CEIS TOR VERGATA Research Papers

Università di Roma Tor Vergata

ISSN 2610-931X

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 23, Issue 5, No. 603 – June 2025

A Foundation for Universalisation in Games

Enrico Mattia Salonia

A Foundation for Universalisation in Games

Enrico Mattia Salonia*

May 27, 2025

Abstract

I study the behaviour of individuals who have preferences for universalisation. When considering a course of action, they evaluate the consequence that would occur if everyone else acted equivalently, according to some criterion of equivalence. That is, they universalise their behaviour. I develop and axiomatise a model for individuals who value their choices in light of the consequences they induce when their action is universalised. The key behavioural prediction is that the independence axiom is satisfied only among actions that are universalised equivalently. I impose conditions to single out the most prominent models of universalisation, compare them, highlight and arguably overcome their limitations. I propose a unifying model of universalisation inspired by the equal sacrifice principle.

JEL Classification: C70, C72, D01, D03, D90.

Keywords: Universalisation reasoning, Non-consequentialism, Kantian preferences, Equal sacrifice principle, Axiomatic model.

*Toulouse School of Economics. mattia.salonia@tse-fr.eu. I am indebted to Ingela Alger and François Salanié for countless discussions, comments, and invaluable guidance throughout various stages of this project. I also thank, in random order, Karine Van Der Straeten, Mostapha Diss, Alberto Grillo, Pau Juan Bartroli, Matteo Broso, Annalisa Costella, Philippe De Donder, Giacomo Rubbini, Peter Hammond, Franz Dietrich, two anonymous referees and participants at various workshops and conferences for helpful feedback. I acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 789111 - ERC EvolvingEconomics)

1 Introduction

What would I get if everyone acted as I do? An individual who acts based on the answer to this question exhibits universalisation reasoning. In group interactions, universalisation reasoning prescribes that individuals consider what would happen if everyone chooses the same action as them. Universalisation has been shown to have evolutionary foundations (Alger & Weibull, 2013) and aligns with behaviour observed in experiments (Levine et al., 2020; Miettinen et al., 2020; Van Leeuwen & Alger, 2024). Furthermore, it leads to desirable allocations under several normative criteria (Roemer, 2010).

Universalisation appears in the literature in various forms, with two prominent formulations being Homo Moralis preferences (Alger & Weibull, 2013) and the Kantian equilibrium concept (Roemer, 2019). Nevertheless, these models lack choice-theoretic foundations, complicating their unification and empirical testing. Without such foundations, extending the models beyond symmetric settings becomes challenging. It is unclear what "behaving in the same way" means in asymmetric contexts. Furthermore, the conceptual relationship between universalisation and other pro-social preferences remains unexplored. More worryingly, the models' predictions depend on the labels assigned to the primitive objects of choice, namely, actions in games. Universalisation prescribes considering what happens when everyone chooses the same action, therefore, changing the names of actions alters the predictions of these models. I show that developing choice-theoretic foundations for universalisation allows resolution of these issues.

I develop a model and introduce axioms that characterise preferences for universalisation. This characterisation enables the unification of previous models, rationalises existing empirical identification practices, and provides new testable predictions. I also introduce a new class of preferences for universalisation that are applicable to asymmetric settings. These preferences generalise the symmetric models, and their predictions are independent of the labelling of actions in games.

The main difficulty in modelling universalisation is that it is a non-consequentialist motivation. Preferences over actions do not depend on the material consequences these induce. Therefore, it is not straightforward to identify preferences for universalisation from choices over material consequences.[1] Economics is often resistant to considering non-consequentialist motivations (Fleurbaey, 2019). The classical models of Anscombe & Aumann (1963) and Savage (1972) illustrate this resistance. In these models, individuals rank mappings from uncertain states to consequences, usually referred to as "acts". Preferences for an act inducing a sure consequence are equivalent to preferences for that consequence. It is impossible to rank acts according to a criterion that does not depend on their induced consequences without trivializing such a notion, for example, by including the chosen act in the description of consequences. Thus, the question is whether universalisation, as a form of non-consequentialism, can be reconciled with the consequentialist

[1] Sen (1973) suggested that non-consequentialism poses a challenge for revealed preference theory.

approach of choice theory without resorting to ad hoc solutions.

I show that it is fruitful to study non-consequentialist decision criteria by taking a ranking over actions in a game, not consequences, as the primitive. An example in Section 1.1 illustrates that behaviour consistent with preferences for universalisation in a game cannot be rationalised by a preference ranking over material consequences. This motivates the use of Luce & Raiffa (1957)'s model, where the object of choice is an element of an action set. In a two-player game, an action induces an act, a mapping between the opponent's action and a consequence of the game. The novelty is that preferences over actions are not equivalent to preferences over the induced acts. In particular, the individual cares about the consequence that would obtain if his action is universalised. For instance, in a symmetric game, the individual considers what would happen if his opponent chooses the same action as he does. To capture more general criteria for universalisation, I introduce a universalisation function that maps an individual's action to an opponent's action, given a reference profile of actions. As an example, in Multiplicative Kantian Equilibrium (Roemer, 2019), individual actions that deviate from the reference profile by a specific proportion are universalised to opponent's action that deviates by the same proportion.

The main result, Theorem 1, provides a representation of preferences over mixed actions. The representation is a convex combination of two components. The first is the usual subjective expected utility. The second is the expected utility over the distribution of consequences obtained if the action is universalised. Therefore, preferences are a generalisation of subjective expected utility. The theorem allows to identify preferences over sure consequences up to the usual affine transformations and beliefs uniquely. The key deviation from expected utility is a violation of the independence axiom. In particular, independence holds only among actions that are universalised equivalently. Such violation of independence also constitutes a novel behavioural prediction. A direct test of a specific model of universalisation requires observing a violation of independence among actions that are universalised equivalently.

Because the theorem is silent on the shape of preferences over sure consequences, it reveals that universalisation and pro-social preferences are distinct assumptions; it is possible for an individual to exhibit both, consistent with empirical evidence (Van Leeuwen & Alger, 2024). Moreover, the theorem implies that welfare analysis for individuals with preferences for universalisation cannot use material consequences as a currency, contrary to the standard practice in Kantian Equilibrium models (Roemer, 2019). An individual with consequentialist preferences can always be compensated with material payoff, such as money, to refrain from taking a specific action. This is not true for individuals with preferences for universalisation, as they desire to induce a specific consequence as a result of their action being universalised. Non-consequentialist individuals thus suffer when they cannot choose the action they prefer, regardless of any material compensation. I thus argue that welfare criteria for non-consequentialist preferences may encompass a

form of freedom of choice.[2]

By specifying the universalisation function, I provide a choice-theoretic foundation for Homo Moralis preferences for universalisation à la Alger & Weibull (2013) and the various definitions of Kantian Equilibrium by Roemer (2019), both of which constitute a generalisation of the model by Laffont (1975). I comment on the difference between my foundation for Kantian Equilibrium and that of Roemer. He suggests that his model does not assume preferences that deviate from selfish material satisfaction, but rather a different "optimisation protocol". I argue that his model's properties can be preserved by abandoning the distinction between the optimisation protocol and preferences, resulting in a more parsimonious framework in line with classical choice theory.

I develop a novel concept of universalisation inspired by the equal sacrifice principle (Mill, 1885; Young, 1988). Consider an individual with any given aim. Given a profile of actions in a game, the individual evaluates a deviation by considering the consequence that would occur if their opponents also deviated to induce an equivalent difference in aim satisfaction, that is, an equal sacrifice. I show that this form of universalisation is equivalent to that of Homo Moralis and Kantian Equilibrium in symmetric games. Moreover, its predictions do not depend on the labelling of actions, nor does its definition require the veil of ignorance construct used to define Homo Moralis in asymmetric contexts.

The paper is organized as follows: In Section 2, I introduce the primitives of the model and the axioms. The main theorem is presented in Section 3. In Section 4, I show how assumptions on the universalisation function allow me to derive various models of universalisation. Equal sacrifice universalisation is introduced in Section 5. Section 6 concludes the paper. A literature review and illustrative example follow.

Related literature. In this paper, I study a decision problem as modelled in Luce & Raiffa (1957). The analysis builds on results by Battigalli et al. (2017), who study Luce & Raiffa decision problems and connect them to the approach in Anscombe & Aumann (1963).

The model here is reminiscent of context-dependent preferences by Gilboa & Schmeidler (2003). They study collections of individuals' preferences, one for each possible belief, over their actions and an uncertain state. As in this paper, the state is interpreted as opponents' choices. They also start from a primitive ranking over individuals' actions and obtain an expected utility representation in games. However, I here study a subjective beliefs setting, where these are derived from behaviour.

The intuition that non-consequentialist individuals do not care about an act because of its consequences has been highlighted by Chen & Schonger (2022), who develop a choice-theoretic model to guide an experiment testing for the presence of non-consequentialist preferences. They argue that, to identify non-consequentialism from choice, individuals must face the possibility that their actions will not be implemented or observed by the experimenter. Their model has a different interpretation compared to mine. In their

[2]See, for example, Fleurbaey (2008, Ch. 10).

experiment, subjects knew that there was a chance that their action would not have been implemented, whereas here there is no such possibility.

The model in this paper allows me to distinguish universalisation from the related concept of magical thinking, studied from a choice-theoretic perspective by Daley & Sadowski (2017). An individual exhibits magical thinking if he expects the probability the opponent selects a specific action to increase if he chooses that action. They provide axioms on behaviour in symmetric games that characterize magical thinking. I show that magical thinking and universalisation are different from a choice-theoretic perspective. An individual with preferences for universalisation does not believe he affects opponents' choice.

In this paper, I provide a choice-theoretic foundation for various models of universalisation. The two main alternatives are Homo Moralis preferences by Alger & Weibull (2013, 2016); Alger et al. (2020) and Kantian Equilibrium by Roemer (2010, 2015, 2019). In two-player games, Homo Moralis maximises a convex combination of his payoff and the payoff he would obtain if his opponent behaved as he does. The authors show that, among the set of continuous preferences, Homo Moralis is the only one that is evolutionary stable for all the game protocols their model covers, when interactions take place under incomplete information, and there is assortativity in the process. The result is generalised to multiplayer games and structured populations by Alger & Weibull (2016) and Alger et al. (2020). Roemer (2019) introduces a new solution concept, Kantian Equilibrium. He argues that, if individuals are Kantian rather than Nash optimisers, when considering deviating from an action profile they assume other players will deviate in an equivalent manner, where "equivalent" is defined in various ways. Alger & Weibull derive novel preferences from evolutionary analysis and Roemer changes the equilibrium concept, when compared with selfish/Nash individuals. I comment on the relation between these two models in the body of the paper.

This paper relates to studies of universalisation and other non-consequentialist motivations in various settings. Some of these study moral attitudes or their relation with pro-social preferences, as Dewatripont & Tirole (2024), Ellingsen & Mohlin (2024), Fleurbaey et al. (2023) and Laslier (2022). Others are applications in economic environments, including bargaining (Dizarlar & Karagözoğlu, 2023; Juan-Bartroli & Karagözoğlu, 2024), contract theory (Sarkisian, 2017, 2021a,b), public goods (Brekke et al., 2003), social norms (Juan-Bartroli, 2024), taxation (Sobrado, 2022), vaccination (De Donder et al., 2025) and voting (Alger & Laslier, 2022; Dierks et al., 2024; Grillo, 2022). Finally, there is interest in choice-theoretic models of individual moral attitudes. For example, Ponthiere (2024a), Ponthiere (2024b) and Shi (2024) study, respectively, Epictetianism, Stoicism and preference for a social minimum consumption level.

1.1 Illustrative Example

I briefly illustrate the contribution of this paper through an example. I show that preferences for universalisation cannot be rationalised by a preference ranking over material consequences, and that predictions of previous models depends on the labeling. I then discuss the solution I propose and how it relates to the existing literature.

Two individuals play the following game. They can go left (ℓ), middle (m) or right (r). The numbers in the table are monetary rewards.

μ'		$1/2$	$1/2$
μ	$1/2$	$1/2$	
	ℓ	m	r
ℓ	$1,1$	$0,0$	$0,0$
m	$0,0$	$0,0$	$1,1$
r	$0,0$	$1,1$	$0,0$

Table 1: Preference reversal.

Assume the row player has beliefs μ in Table 1 and thus conjectures his opponent will play ℓ or m, each with probability $\frac{1}{2}$. By choosing a mixed action, the row player can induce any distribution over consequences that mixes between $(0,0)$ for sure and $(1,1)$ or $(0,0)$ with equal probability. If the row player has preferences for universalisation, he will choose ℓ, since it is the action that, if implemented by everyone in this game, maximises his monetary payoff. From a revealed preference perspective, it is inferred that he prefers the lottery $\frac{1}{2}(1,1) + \frac{1}{2}(0,0)$ to the sure consequence $(0,0)$. Now, consider a second scenario where the same individual has beliefs μ' in Table 1, according to which his opponent plays m or r with probability $\frac{1}{2}$. The feasible set of lotteries over consequences is the same as before. Actions m and r induce the midpoint between $(0,0)$ and $(1,1)$ whereas ℓ induces the sure consequence $(0,0)$. The row player still chooses ℓ, as it is again the action that maximises his payoff if implemented by everyone. When $(0,0)$ was available, he revealed to prefer $\frac{1}{2}(1,1) + \frac{1}{2}(0,0)$. Nevertheless, he exhibits a preference reversal in the second scenario, thus violating the weak axiom of revealed preference. There is no complete and transitive preference relation on lotteries consistent with this choice pattern. This impossibility does not occur for consequentialist preferences defined on distributions of material consequences, such as selfishness, altruism, inequity aversion, or maximin. Therefore, functional forms for preferences for universalisation in the literature represent orderings over objects that are different from distributions over material consequences. This implies that preferences over material consequences should not be the relevant measure for welfare analysis of an individual exhibiting universalisation reasoning, contrary to what Roemer (2019) proposes.

This example also shows that the predictions of models of universalisation depend on

the labelling of actions. To avoid the preference reversal, it would suffice to swap the labels of one individual's actions, changing m to r and vice versa. Indeed, Roemer (2019) discusses in multiple instances how to change the label of actions to define and employ universalisation. In Section 5, I present a novel definition of universalisation, relying on the general theory, that is equivalent under any redescription of actions.

2 Model

In this section, I introduce the primitives of the model and the axioms I consider. For any set Y, I denote with $\Delta(Y)$ the set of finite probability distributions over Y.

Primitives. I focus on two-player games, defined as follows.

Definition 1. *A **two-player game** is a list $G = (\{1,2\}, (A_i, \succsim_i)_{i\in\{1,2\}}, X, \rho)$, featuring:*[3]

- *a finite set of actions A_i for each player i;*
- *a common set of consequences X;*
- *a consequence function $\rho\colon A_i \times A_{-i} \to X$;*
- *player i's preferences over mixed actions $\succsim_i$, for each player i.*

Each pair of pure actions (a_i, a_{-i}) induces a consequences $x = \rho_{a_i,a_{-i}}$ where $x \in X$. Any mixed action $\alpha_i \in \Delta(A_i)$ induces an Anscombe & Aumann act denoted with $\rho_{\alpha_i}\colon A_{-i} \to \Delta(X)$ leading to consequence x under opponent's action a_{-i} with probability $\rho_{\alpha_i,a_{-i}}(x) = \alpha_i(\{a_i \in A_i \mid \rho_{a_i,a_{-i}} = x\})$. Each pair of mixed actions (α_i, α_{-i}) induce a distribution over consequence, i.e., the constant act $\rho_{\alpha_i,\alpha_{-i}} \in \Delta(X)$. I say that a mixed action α_i induces a constant act if $\rho_{\alpha_i,a_{-i}}(x) = \rho_{\alpha_i,a'_{-i}}(x)$ for each pair a_{-i}, a'_{-i} and x. I assume there exist mixed actions that, under various opponent's actions, can induce every possible distribution of consequences. A sufficient condition for this to hold is that for each consequence x there exists an action a_i such that $\rho_{a_i,a_{-i}} = x$ for each opponent's action a_{-i}. I need such richness assumption to identify preferences, as usual in decision theory. However, in examples of games in this paper, usually only a subset of A_i of feasible actions is available.[4]

I now introduce a primitive instrumental to capture universalisation reasoning. The idea of universalisation is that, when individual i is evaluating mixed action α_i, he considers the distribution over consequences that would occur if his opponent plays equivalently, under some notion of equivalence. As an example, when the game is symmetric and the action set is the same for both players, he might consider the distribution of consequences

[3]The textbook by Bonanno (2018) discusses games whose primitives are ordinal preferences.

[4]See Dekel & Siniscalchi (2015, p. 631) and references therein for a discussion on elicitation of preferences from bets in games.

induced when his opponent also plays α_i. Fix a reference mixed action profile $(\alpha_i^*, \alpha_{-i}^*)$. A universalisation function $T_{(\alpha_i^*,\alpha_{-i}^*)}\colon \Delta(A_i) \to \Delta(A_{-i})$, maps individual i's mixed action to an opponent's mixed action, given a reference profile. For each mixed action α_i, the corresponding $-i$ universalised action is $T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i]$.

Consider two pure actions inducing the same act. If the individual is consequentialist, he should be indifferent between these two actions, as they induce the same consequences under each opponent's action. Under consequentialism, it would be without loss of generality to study a game in which two actions inducing the same act are identified as the same action. However, preferences for universalisation are not consequentialist, they cannot be reduced to preferences over acts. Therefore, I impose a different notion of equivalence between actions.

I refer to two actions a_i and a_i' as realisation equivalent if

$$\left(\rho_{a_i}, \rho_{a_i, T_{\alpha_i^*,\alpha_{-i}^*}[a_i]}\right) = \left(\rho_{a_i'}, \rho_{a_i', T_{\alpha_i^*,\alpha_{-i}^*}[a_i']}\right).$$

Namely, two actions are realisation equivalent when they induce the same act and the same distribution over consequences when they are universalised. Consider an individual who cares only about the act he induces and the constant act induced under universalisation reasoning. Then, he would be indifferent between two realisation equivalent actions. A game is **reduced** if realisation equivalent actions are the same action.

I study preferences over mixed actions $\succsim_i$ of a generic individual i. I introduce axioms on $\succsim_i$ that characterise the following functional representation.

Definition 2. *A ranking $\succsim_i$ is a **Universalisation Preference** (UP) with respect to the universalisation function $T_{\alpha_i^*,\alpha_{-i}^*}$ if it is represented by*

$$\begin{aligned} U_i(\alpha_i) = (1-\kappa) \sum_{a_i, a_{-i}} \alpha_i(a_i)\mu_i(a_{-i})u_i(\rho_{a_i,a_{-i}}) \\ + \kappa \sum_{a_i, a_{-i}} \alpha_i(a_i) T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i](a_{-i})u_i(\rho_{a_i,a_{-i}}), \end{aligned} \tag{1}$$

for some utility function $u_i\colon X \to \mathbb{R}$ and belief $\mu_i \in \Delta(A_{-i})$.

A *UP* is a linear combination of two components. The first component, weighted by $1-\kappa$, is a standard subjective expected utility. The individual computes the probability the action profile (a_i, a_{-i}) realises, which depends on his mixed action α_i and his belief over opponent's actions μ_{-i}. Then, he evaluates the consequence $\rho_{a_i,a_{-i}}$ obtained according to the utility u_i. The second component, weighted by κ, is the result of universalisation reasoning. Instead of evaluating the probability an opponent's action realises according to the belief μ_i, the individual considers the opponent's mixed action that results from universalising his action via the universalisation function $T_{\alpha_i^*,\alpha_{-i}^*}$. As an example, when the game is symmetric, and then $A_i = A_{-i}$, one can define the identity universalisation function as $T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i] = \alpha_i$ for each α_i, regardless of the reference profile. I show in

Section 4 that the identity universalisation function singles out Homo Moralis preferences from Equation (1).

Axioms. I now introduce the axioms that characterise *UP*. I start with standard axioms allowing me to obtain a utility representation of preferences over mixed actions.

Axiom 1. *(**Weak order**) Preferences $\succsim_i$ are a continuous weak order.*

Axiom 2. *(**Non-triviality**) There exist α_i, α_i' such that $\alpha_i \succ \alpha_i'$.*

I now proceed with axioms that characterise preferences for universalisation. First, the individual only satisfies independence among actions that are universalised equivalently. The intuition is as follows. If the individual was a consequentialist, he would satisfy independence, which would result in the standard independence condition among acts. However, the individual is also interested in the distribution of consequences induced when his action is universalised. A mixture of two actions induce a mixture in their corresponding universalised action. Therefore, when mixing, the distribution over consequences induced by the action and its universalised counterpart is not guaranteed to change linearly. Linearity is guaranteed only if the two actions are universalised equivalently.

Axiom 3. *(**Universalisation Independence**) If*

$$T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i] = T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i'] = T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i''],$$

then, for all $\lambda \in (0,1)$,

$$\alpha \succsim_i \alpha_i' \implies \lambda\alpha_i + (1-\lambda)\alpha_i'' \succsim_i \lambda\alpha_i' + (1-\lambda)\alpha_i''.$$

The next axiom states that, when two actions induce the same act, then their ranking depends on the distribution of consequences they induce when universalised. It restricts attention to preferences over actions that not only depend on the induced act, but also on the distribution of consequences induced by the universalised action.

Axiom 4. *(**Universalisation evaluation**) If $\rho_{\alpha_i} = \rho_{\alpha_i'}$, then*

$$\alpha_i \succsim_i \alpha_i' \text{ if and only if } \rho_{\alpha_i, T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i]} \succsim_i \rho_{\alpha_i', T_{\alpha_i^*,\alpha_{-i}^*}[\alpha_i']}.$$

Lastly, I assume the individual satisfies independence among actions inducing constant acts. The reason is the following. A constant act induces the same distribution of consequences regardless of the opponent's action. Therefore, regardless of how the action is universalised, the opponent is not able to affect the distribution of consequences. There is therefore no reason to violate independence when considering constant acts.

Axiom 5. *(**Lotteries independence**) If α_i, α_i' and α_i'' induce constant acts, then for all $\lambda \in (0,1)$,*

$$\alpha \succsim_i \alpha_i' \implies \lambda\alpha_i + (1-\lambda)\alpha_i'' \succsim_i \lambda\alpha_i' + (1-\lambda)\alpha_i''.$$

As an alternative, one could dispense from Lotteries independence and assume that all actions inducing constant acts are universalised equivalently. Then, Universalisation Independence would imply Lotteries independence. The next section studies the implication of imposing these axioms on preferences over mixed actions.

3 Functional Representation

The main result of this paper shows that the axioms in the previous section are necessary and sufficient to characterise *UP*.[5]

Theorem 1. *A ranking $\succsim_i$ satisfies Weak order, Non-triviality, Universalisation Independence, Universalisation evaluation and Lotteries independence in a reduced game if and only if it is a UP. Moreover, the utility function u_i is unique up to positive affine transformations and beliefs μ_i are unique.*

Theorem 1 states that choices of mixed actions satisfying the axioms are consistent with the following utility function: when choosing the mixed action α_i, the individual evaluates the probability that each opponent's action a_{-i} realises according to his subjective belief μ_i. However, he also considers the distribution of consequences induced by his mixed action α_i and the universalised action according to the universalisation function $T_{\alpha_i^*, \alpha_{-i}^*}$. The two components are aggregated linearly.

I do not derive the form of u_i, the individual may have any preferences over consequences. This fact clarifies the difference between my exercise and, as an example, that of Rohde (2010). Rohde (2010) establishes conditions on a ranking over collective monetary rewards that characterise inequity aversion. In the language of the present paper, she studies the shape of u_i. The axioms here imply nothing about such shape. The representation allows the individual, as an example, to both exhibit preferences for universalisation and, say, inequity aversion, as captured by u_i. Then, in a game, the individual would choose the action that, if universalised, satisfies his inequity averse preference. Theorem 1 thus clarifies that pro-social and non-consequentialist preferences are not exclusive. On the contrary, these two can coexist.

Lastly, Theorem 1 allows marking the difference between universalisation and magical thinking. An individual exhibiting magical thinking believes he affects the opponent's probability to choose an action by choosing it himself. An individual with preferences for universalisation, instead, develops standard subjective beliefs about opponents' actions, and his behaviour does not affect them. Since beliefs are standard, their updating

[5] All proofs are in Appendix B.

should be consistent with Bayes rule in dynamic settings. The first component of the utility function, representing preferences over the induced act, is standard, and therefore results on Bayesian updating holds.[6] In the next section, I study different forms of the universalisation function corresponding to particular preferences in the literature.

4 Preferences for Universalisation

In this section, I study conditions on the universalisation function under which *UP* preferences are equivalent to various notions of universalisation in games. I start with Simple Kantian Equilibrium by Roemer (2019), to later proceed with Homo Moralis by Alger & Weibull (2013) and conclude with Multiplicative Kantian Equilibrium by Roemer (2019). I supplement results with discussions on the interpretation of these concepts and the relation between them.

4.1 Homo Kantiensis and Simple Kantian equilibrium

In this section, I restrict attention to games with common action sets, where $A_1 = A_2 = A$. Simple Kantian Equilibrium is defined as follows.

Definition 3. *An action profile (α, α) constitutes a **Simple Kantian Equilibrium** (SKE) of a game with common action sets if, for all players i and actions α'*

$$\sum_{a_i, a_{-i}} \alpha(a_i)\alpha(a_{-i})u_i(\rho_{a_i,a_{-i}}) \geq \sum_{a_i, a_{-i}} \alpha'(a_i)\alpha'(a_{-i})u_i(\rho_{a_i,a_{-i}}).$$

A symmetric mixed action profile constitutes a *SKE* if it induces the best distribution over consequences over all symmetric mixed action profiles. I show that a *SKE* can be interpreted as a Nash Equilibrium in a game between two players with *Homo Kantiensis* preferences.

Definition 4. *A ranking $\succsim_i$ is a **Homo Kantiensis** (HK) preference if it is represented by*

$$U_i(\alpha) = \sum_{a_i, a_{-i}} \alpha(a_i)\alpha(a_{-i})u_i(\rho_{a_i,a_{-i}}), \tag{2}$$

for some utility function $u_i \colon X \to \mathbb{R}$.

When evaluating any mixed action α, a *HK*, first introduced in Laffont (1975), only considers the distribution over consequences induced when his opponent chooses α as well. A *HK* is a particular case of a *UP* preference. If $\kappa = 1$ and $T_{\alpha_i^*, \alpha_{-i}^*}[\alpha] = \alpha$ for each α, then Equation (1) reduces to Equation (2). In other words, a *HK* satisfies the axioms in Theorem 1 with respect to the identity universalisation function.

[6] See e.g. Epstein & Schneider (2003); Ghirardato (2002).

Proposition 1. *An action profile* (α, α) *constitutes a SKE in a game with common action sets if and only if it constitutes a Nash Equilibrium between two HK.*

Proposition 1 thus establishes that *SKE* is a Nash Equilibrium in a game between two players with preferences over mixed actions satisfying the axioms in Theorem 1 with respect to the identity universalisation function. The result allows me to compare the foundation I offer for *SKE* with that of Roemer (2019). He argues that, contrary to other models in economics, he does not assume exotic preferences, but classical self-regarding attitudes.[7] What he varies, instead, is individuals' "optimisation protocol", as he refers to it. He contrasts Nash optimisation with Kantian optimisation. Nash optimisation, he maintains, relies on the counterfactual "what would happen were I to change my action alone?". Instead, Kantian optimisation induces the counterfactual "what would happen were I and all others to deviate equally?" This argument is echoed in the papers employing various declinations of Kantian Equilibrium.[8]

In the following, I argue that, although appealing, such reasoning cannot be backed up by classical choice theory. I do not take any stance on this point. It is legitimate to employ concepts that diverge from standard theory. Nevertheless, this incompatibility is particularly relevant here, as Roemer relies on his distinction between preferences and optimisation protocol to derive welfare statements.

Roemer's description of the Nash counterfactual refers to the logic employed to check whether an action profile constitutes a Nash Equilibrium. Nevertheless, this is only vaguely related to the foundation of the concept.[9] Outside contexts of long repeated interactions and adaptive dynamics, an action in a Nash Equilibrium profile is played by an individual holding correct conjectures about opponents' behaviour.[10] However, players cannot perform the Nash counterfactual exercise, because they do not know what opponents will do, and are unable to evaluate the gain obtained from a unilateral deviation. An individual in a game selects the action that he considers the best one according to his beliefs about what his opponents will do. In turn, the definition of "best" is, in economics, his preference. In choice theory, observed behaviour is interpreted as revealing a preference for an object compared with others available, actions in this case. Optimisation is a mathematical technique employed to compute what the maximal element is given a primitive ranking over the objects of choice, it is not a feature of the individual or of an equilibrium concept. There is no empirical observation able to tell that two individuals have the same preferences but different optimisation protocols. If they choose differently in the same problem, by definition they have different preferences.

I show with Proposition 1 that there is no need to rely on informal arguments regarding how individuals optimise. Behaviour consistent with *SKE* can be interpreted as Nash Equilibrium behaviour in a game between two *HK*. Therefore, Roemer is correct in arguing

[7] See, among many others, Roemer (2019, p. 69).

[8] See the papers in the literature review in Section 1.

[9] Battigalli et al. (2023) offer a thorough discussion on the interpretation of Nash Equilibrium.

[10] See Perea (2012) or Dekel & Siniscalchi (2015) and references therein.

that assuming individuals behave according to *SKE* is different from saying that they are pro-social. Nevertheless, this does not mean that they optimise differently.

The critique above has implications for welfare analysis. Roemer's argument according to which, in *SKE*, individuals have selfish preferences over material consequences but the optimisation protocol is different from Nash, generates confusion. As I showed in the motivating example, it is possible that an individual who plays according to *SKE* does not have a complete and transitive preference, and hence a utility representation, over material consequences. I believe the closest reformulation of Roemer's point is that one can have preferences for universalisation even if the utility index in Theorem 1 for material consequences u_i is the same as that for consequences induced by universalised actions, as in *UP* preferences in Equation (1). Nevertheless, this equivalence does not imply the individual would be indifferent between receiving a monetary amount and acting to induce it as a consequence of universalisation reasoning. Great care must be devoted to make welfare statements for non-consequentialist preferences over actions. Given that universalisation is a preference over actions, one interesting avenue is to consider that welfare should be evaluated in terms of the freedom the individual has in choosing an action he prefers.[11]

Proposition 1 also offers a novel rationale for using mixed actions. Under expected utility, there is always a pure action in the set of best replies to probabilistic conjectures regarding opponents' behaviour. The Nash equilibrium mixed action of player i can be interpreted as strategic uncertainty from player $-i$'s perspective. Nevertheless, a *HK* who plays a mixed action in a *SKE* profile has no interest in being difficult to be predicted by his opponents. In his best reply set, there may be no pure actions. A rationale for employing mixed actions is therefore the adherence to a non-consequentialist attitude.

4.2 Homo Moralis

In this section, I exploit the representation in Theorem 1 to derive Homo Moralis preferences as a special case of Equation (1). I again restrict attention to games with common action sets, to define Homo Moralis preferences as follows.

Definition 5. *A ranking $\succsim_i$ is a **Homo Moralis** (HM) preference if it is represented by*

$$\begin{aligned} U_i(\alpha) = (1-\kappa) \sum_{a_i, a_{-i}} \alpha(a_i)\mu_i(a_{-i})u_i(\rho_{a_i,a_{-i}}) \\ + \kappa \sum_{a_i, a_{-i}} \alpha(a_i)\alpha(a_{-i})u_i(\rho_{a_i,a_{-i}}), \end{aligned} \tag{3}$$

for some utility function $u_i \colon X \to \mathbb{R}$ and belief $\mu_i \in \Delta(A_{-i})$.

A Homo Moralis maximises a convex combination between subjective expected utility and expected payoff when both individuals play his action. Contrary to *SKE*, *HM* is

[11] Laslier et al. (1998) offer a review of approaches on how to conceptualise freedom in economics.

a preference, not a property of action profiles. A *HM* with $\kappa = 1$ is a *HK* and his preferences are represented by Equation (2). A *HM* is a *UP* where the universalisation function is the identity, as in *HK*. If $T_{\alpha_i^*,\alpha_{-i}^*}[\alpha] = \alpha$ for each α, then Equation (1) coincides with Equation (3).

A *HM* is not only interested in the consequence obtained when his action is universalised, but trades off consequentialist and non-consequentialist motives. Since *HM* is partially strategic, he also cares about his opponent's action and thus his beliefs matter. However, contrary to magical thinking, a *HM* exhibits standard beliefs that do not depend on his action. Such distinction rationalises current experimental practices relying on identifying beliefs of individuals with preferences for universalisation (Van Leeuwen & Alger, 2024). Moreover, my analysis offers a new behavioural prediction to test for the presence of *HM*, and therefore also of *HK*: they both violate independence between mixed actions that are not universalised equivalently, according to the identity universalisation function.

Both *SKE* and *HM* are well-defined only in games with common action sets. Alger & Weibull (2013) suggest a way to employ *HM* preferences in asymmetric games. They propose to consider an incomplete information expansion of the basic game where players are not aware of their role, reminiscent of the veil of ignorance of Harsanyi (1955) and Rawls (1971). Such incomplete information game is a symmetric interaction where a strategy is a map between role and action. Universalisation can then be defined as strategies are common across players. The authors refer to this preference as *Ex-ante Homo Moralis*. Another definition of universalisation in asymmetric games is Multiplicative Kantian Equilibrium by Roemer (2019). In the next section, I discuss Multiplicative Kantian Equilibrium, postponing observations on *Ex-ante HM* to Section 5.

4.3 Multiplicative Kantian equilibrium

In this section, I discuss the relationship between Multiplicative Kantian Equilibrium and *UP*.[12] The solution concept is defined in games where the action space has a linear structure. It is employed when players can choose a number from the real line. For simplicity, I follow the recommendation in Roemer (2019, p. 42) and consider the mixed extension of two-player two-actions games, though developing a generalisation of Multiplicative Kantian Equilibrium to multiple actions games is not trivial.

I remove the restriction to games with common action sets and assume players only have two pure actions available. For each α_i and $r \geq 0$, I denote with $r \cdot \alpha_i$ an operation that affects α_i by the multiplicative factor r and $1 - \alpha_i$ by the complementary weight to obtain a probability distribution on pure actions, that is

$$r \cdot \alpha_i := \frac{r\alpha_i}{r\alpha_i + (1 - \alpha_i)} .$$

[12] An equivalent analysis delivers similar results for Additive Kantian Equilibrium (Roemer, 2019).

Definition 6. *An action profile* (α_i, α_{-i}) *constitutes a* ***Multiplicative Kantian Equilibrium*** *(MKE) of a game if, for all players* i *and real numbers* $r \geq 0$

$$\sum_{a_i, a_{-i}} \alpha_i(a_i)\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}) \geq \sum_{a_i, a_{-i}} r \cdot \alpha_i(a_i) r \cdot \alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}).$$

A mixed action profile constitutes a *MKE* if it induces the best distribution over consequences when compared with all mixed action profiles that can be obtained by multiplying both actions by r.[13] The notion is well defined as I am restricting attention to two actions. A multiplicative deviation is equivalent to moving weight from one action to the other.

Paralleling the analysis for *SKE*, I show that *MKE* can be interpreted as a Nash Equilibrium in a game between two individuals with Multiplicative Homo Kantiensis preferences, defined as follows.

Definition 7. *A ranking* $\succsim_i$ *is a* ***Multiplicative Homo Kantiensis (MHK)*** *preference relative to the profile* (α_i, α_{-i}) *if it is represented by*

$$U_i(r \cdot \alpha_i) = \sum_{a_i, a_{-i}} r \cdot \alpha_i(a_i) r \cdot \alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}), \tag{4}$$

for some utility function $u_i \colon X \to \mathbb{R}$.

The difference between *MHK* and *HK* is how actions are universalised, as illustrated in Figure 1. A *HK* only conceives both players choosing the same action. In two-player symmetric games with two actions, these profiles correspond to the diagonal of the square representing mixed actions. Instead, *MHK* considers actions as multiplicative deviations from a specific profile. Universalised actions lie on the line connecting the origin and the reference profile, i.e., all the pairs in which the ratio between the two actions is preserved. A profile (α_i, α_{-i}) constitutes a *MKE* if it is the preferred one for both players compared with any other on the line joining the origin and (α_i, α_{-i}). If (α_i, α_{-i}) lays on the 45° line, the two universalisation criteria are identical. In the next section, I discuss a novel version under which the universalisation criterion is endogenous and depends on the game at hand.

[13] According to this definition, $(0, 0)$ is always a *MKE* if it is available.

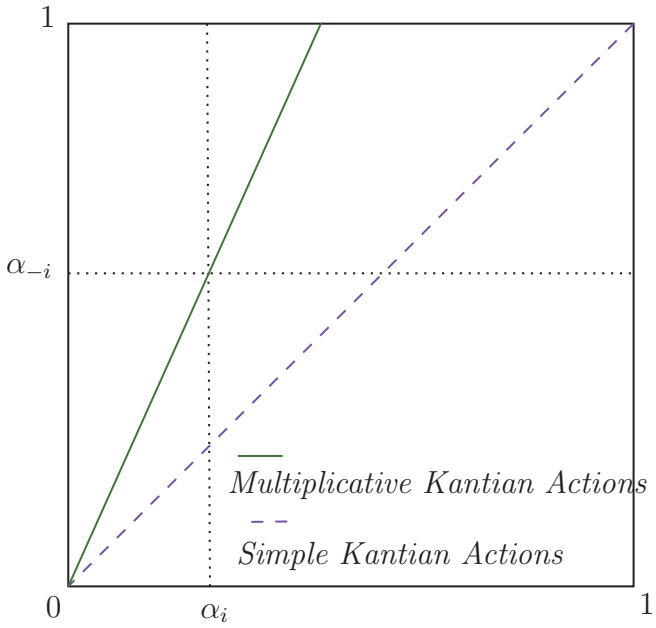

Figure 1: Universalised action profiles of Simple and Multiplicative Kantian Equilibria.

A *MHK* is a particular case of *UP* in which the actions are universalised through a multiplicative logic. If $\kappa = 1$ and $T_{\alpha_i,\alpha_{-i}}[r \cdot \alpha_i] = r \cdot \alpha_{-i}$ for each r, then Equation (1) reduces to Equation (4).

Proposition 2. *An action profile* (α_i, α_{-i}) *constitutes a MKE in a game if and only if it constitutes a Nash Equilibrium between two MHK relative to the profile* (α_i, α_{-i}).

Proposition 2 has an interpretation on the lines of its parallel result for *SKE*. It establishes that *MKE* is a Nash Equilibrium in a game between two players with preferences over mixed actions satisfying the axioms in Theorem 1 with respect to the multiplicative universalisation function. Contrary to *SKE*, *MKE* can be defined when action sets are not common and allows individuals to choose heterogeneous actions, as *Ex-ante HM* does. In the next section, I develop a new concept that takes a different route to define universalisation in games with general action sets.

5 Equal Sacrifice Universalisation

In this section, I elaborate on the concept of universalisation and present a new notion, inspired by the equal sacrifice principle (Mill, 1885; Young, 1988). The model I propose has several features: its definition does not depend on the label of actions; it can be defined in asymmetric games; in symmetric games it is equivalent to *HM*.

Universalisation requires the definition of two objects. First, it must be transparent what "doing the same thing" is. Second, it must be equally clear what "deviating in the same manner" means. For these two concepts to be defined, a common currency must exist for the adjective "same" to have meaning. Previous ideas employed the label

of actions in games and a notion of distance between them when the action space is structured. Such an approach, I argue, is partially lacking. In most economic models, the label of actions bears no conceptual relevance and might be misleading to use it as the main ingredient of a model of universalisation. In fact, in many applications where *MKE* gives intuitive results, the action labels have clear conceptual significance, as they represent effort, contribution to a public good, or use of a common resource.

I propose to use the relevant consequence of the game as a currency. In game theory, this is usually players' utility, but it can be any other index of well-being. Then "doing the same thing" and "deviating in the same manner" are interpreted as "inducing the same utility" and "inducing the same difference in utility". The following example illustrates the idea. Consider two individuals playing the prisoners' dilemma. Numbers are Bernoulli utilities for consequences.

$1 \setminus 2$	a_2	a_2'
a_1	$2, 2$	$0, 3$
a_1'	$3, 0$	$1, 1$

Table 2: Prisoners' dilemma.

Row player 1 attains his highest Bernoulli utility in $(3, 0)$, induced by the profile (a_1', a_2). I define α_1^s as a mixed action which, compared to a_1', given a_2, induces a reduction in expected Bernoulli utility by s, i.e., $3 - [2\alpha_1^s + 3(1 - \alpha_1^s)] = s$. The profile leading to the highest Bernoulli utility for column player 2 is instead (a_1, a_2'). As for player 2, his mixed actions α_2^s induces the difference $3 - [2\alpha_2^s + 3(1 - \alpha_2^s)] = s$. Assume that player 1, when picking any action α_1^s, considers a scenario where 2 chooses α_2^s. He envisions the consequence that would obtain if his opponent chooses the action that, if considered a unilateral deviation from (a_1, a_2'), generates the same difference in Bernoulli utility. In this game, $\alpha_1^s = \alpha_2^s = s$ for all s. Whenever they deviate from the action profile that yields their preferred material outcome, both players consider a scenario in which their opponents also deviate by choosing the same action. Therefore, the universalisation is identical to the one of *HK* and *HM* in this example. The actions that lead to the highest Bernoulli utility under this universalisation reasoning are a_1 for 1 and a_2 for 2, i.e., $\alpha_1^s = \alpha_2^s = 1$, the same optimal actions of *HK* under proper re-labelling of actions.

Evaluating differences from the maximum attainable payoff is reminiscent of the equal sacrifice principle of Mill (1885) in the context of taxation. Hence, I dub this concept *equal sacrifice universalisation* (*ESU*). An individual with *ESU* preferences first identifies the profile of actions inducing his preferred consequence. Second, he evaluates each action considering the induced difference in payoff compared with the optimal action computed previously. Third, he individuates the collection of opponents' deviations that, compared with their maximal action profiles in the material dimension, lead to obtain the same absolute difference.

To ease the exposition, I here focus on equal absolute sacrifice (Young, 1988). In Appendix A, I consider general equal sacrifice rules. The results and arguments in this section hold for any equal sacrifice rule. Under Weak order, Non-triviality and Lotteries independence, preferences over actions inducing constant acts satisfy the standard Anscombe & Aumann conditions. These axioms guarantee the existence of two distributions over consequence $\overline{\gamma}, \underline{\gamma} \in \Delta(X)$ such that $\overline{\gamma} \succsim_i \gamma \succsim_i \underline{\gamma}$ for all γ. Then, one could define a distribution over consequences inducing sacrifice s as[14]

$$\gamma^s := \lambda_s \overline{\gamma} + (1 - \lambda_s)\underline{\gamma}.$$

Then, define $\alpha_i^*, \alpha_{-i}^*$ as any two actions such that $\rho_{\alpha_i^*, \alpha_{-i}^*} = \overline{\gamma}$, and α_i^s as any action such that $\rho_{\alpha_i^s, \alpha_{-i}^*} = \gamma^s$. Actions α_i^s leading to absolute sacrifice s by definition satisfy

$$\sum_{a_{-i}} \alpha_{-i}^*(a_{-i}) \sum_x \rho_{\alpha_i^*, a_{-i}}(x) u_i(x) - \sum_{a_{-i}} \alpha_{-i}^*(a_{-i}) \sum_x \rho_{\alpha_i^s, a_{-i}}(x) u_i(x) = s. \tag{5}$$

Neither $(\alpha_i^*, \alpha_{-i}^*)$ nor any α_i^s are guaranteed to be unique. For the sake of the following definition, I assume they are.[15]

Definition 8. *A ranking $\succsim_i$ is an **Equal Sacrifice Universalisation** preference if it is represented by*

$$\begin{aligned} U_i(\alpha_i^s) = (1 - \kappa) &\sum_{a_i, a_{-i}} \alpha_i^s(a_i) \mu_i(a_{-i}) u_i(\rho_{a_i, a_{-i}}) \\ + \kappa &\sum_{a_i, a_{-i}} \alpha_i^s(a_i) \alpha_{-i}^s(a_{-i}) u_i(\rho_{a_i, a_{-i}}), \end{aligned} \tag{6}$$

for some utility function $u_i \colon X \to \mathbb{R}$ and belief $\mu_i \in \Delta(A_{-i})$.

When choosing α_i^s, the individual evaluates the scenario where his opponent deviates from the action in the profile leading to the highest Bernoulli utility to induce the same sacrifice. *ESU* is a *UP* in which $T_{\alpha_i^*, \alpha_{-i}^*}[\alpha_i^s] = \alpha_{-i}^s$ for all s, so that Equations (1) and (6) coincide.

[14] The weight λ_s depends on the normalisation of utility. As an example, if

$$\sum_x \overline{\gamma}(x) u_i(x) = S \text{ and } \sum_x \underline{\gamma}(x) u_i(x) = 0,$$

then $\lambda_s = 1 - \frac{s}{S}$. In fact, by linearity $\sum_x \gamma^s(x) u_i(x) = \lambda_s S$, and I would like that $\sum_x \gamma^s(x) u_i(x) = S - s$, so that

$$\lambda_s S = S - s \;\Rightarrow\; \lambda_s = 1 - \frac{s}{S}.$$

[15] Therefore, I am implicitly considering a restriction of preferences or games.

The key difference between *ESU* and previous concepts is that universalisation reasoning depends on the game at hand. I illustrate this point in the battle of the sexes. Numbers are again Bernoulli utilities for consequences.

1 \ 2	a_2	a'_2
a_1	$2,1$	$0,0$
a'_1	$0,0$	$1,2$

Table 3: Asymmetry.

1 \ 2	a	a'
a	$2,1$	$0,0$
a'	$0,0$	$1,2$

Table 4: Same Actions.

1 \ 2	a	a'
a	$0,0$	$2,1$
a'	$1,2$	$0,0$

Table 5: Symmetry.

Consider the game in Table 3 on the left and assume throughout that $\kappa = 0$ for simplicity. The table represents a standard battle of the sexes, in which player 1 would like to coordinate in the top-left corner, while player 2 would like to coordinate on the bottom-right corner. The greatest achievable payoff of both players is 2, in (a_1, a_2) and (a'_1, a'_2). The action α_1^s of player 1 inducing a sacrifice of s solves $2 - 2\alpha_1^s = s$ and hence $\alpha_1^s = \frac{2-s}{2}$. The equivalent for player 2 is $2 - (2 - 2\alpha_2^s) = s$ which implies $\alpha_2^s = \frac{s}{2} = 1 - \alpha_1^s$. The optimum for *ESU* is reached at $s = \frac{2}{3}$ with $\alpha_1^s = \alpha_2^s = \frac{1}{2}$ which, if picked by both players, leads to a common expected payoff of $\frac{3}{2}$.

This simple example allows me to discuss important differences between *ESU* and previous concepts. First, even if we were to relabel the actions (a_2, a'_2) to (a, a') for employing *SKE*, as in the table in the middle, one would not exist anyway. The optimal action is not common, as it is a for 1 and a' for 2. Nevertheless, I argue that the problem here is not existence. It is possible to define universalisation from an individual perspective and obtain the profile composed by subjectively optimal actions (a, a'). This is indeed what would happen assuming both players are *HK*. The issue is that it is meaningless to define "the same thing" as "the same action" in this scenario. The relabelling of actions from Table 3 to 4 is arbitrary as any other, it is not surprising that it does not lead to intuitive results.

As a solution, Roemer (2019, p. 26) suggests to relabel the game as in Table 5 on the right, to make it symmetric. Now actions are interpreted as "do the favourite thing" and "do the least favourite thing". The *SKE* of this reformulation of the game is $(\frac{1}{2}, \frac{1}{2})$, i.e., the optimal actions of *ESU*. Not only the optimal profile coincides, but also the set of profiles considered in the universalisation evaluation is identical. The re-labelling of actions from the first to the third table amounts to changing any mixed action α_2^s to $1 - \alpha_1^s$, which leads to $a_2 = a'_1$ and $a'_2 = a_1$ and switching columns. This is exactly the *ESU* universalisation reasoning.

Now consider the difference between *ESU* and *Ex-Ante HM*. *Ex-Ante HM* is defined in an incomplete information expansion of the game in which players do not know whether they will be the row player or the column player. When $\kappa = 1$, it prescribes players to choose the strategy, in this case mapping between identity and action, that ex-ante, before identities are revealed, maximises expected utility over material consequences. The

optimal strategies according to such criterion are (a_1, a_2) and (a_1', a_2'). Contrary to what is implemented if both players exhibit *ESU*, these two profiles are Pareto-Efficient. It is already known that *Ex-Ante HM* is related to utilitarian altruism (Laslier, 2022). Hence, it is possible that *ESU* delivers an inefficient allocation in terms of material payoff. By contrast, *Ex-ante HM* is always efficient, but is indifferent to inequality.

The following result establishes that optimal actions under *ESU* are always optimal actions under *HK* in symmetric games and therefore the first is a generalisation of the second. The result holds for any equal sacrifice rule, not only absolute sacrifice, as shown in the proof in Appendix B.

Proposition 3. *Assume the game is symmetric. Then, if an action is optimal under ESU with $\kappa = 1$, it is also optimal under HK.*

The result may be interpreted as a conceptual robustness check. In games where "same action" has meaning, because of symmetry, *ESU* delivers the intuitive universalisation evaluation of previous concepts. In asymmetric games, the universalisation evaluation depends on the equal sacrifice conception of the individual.

I conclude by addressing possible critiques to *ESU*. First, it relies on interpersonal comparisons of utility, and thus is less parsimonious compared with previous concepts. I acknowledge the issue, but I argue that universalisation always relies on some form of interpersonal comparison and hence the problem is not idiosyncratic to *ESU*. *Ex-ante HM* also relies on the same informational requirement, as it employs the veil of ignorance construct, and thus relies on the same interpersonal comparisons of Harsanyi's utilitarianism. As for the various forms of Kantian Equilibrium, these rely on interpersonal comparisons of actions, as argued by Sher (2020), as actions need to have a cardinal interpretation common to all players. Some form of interpersonal comparisons is therefore needed also in previous conceptions.

The issue is deeper. It is not that universalisation needs some form of interpersonal comparison outside symmetric environments. It always does, but under symmetry, both concepts of "same action" and "same utility" have meaning, so comparisons of actions and utility are easy to deal with. Universalisation becomes problematic without symmetry not because of labels, but because of heterogeneity among players. The implicit suggestion of *Ex-ante HM* is to solve such heterogeneity by aggregating preferences in the utilitarian fashion. MKE, instead, suggests to give actions a cardinal meaning. *ESU* offers a third way.[16]

A second issue is that *ESU* might lead to corner solutions. The problem is related to the previous one. It is possible that utility indexes across players have different scales and range and this makes it hard for equal sacrifice of utility to be feasible. A partial solution is to perform a proper rescaling of utility.[17] When this is not enough, constrained

[16] Incidentally, *ESU* is reminiscent of Kalai & Smorodinsky (1975) bargaining solution.

[17] Interpersonal comparisons of utility are widely discussed in social choice theory. Binmore (1994, Ch. 4) and Sen (2017, Ch. 7) offer critical overviews of approaches to perform this exercise.

versions of equal sacrifice, developed by Stovall (2020), can be employed.

6 Conclusion

I developed and axiomatically characterised a model to account for non-consequentialist preferences for universalisation. The main behavioural prediction is that independence is satisfied only among actions that are universalised equivalently. I showed that the general model unifies the two most prominent models of universalisation, namely Homo Moralis and Kantian Equilibrium. Lastly, inspired by the equal sacrifice principle, I proposed a novel concept of universalisation that does not rely on the labelling of actions, is equivalent to the previous models under symmetry, and can be defined in asymmetric games. I showed how the results shed light on the conceptual underpinnings of universalisation, guide empirical work, and inform the evaluation of welfare statements. In the last paragraphs, I discuss two points regarding the methodology and implications of this paper.

I am not the first to propose changing the set of consequences to account for apparent paradoxes. Baccelli & Mongin (2022), among others, have criticised this practice, as a redescription of the problem might solve technical but not conceptual issues. They argue that it is more reasonable to capture non-material determinants of utility in the evaluation of consequences, without affecting their definition. In this paper, I adhere to this principle. I do not need to alter the set of consequences by including other features in the game. The key is to introduce a link between actions and consequences without changing these two primitives. As my introductory example shows, universalisation cannot be rationalised without assuming that the individual cares about something unrelated to the material consequences of the game. An expansion of the consequence domain is necessary. A second possibility is to include the chosen action in the description of the consequence. It would then be easy to formalise a trade-off between selecting the preferred action and maximising material payoff. This has been done in empirical work on moral preferences, notably by Cappelen et al. (2007). By contrast, my universalisation theory does not rely on assuming that an action is optimal but explains why, i.e., because it induces the preferred universalised consequence.

The final point concerns the nature of preferences for universalisation. I have denoted these as non-consequentialist, and the literature refers to them as moral. Nevertheless, I show that universalisation satisfies consequentialism under an appropriate redefinition of consequences which consider what is induced by the universalised action. What, then, is the difference between universalisation and consequentialist pro-social attitudes? John Broome argues, in Bradley & Fleurbaey (2021, p. 120), that "*a very specific version of consequentialism is a view I call distribution (it is often called welfarism), which is the view that the goodness of an act is determined by the goodness of the distribution of well-being that results from it*". Universalisation is, strictly speaking, not a welfarist attitude,

as the optimal action is unrelated to the distribution of well-being it induces.

References

Alger, I., & Laslier, J.-F. (2022). Homo moralis goes to the voting booth: Coordination and information aggregation. *Journal of Theoretical Politics*, *34*(2), 280–312. 5

Alger, I., & Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica : journal of the Econometric Society*, *81*(6), 2269–2302. 2, 4, 5, 11, 14

Alger, I., & Weibull, J. W. (2016). Evolution and kantian morality. *Games and Economic Behavior*, *98*, 56–67. 5

Alger, I., Weibull, J. W., & Lehmann, L. (2020). Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory*, *185*, 104951. 5

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of mathematical statistics*, *34*(1), 199–205. 2, 4, 7, 18

Baccelli, J., & Mongin, P. (2022). Can redescriptions of outcomes salvage the axioms of decision theory? *Philosophical Studies*, *179*(5), 1621–1648. 21

Battigalli, P., Catonini, E., & De Vito, N. (2023). *Game theory: Analysis of strategic thinking* [Lecture Notes]. 12

Battigalli, P., Cerreia-Vioglio, S., Maccheroni, F., & Marinacci, M. (2017). Mixed extensions of decision problems under uncertainty. *Economic Theory*, *63*(4), 827–866. 4, 27

Binmore, K. (1994). *Game Theory and the Social Contract: Playing fair*. Cambridge, MA: MIT Press. 20

Bonanno, G. (2018). *Game theory* (2nd ed.). North Charleston, SC: CreateSpace Independent Publishing Platform. 7

Bradley, R., & Fleurbaey, M. (2021). John Broome. *Conversations on Social Choice and Welfare Theory-Vol. 1*, 115–127. 21

Brekke, K. A., Kverndokk, S., & Nyborg, K. (2003). An economic model of moral motivation. *Journal of public economics*, *87*(9-10), 1967–1983. 5

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., & Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, *97*(3), 818–827. 21

Chateauneuf, A., & Wakker, P. (1993). From local to global additive representation. *Journal of Mathematical Economics*, *22*(6), 523–545. 28

Chen, D. L., & Schonger, M. (2022). Social preferences or sacred values? Theory and evidence of deontological motivations. *Science Advances*, *8*(19). 4

Daley, B., & Sadowski, P. (2017). Magical thinking: A representation result. *Theoretical Economics*, *12*(2), 909–956. 5

De Donder, P., Llavador, H., Penczynski, S. P., Roemer, J. E., & Vélez-Grajales, R. (2025). Nash versus Kant: A game-theoretic analysis of childhood vaccination behavior. *Journal of Economics*. 5

Dekel, E., & Siniscalchi, M. (2015). Epistemic game theory. In *Handbook of Game Theory with Economic Applications* (Vol. 4, pp. 619–702). Elsevier. 7, 12

Dewatripont, M., & Tirole, J. (2024, August). The Morality of Markets. *Journal of Political Economy*, *132*(8), 2655–2694. 5

Dierks, K., Alger, I., & Laslier, J.-F. (2024). Does universalization ethics justify participation in large elections? *TSE Working Paper*. 5

Dizarlar, A., & Karagözoğlu, E. (2023). Kantian equilibria of a class of Nash bargaining games. *Journal of Public Economic Theory*, *25*(4), 867–891. 5

Ellingsen, T., & Mohlin, E. (2024). A model of social duties. *Working Paper*. 5

Epstein, L. G., & Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, *113*(1), 1–31. 11

Fleurbaey, M. (2008). *Fairness, responsibility, and welfare*. Oxford: Oxford University Press. 4

Fleurbaey, M. (2019). Economic theories of justice. *Annual Review of Economics*, *11*, 665–684. 2

Fleurbaey, M., Kanbur, R., & Snower, D. J. (2023, December). *An analysis of moral motives in economic and social decisions* (CEPR Discussion Paper No. DP18682). London: Centre for Economic Policy Research (CEPR). 5

Ghirardato, P. (2002). Revisiting Savage in a conditional world. *Economic theory*, *20*, 83–92. 11

Gilboa, I., & Schmeidler, D. (2003). A derivation of expected utility maximization in the context of a game. *Games and Economic Behavior*, *44*(1), 172–182. 4

Grillo, A. (2022). Ethical Voting in Heterogenous Groups. *Working Paper*. 5

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, *61*(5), 434–435. 20

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, *63*(4), 309–321. 14

Juan-Bartroli, P. (2024). On Injunctive Norms: Theory and Experiment. *TSE Working Paper, n. 24-1515*. 5

Juan-Bartroli, P., & Karagözoğlu, E. (2024). Moral preferences in bargaining. *Economic Theory*, 1–24. 5

Kalai, E., & Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica : journal of the Econometric Society*, 513–518. 20

Laffont, J.-J. (1975). Macroeconomic constraints, economic efficiency and ethics: An introduction to Kantian economics. *Economica*, *42*(168), 430–437. 4, 11

Laslier, J.-F. (2022). Universalization and altruism. *Social Choice and Welfare*, 1–16. 5, 20

Laslier, J.-F., Fleurbaey, M., Gravel, N., & Trannoy, A. (1998). Freedom in Economics. *New Perspectives in Normative Analysis*. 13

Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences*, *117*(42), 26158–26169. 2

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York: John Wiley & Sons. 3, 4

Miettinen, T., Kosfeld, M., Fehr, E., & Weibull, J. (2020). Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *Journal of Economic Behavior & Organization*, *173*, 1–25. 2

Mill, J. S. (1885). *Principles of political economy*. New York: D. Appleton and Company. 4, 16, 17

Perea, A. (2012). *Epistemic game theory: Reasoning and choice*. Cambridge: Cambridge University Press. 12

Ponthiere, G. (2024a). Epictetusian rationality. *Economic Theory*, *78*(1), 219–262. 5

Ponthiere, G. (2024b, March). *Stoicism and the tragedy of the commons* (GLO Discussion Paper No. 1405). Essen: Global Labor Organization (GLO). 5

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: The Belknap Press of Harvard University Press. 14

Roemer, J. E. (2010). Kantian equilibrium. *Scandinavian Journal of Economics*, *112*(1), 1–24. 2, 5

Roemer, J. E. (2015). Kantian optimization: A microfoundation for cooperation. *Journal of Public Economics*, *127*, 45–57. 5

Roemer, J. E. (2019). *How we cooperate: A theory of kantian optimization*. New Haven, CT: Yale University Press. 2, 3, 4, 5, 6, 7, 11, 12, 13, 14, 19

Rohde, K. I. (2010). A preference foundation for Fehr and Schmidt's model of inequity aversion. *Social Choice and Welfare*, *34*(4), 537–547. 10

Sarkisian, R. (2017). Team Incentives under Moral and Altruistic Preferences: Which Team to Choose? *Games*, *8*(3), 37. 5

Sarkisian, R. (2021a). Optimal Incentives Schemes under Homo Moralis Preferences. *Games*, *12*(1), 28. 5

Sarkisian, R. (2021b). Screening Teams of Moral and Altruistic Agents. *Games*, *12*(4), 77. 5

Savage, L. J. (1972). *The foundations of statistics* (2nd rev. ed.). New York: Dover Publications. 2

Sen, A. (1973). Behaviour and the Concept of Preference. *Economica*, *40*(159), 241–259. 2

Sen, A. (2017). *Collective choice and social welfare: An expanded edition* (Expanded ed.). Cambridge, MA: Harvard University Press. 20

Sher, I. (2020). Normative aspects of kantian equilibrium. *Erasmus Journal for Philosophy and Economics*, *13*(2), 43–84. 20

Shi, Y. (2024). Endogenous Social Minimum. *Working Paper*. 5

Sobrado, E. M. (2022). Taxing moral agents. *CESifo working paper series 9867*. 5

Stovall, J. E. (2020). Equal sacrifice taxation. *Games and Economic Behavior*, *121*, 55–75. 21

Thomson, W. (2013). Game-theoretic analysis of bankruptcy and taxation problems: Recent advances. *International Game Theory Review*, *15*(03), 1340018. 26

Thomson, W. (2019). *How to divide when there isn't enough: From aristotle, the talmud, and maimonides to the axiomatics of resource allocation* (No. 62). Cambridge: Cambridge University Press. 26

Van Leeuwen, B., & Alger, I. (2024, November). Estimating Social Preferences and Kantian Morality in Strategic Interactions. *Journal of Political Economy Microeconomics*, *2*(4), 665–706. 2, 3, 14

Young, H. P. (1988). Distributive justice in taxation. *Journal of Economic Theory*, *44*(2), 321–335. 4, 16, 18

A Equal Sacrifice in Games

I map normal-form games to claim problems.[18] This exercise allows defining equal sacrifice universalisation for any sacrifice rule. I restrict attention to two-player games. Here $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the non-negative and positive real numbers, respectively.

A **claim problem** is an ordered list $(\{1,2\},(x_i)_{i\in I})$ where $\{1,2\}$ is the set of players and $x_i \in \mathbb{R}_{++}$ is the claim of individual i. An **award** is $y_i \in \mathbb{R}_+$ satisfying $0 \leq y_i \leq x_i$ for all i. In my formulation, the claim of each player in a game is the maximal expected utility for consequences he can obtain, which I denote with $\overline{V_i}$.[19] Therefore, $x_i = \overline{V_i}$ for all i. An **allocation rule** maps claims to awards $\pi : \mathbb{R}^2_{++} \to \mathbb{R}^2_+$. An **equal sacrifice function** is a continuous, strictly increasing and hence invertible function $R : \mathbb{R}_{++} \to \mathbb{R}$. The equal sacrifice allocation rule relative to the function R and sacrifice $s \in \mathbb{R}_+$ is

$$\pi_R\left(x_i, x_{-i}\right) := \left(R^{-1}\left(R\left(\overline{V_i}\right) - s\right)\right)_{i\in I}.$$

As an example, the equal loss rule $\pi_R\left(x_i, x_{-i}\right) = \left(x_i - s\right)_{i\in I}$ in the main text has $R\left(x_i\right) = x_i$ for all x_i. In a game, utilities depend on actions, so denote

$$U_i^c\left(\alpha_i, \alpha_{-i}\right) = \sum_{a_i, a_{-i}} \alpha_i(a_i)\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}})$$

for each (α_i, α_{-i}). An action profile inducing the maximal expected utility for i is $\left(\alpha_i^*, \alpha_{-i}^*\right)$ so that $U_i^c\left(\alpha_i^*, \alpha_{-i}^*\right) = \overline{V_i}$. Then, action α_i^{Rs} induces sacrifice s relative to the function R if

$$R^{-1}\left(R\left(U_i^c\left(\alpha_i^*, \alpha_{-i}^*\right)\right) - s\right) = U_i^c\left(\alpha_i^{Rs}, \alpha_{-i}^*\right).$$

A player exhibits equal sacrifice universalisation with respect to R if his universalisation function is $T_{\alpha_i^*, \alpha_{-i}^*}\left[\alpha_i^{Rs}\right] = \alpha_{-i}^{Rs}$.

[18] I redirect the interested reader to Thomson (2019) for a general treatment of claim problems. Notice that the model in this section is not related to game-theoretic analyses of claim problems surveyed by Thomson (2013). The only purpose is to determine players' universalisation functions, not to distribute a given endowment.

[19] In this appendix, I employ the utility representation for simplicity, but everything could be defined only using ordinal preferences $\succsim_i$.

B Proofs

Proof of Theorem 1. I omit necessity and focus on sufficiency. First, for each α_{-i}, consider the set of actions

$$\Delta^{\alpha_{-i}}(A_i) = \left\{ \alpha_i \in \Delta(A_i) \mid T_{\alpha_i^*, \alpha_{-i}^*}(\alpha_i) = \alpha_{-i} \right\}.$$

This is the set of all actions that are universalised to α_{-i}. Since the game is reduced, all actions in $\Delta^{\alpha_{-i}}(A_i)$ that induce the same act are the same action. Moreover, preferences $\succsim_i$ satisfies independence when restricted to $\Delta^{\alpha_{-i}}(A_i)$ by Universalisation Independence. By Weak order, Non-triviality and Theorem 4 in Battigalli et al. (2017), preferences $\succsim_i$ are represented by

$$U_i^{\alpha_{-i}}(\alpha_i) = \sum_{a_i, a_{-i}} \alpha_i(a_i)\mu_i(a_{-i})u_i'(\rho_{a_i,a_{-i}}, \alpha_{-i}), \tag{7}$$

for each $\alpha_i \in \Delta^{\alpha_{-i}}(A_i)$, for some function u_i' unique up to affine transformations and unique beliefs μ_i.

Now consider the set of actions inducing constant acts

$$\Delta^c(A_i) = \left\{ \alpha_i \in \Delta(A_i) \mid \rho_{\alpha_i,a_i}(x) = \rho_{\alpha_i,a_i'}(x) \text{ for each pair } a_{-i}, a_{-i}' \text{ and } x \right\}.$$

The game restricted to action inducing constant acts is also reduced, and by Lotteries independence, preferences $\succsim_i$ restricted to constant acts are a continuous weak order satisfying independence. With a slight abuse of notation, I denote the probability consequence x realises under the constant act induced by action $\alpha_i \in \Delta^c(A_i)$ with $\rho_{\alpha_i}(x)$. Again by Theorem 4 in Battigalli et al. (2017) preferences $\succsim_i$ over actions inducing constant acts can be represented by

$$U_i^c(\alpha_i) = \sum_x \rho_{\alpha_i}(x)\, u_i(x), \tag{8}$$

for each $\alpha_i \in \Delta^c(A_i)$, for some u_i unique up to affine transformations. Equations (7) and (8) imply that $u_i'(x, \alpha_{-i})$ should represent the same ordering as $u(x)$ for each α_{-i}, and therefore that preferences over actions in $\Delta^{\alpha_{-i}}$ are represented by

$$U_i^s(\alpha_i) = \sum_{a_i, a_{-i}} \alpha_i(a_i)\mu_i(a_{-i})u_i(\rho_{a_i,a_{-i}}), \tag{9}$$

for each α_{-i}.

Now consider the set of actions inducing the same act

$$\Delta^f(A_i) = \{\alpha_i \in \Delta(A_i) \mid \rho_{\alpha_i} = f\}.$$

By Universalisation evaluation, preferences $\succsim_i$ over $\Delta^f(A_i)$ must be consistent with preferences with constant acts. By Equation (8) and by Universalisation evaluation, these are represented by

$$U_i^c(\alpha_i) = \sum_{a_i, a_{-i}} \alpha_i(a_i) T_{\alpha_i^*, \alpha_{-i}^*}[\alpha_i](a_{-i}) u_i(\rho_{a_i, a_{-i}}), \tag{10}$$

for each $\alpha_i \in \Delta^f(A_i)$.

In the next step, I will "patch" the functions U_i^s in Equation (9) and U_i^c in Equation (10) into a unique function which is a convex combination of the two. In particular, I exploit the fact that the two functions are mixture linear on overlapping subdomains. Each mixed action α_i induces a pair of acts $\left(\rho_{\alpha_i}, \rho_{\alpha_i, T_{\alpha_i^*, \alpha_{-i}^*}[\alpha_i]}\right)$, where the second is a constant act. Pairs of acts lie in a connected convex domain with a product structure. By Theorem 2.2 in Chateauneuf & Wakker (1993), preferences $\succsim_i$ on the whole domain of mixed actions can thus be represented by

$$U_i(\alpha_i) = (1-\kappa) U_i^s(\alpha_i) + \kappa U_i^c(\alpha_i)$$

for each α_i, and the result follows.

□

Proof of Proposition 1. For the profile (α, α) to be a *SKE*, it must be the case that, for each mixed action α' and player i

$$\sum_{a_i, a_{-i}} \alpha(a_i)\alpha(a_{-i}) u_i(\rho_{a_i, a_{-i}}) \geq \sum_{a_i, a_{-i}} \alpha'(a_i)\alpha'(a_{-i}) u_i(\rho_{a_i, a_{-i}}). \tag{11}$$

Under *HK* preferences, player i evaluates action α by[20]

$$U_i(\alpha) = \sum_{a_i, a_{-i}} \alpha(a_i)\alpha(a_{-i}) u_i(\rho_{a_i, a_{-i}}).$$

For (α, α) to be a Nash Equilibrium in a game between two *HK*, it must be the case that, for each α' and i

$$U_i(\alpha) = \sum_{a_i, a_{-i}} \alpha(a_i)\alpha(a_{-i}) u_i(\rho_{a_i, a_{-i}}) \geq \sum_{a_i, a_{-i}} \alpha'(a_i)\alpha'(a_{-i}) u_i(\rho_{a_i, a_{-i}}) = U_i(\alpha') \tag{12}$$

Equation (11) and (12) are equivalent, one is satisfied only when the other is too, which concludes the proof. □

Proof of Proposition 2. For the profile (α_i, α_{-i}) to be a *MKE*, it must be the case that, for all players i and real numbers $r \geq 0$

[20] Preferences in games are usually represented by utility functions over action profiles. Since *HK* payoff only depends on his action, I can stick with my notation without loss.

$$\sum_{a_i,a_{-i}} \alpha_i(a_i)\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}) \geq \sum_{a_i,a_{-i}} r\cdot\alpha_i(a_i)r\cdot\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}). \tag{13}$$

Under *MHK* preferences relative to the profile (α_i, α_{-i}), player i evaluates action α_i by[21]

$$U_i\left(r\cdot\alpha_i\right) = \sum_{a_i,a_{-i}} r\cdot\alpha_i(a_i)r\cdot\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}).$$

For (α_i, α_{-i}) to be a Nash Equilibrium in a game between two *MHK*, it must be the case that, for all players i and real numbers $r \geq 0$

$$U_i(\alpha_i) = \sum_{a_i,a_{-i}} \alpha_i(a_i)\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}) \geq \sum_{a_i,a_{-i}} r\cdot\alpha_i(a_i)r\cdot\alpha_{-i}(a_{-i})u_i(\rho_{a_i,a_{-i}}) = U_i(r\cdot\alpha_i) \tag{14}$$

Equation (13) and (14) are equivalent, one is satisfied only when the other is too, which concludes the proof. □

I here prove a version of Proposition 3 that holds for all equal sacrifice rules.

Proposition 4. *Assume the game is symmetric. Then, if an action is optimal under ESU with $\kappa = 1$ with respect to any R, it is also optimal under HK.*

Proof of Proposition 4. I employ the notation of Appendix A. Pick a profile implementing the maximal expected utility for material consequences for player i, denoted $\left(\alpha_i^*, \alpha_{-i}^*\right)$. An action α_i^{Rs} inducing sacrifice s for rule R satisfies the following:

$$R^{-1}\left(R\left(U_i^c\left(\alpha_i^*, \alpha_{-i}^*\right)\right) - s\right) = U_i^c\left(\alpha_i^{Rs}, \alpha_{-i}^*\right).$$

Since the game is symmetric, the profile $\left(\alpha_i^*, \alpha_{-i}^*\right)$ also induces a maximal consequence for player $-i$ as $U_i^c\left(\alpha_i^*, \alpha_{-i}^*\right) = U_{-i}^c\left(\alpha_{-i}^*, \alpha_i^*\right)$. Then, the condition for equal sacrifice of $-i$ is equivalent to the one of i, that implies $\alpha_i^{Rs} = \alpha_{-i}^{Rs}$ for every s. The universalisation function of player i, if he has *ESU* preferences, is thus $T_{\alpha_i^*,\alpha_{-i}^*}\left[\alpha_i^{Rs}\right] = \alpha_i^{Rs}$, which is the same as the one of *HK*. Therefore, if an action is optimal under *ESU* with $\kappa = 1$, it is also optimal under *HK*. □

[21] The same point of the previous footnote holds. Since *MHK* payoff only depends on his action, I can stick with my notation without loss.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

On the Estimation of Climate Normals and Anomalies
Tommaso Proietti and Alessandro Giovannelli
CEIS *Research Paper*, 602 June 2025

Meritocracy as an End and as a Means
Enrico Mattia Salonia
CEIS *Research Paper*, 601 May 2025

Nudging Households' Sustainable Investments: Results from a Pilot Lab-in-the-field Experiment In Italy
Beatrice Bertelli, Marianna Brunetti, Costanza Torricelli and Mariangela Zoli
CEIS *Research Paper*, 600 May 2025

Identifying Belief-dependent Preferences
Enrico Mattia Salonia
CEIS *Research Paper*, 599 May 2025

Should I Share or Should I Not? On the Sharing of Information on Past Performance in Procurement
Gian Luigi Albano, Walter Ferrarese, Alberto Iozzi and Roberto Pezzuto
CEIS *Research Paper*, 598 April 2025

When the Going Gets Tough: the Impact of Health Shocks on Divorce
Javier Adrián López Artero, Anna Sanz-de-Galdeano and Daniela Vuri
CEIS *Research Paper*, 597 April 2025

A Stochastic Volatility Approximation for a Tick-By-Tick Price Model with Mean-Field Interaction
Paolo Dai and Paolo Pigato
CEIS *Research Paper*, 596 April 2025

Credit Markets, Corporate Governance and Growth
Emanuele Brancati, Paolo E. Giordani, Maurizio Iacopetta and Raoul Minetti
CEIS *Research Paper*, 595 March 2025

Using Multiple Tools to Enhance Competition in Public Procurement
Clara Calini, Alessandra Catenazzo and Elisabetta Iossa
CEIS *Research Paper*, 594 February 2025

Rage Against the Machine or Humans?
Luca Delle Foglie, Stefano Papa and Giancarlo Spagnolo
CEIS *Research Paper*, 593 February 2025

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

The opinions expressed in these publications are the authors' alone and therefore do not necessarily reflect the opinions of the supporters, staff, or boards of CEIS Tor Vergata.

COPYRIGHT

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595652/01 or by e-mail at piazzi@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact at +39 06 72595601 or by e-mail at segr.ceis@economia.uniroma2.it