern utility function to represent their respective references

-$E[u_n(\cdot)]\ \forall n \in \{1, \dots, N\}$-.

In consequence, we can define a particular transformation such that, in (1) *"a"* would be 0 and all individual would be a particular *"b"* that we call $\alpha_n\ \forall n \in \{1, \dots, N\}$ such that $\sum_{n=1}^{N} \alpha_n = 1$.

The fact that $\alpha_n\ \forall n \in \{1, \dots, N\}$ needs to be positive, because "*b*" needs to be positive, transforms what Harsanyi calls Fleming's Postulate E (considering, with a positive ponderation, the opinion of all individuals) in a mathematical restriction instead of an ethical arbitrary one. He thought that "if we want a formal guaranty that no individual's utility can be given a negative weight, [...] we must add one more postulate", now we see that we don't. To authentically incorporate a person's preferences in a cardinal welfare function, we have no other option than incorporate them with positive weight[1] (1955, pp. 310-315)1.

$$v_n(\cdot) = \alpha_n E[u_n(\cdot)] \text{ with } \sum_{n=1}^{N} \alpha_n = 1\ \forall n \in \{1, \dots, N\} \tag{2}$$

Notice that they are themselves, Von Neumann–Morgenstern utility functions, too. And so, they preserve the preference order after another linear transformation. Then we can define a transformation, like in (1), for $v_1(\cdot)$ and v_2 with $a = 0$ and $b = 1$ (*id est*, having the same two functions). If we sum these two, we are going to get

$$v_{1+2}(\cdot) \equiv v_1(\cdot) + v_2(\cdot) \tag{3}$$

Since neither $v_1(\cdot)$ nor $v_2(\cdot)$ are being object of a transformation that alters their preference order over lotteries -intuitively, because each individual only has control of their own preferences, the utility functions of the other individuals are "constants" for them-, nor 1 nor 2 represent different judgements over uncertainty by their own, even when together they create a different one. Now we can apply the same logic to $v_3(\cdot)$.

$$v_{1+2+3}(\cdot) \equiv v_1(\cdot) + v_2(\cdot) + v_3(\cdot) \tag{4}$$

And, because all the utility functions are being part of linear transformations, even when they are creating a different preference order in conjunction, all of them keep their internal logic separately. It is clear that, if we keep in this fashion, we are going to get a cardinal social utility function -$V(\cdot)$- that, clearly, is unique up to positive linear transformations.

$$V(\cdot) = \sum_{n=1}^{N} \alpha_n E[u_n(\cdot)] \tag{5}$$

An affine combination compound of the particular utility functions of the people. Even though any positive constant could be multiplied or added to $V(\cdot)$, we could also return to the "original" form based on relative va-

1. This could lead to a moral implication about no discrimination.

lues in scale 1, with an analogous transformation. *Ergo*, again, what Harsanyi (1955) calls *Fleming's Postulate D* (in case the welfare of all other, but two, individuals is not being affected, their relative importance is the decision criterion) in a mathematical restriction instead of an ethical arbitrary one. And what Harsanyi calls *Fleming's Postulates A, B and C* are just restriction need to have a utility function (p.311).

That is, we have freed our theorem from the necessity of most ethical judgements. But we can prove the achievement of a pretty weak one, the *Strong Pareto principle*. Consider any two lotteries "l_x" and "l_y" such that

$$E[u_n(l_x)] \geq E[u_n(l_y)] \; \forall n \in \{1, \dots, N\} \tag{6}$$

We just need to multiply both sides of (6) by the respective ponderation and, then, sum on both sides to account for the whole population.

$$\alpha_n E[u_n(l_x)] \geq \alpha_n E[u_n(l_y)] \; \forall n \in \{1, \dots, N\} \tag{7}$$

$$\leftrightarrow$$

$$\sum_{n=1}^{N} \alpha_n E[u_n(l_x)] \geq \sum_{n=1}^{N} \alpha_n E[u_n(l_y)] \tag{8}$$

Clearly, if all $\alpha_n E[u_n(l_x)] \geq \alpha_n E[u_n(l_y)] \; \forall n \in \{1, \dots, N\}$, and if we have strict inequality of at least one person, the properties of sum are going to make $V(l_x)>V(l_y)$. Of course, $V(\cdot)$ also satisfies, Pareto indifference, weak Pareto and semi-strong Pareto properties. The existence of the function and the uniqueness (for discrete probabilities) of its form, as can be concluded from the previous reasoning, is provided, precisely, by the fact that only the positive linear arrangements can preserve the individual preferences characteristics.

But now we have to analyze if we can approach in the same way to the idea of an ordinal utility function. Before continuing with the case of social preferences without uncertainty, we must proof by construction a highly know result.

Lemma: Without uncertainty, a utility function $u(\cdot)$ expresses the same preferences as any other of its crescent monotonic transformations.
$f(\cdot):\mathbb{R} \rightarrow \mathbb{R}$.

Proof: Consider the first partial derivative of a utility function –that depends on an "*M*" dimensional vector which has as all of its entries the consumption of a commodity-, with respect to any merchandise "x_m" that the agent takes into consideration. This is the marginal utility.

$$\frac{\partial u(\cdot)}{\partial x_m}; m \in \{1, \dots, M\} \tag{9}$$

Now consider the Subjective marginal rate of substitution $SMRS_{j,k}^{u}$. The relative value that the individual gives to any pair of commodities "x_j" and "x_k" is the coefficient between their marginal utilities.

$$SMRS_{j,k}^{u} \equiv \frac{\frac{\partial u(\cdot)}{\partial x_j}}{\frac{\partial u(\cdot)}{\partial x_k}}; j, k \in \{1, \dots, m\} \tag{10}$$

Apply the same operations above a crescent monotonic transformation (that is, with a positive first derivative) over the utility function -$f[u(\cdot)]$-.

$$\frac{\partial f[u(\cdot)]}{\partial x_m} = f' \frac{\partial u(\cdot)}{\partial x_m} \begin{cases} > 0 \leftrightarrow \frac{\partial u(\cdot)}{\partial x_m} > 0 \\ = 0 \leftrightarrow \frac{\partial u(\cdot)}{\partial x_m} = 0 \; \forall m \in \{1, \dots, M\} \\ < 0 \leftrightarrow \frac{\partial u(\cdot)}{\partial x_m} < 0 \end{cases} \tag{11}$$

$$SMRS_{j,k}^{f[u]} \equiv \frac{f' \frac{\partial u(\cdot)}{\partial x_j}}{f' \frac{\partial u(\cdot)}{\partial x_k}} = \frac{\frac{\partial u(\cdot)}{\partial x_j}}{\frac{\partial u(\cdot)}{\partial x_k}} \equiv SMRS_{j,k}^{u}; j, k \in \{1, \dots, m\} \tag{12}$$

Briefly, the agent's judgements about whether a commodity is a good, a bad or a neutral commodity, are not altered and the agent's relative judgements aren't altered either.

Let's think about an ordinal "social" utility function to which we do not impose any other restriction than depending only on all of the "*N*" individual utility functions that characterize the members of the population.

Proposition 2: The social utility $U(\cdot)$ function, , doesn't express the same preferences as any other positive monotonic transformation of its individual utility functions.

Proof: The social marginal utility of a commodity "x_m" is, because of the "chain rule", expressed by

$$\frac{\partial U(\cdot)}{\partial x_m} = \sum_{n=1}^{N} \frac{\partial U(\cdot)}{\partial u_n}\frac{\partial u_n}{\partial x_m} \quad (13)$$

Consequently, we can define the –worth the use of oxymoron- Subjective Social Marginal Rate of Substitution between any pair of commodities "x_j" and "x_k" as

$$SMRS_{j,k}^{u} \equiv \frac{\frac{\partial u(\cdot)}{\partial x_j}}{\frac{\partial u(\cdot)}{\partial x_k}}; j,k \in \{1,\dots,m\} \quad (14)$$

This is enough to see that *Samuelson's condition* of efficiency for public goods provision does not describe the equality between $SMRS_{j,k}^{U}$ and the Marginal Rate of Transformation, against what Maté & Pérez describe (2007, p. 169).

Now, if we apply a crescent monotonic transformation -$f_n(\cdot)\forall n \in \{1,\dots,N\}$- to each individual utility function, the new social welfare function $\tilde{U}(\cdot) \coloneqq U(f_1[u_1(\cdot)],\dots,f_N[u_N(\cdot)])$ should express the same preferences as $U(\cdot)$. But if we ask for the relative judgements of society

$$SMRS_{j,k}^{\tilde{U}(\cdot)} \equiv \frac{\sum_{n=1}^{N}\frac{\partial U(\cdot)}{\partial f_n}\frac{\partial f_n}{\partial u_n}\frac{\partial u_n}{\partial x_j}}{\sum_{n=1}^{N}\frac{\partial U(\cdot)}{\partial f_n}\frac{\partial f_n}{\partial u_n}\frac{\partial u_n}{\partial x_k}} \quad (15)$$

And we can see that, in general,

$$\frac{\sum_{n=1}^{N}\frac{\partial U(\cdot)}{\partial f_n}\frac{\partial f_n}{\partial u_n}\frac{\partial u_n}{\partial x_j}}{\sum_{n=1}^{N}\frac{\partial U(\cdot)}{\partial f_n}\frac{\partial f_n}{\partial u_n}\frac{\partial u_n}{\partial x_k}} \neq \frac{\sum_{n=1}^{N}\frac{\partial U(\cdot)}{\partial u_n}\frac{\partial u_n}{\partial x_j}}{\sum_{n=1}^{N}\frac{\partial U(\cdot)}{\partial u_n}\frac{\partial u_n}{\partial x_k}} \therefore SMRS_{j,k}^{\tilde{U}(\cdot)} \neq SMRS_{j,k}^{U(\cdot)}; \quad (16)$$

$$j,k \in \{1,\dots,m\}$$

Proposition 2 might be interpreted in several ways. The clearest one is that utility is a subjective measure. So it's wrong what Mas-Colell, Whinston & Green (1995) say about the possibility of obtaining a pure utilitarian welfare function, simply by convex transformations of individual utilities in a generalized utilitarian one. The analysis also applies for not differentiable welfare functions, like in a Rawlsian one, a monotonic transformation is all we need to change who is the poorest person in "utils" (pp. 827-829).

If we think about the welfare function as the one created by each individual concept of justice, Proposition 2 shows us the limits of empathy: we can't incorporate "properly" –i.e. read minds- the happiness –all the forms of their utility functions- of others into our judgement, we can simply incorporate the idea that we have of theirs –even more, it is known that the inverse function of a crescent function is also crescent, so we can think of $f_n^{-1}(\cdot)\forall n \in \{1,\dots,N\}$ as a transformation on $\tilde{U}(\cdot)$ and see that not even $U(\cdot)$ expresses accurately the personal preferences.

But if we think of the welfare function as the utility function of the society, Proposition 2 provides us with an alternative demonstration of *Arrow's impossibility paradox* and if we think of the welfare function as the utility function of the benefactor despot, Proposition 2 confirms the approximation (see Hayek, 1935), to *the Austrian socialism impossibility theorem*, that argues the incapacity of any government to accede completely to the tacit information and preferences that characterize the people.

With the concept of benefactor despot, we can consult behavioral economics to see that preference reversals appear in individuals in situations with or without uncertainty (see Thaler, 2016) and remember the public choice theory principle according to which, being the state the conjunction of individual wills it can't be wiser than the conjunction of their knowledge.

So it is trivial that, when individuals present preference reversals, it is impossible to construct nor a cardinal neither an ordinal utility function. *Von Neumann–Morgenstern's theorem* dictates a preference relation over lotteries can be represented by a cardinal utility function if and only if it is a "rational" preference relation and *Debreu's theorem* says the proper about preference relations over certain baskets and an ordinal utility function. Since under preference reversals, at least the *completeness axiom* of preferences is broken, the demonstration is immediate because we can't create a –welfare- function of –utility- functions if the latters don't exist.

According to the standard point of view, we can come up with an even simpler example that provides the same: if a human presents *the endowment effect*, the same basket would yield two different numbers of "utils" (depending on whether the individuals already own the basket or not). If we have two dependent results for the same independent variable, we don't have a function. And again, since a utility function doesn't exist for this individual, we can't create a social utility function that incorporates it.

This last reasoning ends up by enforcing the lesson of Proposition 2: politicians are also humans. Even with the best intentions, politicians suffer from biases that could make their labor inefficient –the trends to value excessively the achievements of the administration, to set proposals that are "bread for today and hunger for tomorrow" or to have highly optimistic expectations that then lead to resistance to change course- (Kahnemann, 2012, pp. 342-347).

LIST OF REFERENCES

Harsanyi, J. (1955). Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, *63*(4), 309-321.

Hayek, F. (1935). The present state of the debate. In: Hayek, F. (ed). *Collectivist economic planning* (1st ed.). Pp 201-243. London, UK: Routledge & Kegan Paul LTD.

Kahnemann, D. (2012). *Pensar rápido, pensar despacio*. Barcelona, España: Debols!llo.

Mas-Colell, A.; Whinston, M. & Green, J. (1995). *Microeconomic theory*. New York, United States: Oxford University Press.

Maté, J. & Pérez, C. (2007). *Microeconomía avanzada: cuestiones y ejercicios resueltos*. Madrid, España: Pearson Prentice Hall.

Thaler, R. (2016) *Todo lo que aprendido con la psicología económica*. Barcelona, España: Booket.