THE UNIVERSITY OF CHICAGO

An Endogenous and Economic Model of Morality: Internalizing Externalities, Traditions, Nativism, and Moral Engineering

By

Thomas Zhang

July 2021

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Ryan Fang
Preceptor: Ryan Fang

Abstract

This study constructs a two-stage game in which players endogenously choose their morality in the first stage, and then actions in the second stage. It regards morality as a function that converts each player's actions into value judgments, and assumes that individuals prefer higher to lower value judgments. Players have an additional incentive for conducting actions related to their own and others' moralities. Players choosing moralities have an incentive to assign high value judgments to actions that have a positive externality on themselves, their own preferred actions, or actions that others' morality encourages. This study shows that the moral mechanism can be a handy tool in internalizing externalities. Allowing for moralities on others' choice of moralities yields traditions, expanding the equilibrium action allocation. The moral mechanism has a natural tendency toward varying extents of nativism, such as family values, nationalism, etc. This model provides a method for quantifying morality, as well as scientific moral engineering.

JEL codes:

Keywords: moral mechanism; value judgments; moral tolerance; moral allocation; morality on morality; externalities; nativism; moral engineering.

Morality and ethics emerge as a great interest in the field of economic analysis. However, most consider morality in absolute terms, such as in altruism (Rushton et al., 1981, 1984) and reciprocity (Sobel, 2005), or how Manski's (1993) social norm model assumes a natural tendency for people's actions to conform. Becker's (1974) theory of social interaction treats moral concerns as inherent in utilities. In contrast, this study only assumes inherent tastes, void of moral concerns; it derives moralities by having individuals **endogenously** select them, driven

by their desire for externalities from the actions of others, or for higher value judgments, such as recognition and applause from self and others. Philosophically, it takes a rather biological and materialistic approach that is inspired by Alexander's (1987) book, *The Biology of Moral Systems*, as I only consider private interest and the desire for recognition as inherent in humans, and then later derive moralities based on them. This is very different from how Brekke et al. (2003) build a moral motivation model that assumes inherent tastes for responsibility.

Let $i, j, k \in \{1,2, \dots, N\} = N$ denote the players. $x_i \in \mathbb{R}^k$ denotes the action vector of player i, containing k actions. Let the biological or genetic preference from actions be denoted as *natural utility* (NU), $NU_i(x_1, x_2, \dots, x_n)$, such that externalities are considered. Let $MU_i(\{m_j\}_{j \in N}, x_i) = \sum_{j \in N} w_{ij} m_j(x_i)$ denote the *moral utility* (MU) that captures the preference for higher *value judgment*, $v_{ij} = m_j(x_i)$, from j to i, where $m_j | \mathbb{R}^k \to \mathbb{R} \in M$ is the *morality* of j, M the moral choice set, and w_{ij} the *moral weight*/influence j has on i, or how much i cares about j's value judgments. Each player has utility function, $U_i(NU_i, MU_i)$, that captures both their NU directly from actions as well as their pursuit of higher value judgments from self and others. The *moral game* has two stages, wherein each player, i, in the first stage chooses m_i, while in the second stage, they choose x_i. Simple examples to facilitate a better understanding of the model are provided directly following the introduction.

This model is unique in the sense that it is based on the foundational assumption that people prefer higher value judgments from others, while each person, more or less, has the ability to provide such value judgments. Morality, then, in each player's hand, can be treated as some sort of tool endowed by nature, which serves to reward and punish own and others' actions. It provides an additional incentive for actions that are associated with goodness, honor,

acceptability, rather than for those deemed bad, evil, or unacceptable. Each person has this almost-free tool that can influence society's choice of action to a limited extent. One can draw a parallel between this model and the Coase theorem (1960), whereas moralities are birth-free, use it or toss it "side-payments" that can also be negative.

As with the Coase theorem, the moral mechanism can be a genetically endowed tool for human societies to internalize the externalities of most social interactions. Better than the Coase theorem, it can deal with some very unconventional and money-untransferable social actions, such as slavery, racism, and monarchy, to name a few significant actions that have involved moralities historically. Although people can quite costlessly assign value judgments to one another, the effect is limited by how much the other cares about their value judgment. Such limits then constrain how well the moral mechanism can perform in terms of internalizing large amounts of externalities. Panchanathan et al. (2004), Fershtman et al. (1998), and Kallbekken et al. (2010) show similar results in deriving that a certain social norm or social reward mechanism can lead toward Pareto efficiency.

Allowing players to assign value judgments based on how others assign value judgments, or moralities on moralities (MoM), creates simultaneous effects that expand the equilibrium moral and action allocation set, depending on what constraints are imposed on MoMs. If no constraint is imposed, then MoM allows limited deviation from the MoM, absent case actions equilibrium. If a linear constraint is imposed, then MoM leads to extremism and polarization. MoM may also contribute to conformity of actions.

This study provides a logic in why and how moralities and ethics change overtime. When certain actions that once served those with the most moral weight no longer serve new individuals with equal weight, then the old moralities will be replaced and overthrown by new ones. I like to use

the example whereby slavery, which once served the feudal lords best under an agricultural economy, then impeded the functioning of the labor market that best served the capitalists under an industrial economy. Consequently, slavery, as a moral standard, was overthrown and replaced, as the capitalists grew more influential than the old feudal lords and plantation owners. This study introduces concepts such as *range of moral tolerance* and *range of moral allocations* to describe the possible set of externalities and moral weight structures that is allowed under a specific moral allocation, and the set of moral allocations that is allowed under a specific externality and weight structure, respectively. A range of moral tolerance provides hints on how long and durable a certain moral allocation will last, while a range of moral allocations represents all the possible moral allocations in a specific historical setting.

The moral mechanism, together with MoM, causes natural rises of nativism to varying extents, such as family values and nationalism, when people close together assign each other higher value judgments. This relates back to how MoM forces conformity in relatively closed and connected societies, and how a morality that embraces all members in this society faces less resistance and gains more support. Nativist moralities induce morality utility efficiency, where everyone receives more weighted value judgments, or a better image of themselves.

Last, this study provides the possibility for scientific moral engineering, and some tools and ideas for it. Moral engineering can enhance the performance of the moral mechanism in solving externalities, and aids the interest group that has this tool on the moral battlefield.

Table 1: An Illustration of the Different Tendencies of Morality

	NU-Efficiency	MU-Efficiency	Moral Change	Others

Inherent/Natural Moral tendencies	Uncertain	Uncertain	Remains throughout	Uncertain
NU-driven Moralities	Yes	Uncertain/ No influence	None	None
MU-driven Moralities	Uncertain/No influence	A strong tendency	Uncertain	Nativism
Moralities on Moralities	Limited deviation, Extremism, Polarization	Yes, through Nativism	Traditions, Diversity	Conformity

Table of Contents

I. MODEL SET-UP

Example: Prisoner Dilemma with Simple Morality

To facilitate a better understanding of how the moral mechanism model functions, I provide two simple examples, one without MoM, and one with it.

Let $N = \{1,2\}$. In Period 1, each player, i, chooses $m_i(A) \in M \equiv \{f|A \to \left[-\frac{1}{2}k, \frac{1}{2}k\right] \subseteq \mathbb{R}, k \in \mathbb{R}_+ \ \forall i$. Each player in Period 2 chooses $a_i \in A \equiv \{C, D\}$, trying to maximize $U_i = NU(a_1, a_2) + w_{i1}m_1(a_i) + w_{i2}m_2(a_i)$; the NU payoff matrix is shown in Table 2.

Table 2: Prisoner Dilemma with Simple Morality - NU payoff matrix

1\2	C	D
C	(1,1)	(-1,2)
D	(2,-1)	(0,0)

For now, let as assume $w_{ij} = 1 \ \forall i \neq j \in N, w_{ii} = 0 \ \forall i \in N$. Let $a_1(A) = \begin{cases} \alpha_1 \ if \ C \\ \beta_1 \ if \ D \end{cases}, m_2(A) = \begin{cases} \alpha_2 \ if \ C \\ \beta_2 \ if \ D \end{cases}$; then, the subgame payoff matrix is as illustrated in Table 3.

Table 3: Prisoner Dilemma with Simple Morality - Subgame Payoff Matrix

1\2	C	D
C	$(1 + \alpha_2, 1 + \alpha_1)$	$(-1 + \alpha_2, 2 + \beta_1)$
D	$(2 + \beta_2, -1 + \alpha_1)$	(β_2, β_1)

Let $i \neq j$. We first observe that i always prefers j to play C rather than D, as i receives an additional payoff of 2, regardless of what action they take. We then observe that, as long as $\alpha_i -$

$\beta_i \geq 1$, j will choose to play C rather than D. Therefore, the equilibrium is $\{(\{m_i|\alpha_i - \beta_i \geq 1\})_{i=1,2}, \{C_1, C_2\}\}$, if $k \geq 1$, or $\{(m_i)_{i=1,2}, \{D_1, D_2\}\}$, if $k < 1$. In other words, if people have a large enough influence on one another, then the moral mechanism Pareto-improves the action allocation. If the influence is minimal, however, nothing will change. If, instead, the actions are divisible into smaller units, then the new outcome with the moral mechanism still Pareto-improves the action allocation, albeit to a smaller extent.

Moreover, fixing $k = 1$, and, if we have $w_{ij} > 1$ while $w_{ji} < 1$, keeping $w_{ii} = w_{jj} = 0$, then, whoever has the more moral weight is able to force others to choose C, while they themselves choose D. Consequently, the larger the projecting moral weight one has, the better outcome they can choose from.

In this simple model, we see how the moral mechanism works. The morality each has serves as a free additional incentive that they can impose on another. If they can, in turn, gain from others by assigning actions that bring them positive externalities with high value judgments, they will do so. The effect, however, is limited by how much the others care about the former's value judgment.

Example: Prisoner Dilemma with MoM

Let $N = \{1,2\}$. In Period 1, each player chooses $c_i \in [-k, k]$, $k \in \mathbb{R}_+$, while in Period 2, they choose $a_i \in A \equiv \{0,1\}$ (contribution to the public good), trying to maximize $U_i = NU(a_1, a_2) + c_{j\neq i}(a_i + c_i)$, where the NU matrix is the same as in Figure 1, with 0 corresponding to D, and 1 corresponding to C. Observe that, if both c_1 and c_2 are positive or negative, they feed off each other; else, there is a penalty for going against the moral direction of the other. Take the example

of morality regarding rape: if one considers rape evil, then, even if the other player has not committed rape, but actively supports it, that person receives the evil mark as well.

The subgame payoff is illustrated in Table 4.

Table 4: Prisoner Dilemma with MoM - Subgame Payoff Matrix

1\2	1	0
1	$(1 + c_2(1 + c_1), 1 + c_1(1 + c_2))$	$(-1 + c_2(1 + c_1), 2 + c_1c_2)$
0	$(2 + c_2c_1, -1 + c_1(1 + c_2))$	(c_2c_1, c_1c_2)

Let $j \neq i$. First, observe that i will play 1 only if $c_j \geq 1$. Then, observe that i gains 2 from j's playing 1 instead of 0. At least, i loses $-\frac{\partial U_i}{\partial c_i} = -c_j$ from playing $c_i \geq 1$. We thus conclude that i will play $c_i \geq 1 \leftrightarrow c_j \geq -2$. Consequently, under equilibrium, either both will play $c_i \geq 1$ or $c_i \leq -2$, and the corresponding equilibrium action allocations are $(1,1)$ and $(0,0)$, respectively. The equilibrium allocations are $\{(k,k),(1,1)\}$ or $\{(-k,-k),(0,0)\}$, if $k \geq 2$, $\{(k,k),(1,1)\}$, if $1 \leq k < 2$,and $\{(-k,-k),(0,0)\}$, if $k \leq 1$.

The greater the gains from a public good (2 here), the more "entrenched," in terms of morality, the requirement to keep the alternative tradition. Moreover, if k is too large, people may be stuck in a tradition or place forever. If k is too small, the moral mechanism fails to work; only in the middle does it guarantee efficiency while simultaneously preventing people being stuck at an inefficient allocation.

Math Set-up

Let $\{1,2,\dots,N\} \in N$, $N \geq 2$ denote the set of players. For convenience, let $i,j,k \in N$ denote individual players.

Let $x_i \in \mathbb{R}^t$ denote the **action** set of Player i, where $t \in \mathbb{I}$ is the number of different types of actions. Let $c_i(x_i) \leq I_i$ be the budget constraint for Player i, where c_i is a continuous, differentiable, and strictly positive monotonic cost function, while I_i is a positive real budget. Let $\theta_i \in \Theta$ be the **information vector** of i, where $p \in I$ is the number of information pieces. Actions can be any actions, such as crime, education level, research published, slave duties, other forms of tribute or redistribution, family responsibility, nationalist support, and so on. In contrast to actions, information is relatively fixed and determined, while the players cannot choose it in Stage two. Additionally, information does not enter into NU. Examples include skin color, sex, place of birth, and so on.

Let $m_j(x, \theta)|\mathbb{R}^t * \mathbb{R}^p \rightarrow \mathbb{R}_{\geq 0}, m \in M$ be the **morality** of *Judger j* that converts the actions and information of the *judged, i*, into value judgments, and $v_{ij} = m_j(x_i, \theta_i)$ be the **value judgment** of j on i. Let M be the moral choice set wherein all moralities are continuous, differentiable, and monotonic, with constraint $mc_{low} \leq m(x_i, \theta_i) \leq mc_{up}$, $mc_{low}, mc_{up} \in \mathbb{R}, \forall x_i$ feasible, $\forall \theta_i$. When two players appear in the subscript, the judged always appears on the left, and the judger on the right; when there is only one player in the subscript, that player always refers to the judged.

I say that a morality, m, is **based on** action or information, x^q, if and only if $\frac{\partial m}{\partial x^q} \neq 0 \; \forall x^q \in \mathbb{R}$.

Optionally, let $aV_{ij} = \frac{v_{ij}}{\sum_{k \in N} v_{kj}}$ denote the **adjusted value judgment** of j on i, to account for how each cares about their relative stance to others. Relative identity comparison is discussed by Tajfel et al. (1979). For adjusted value judgments, modify the moral choice set and shift the value judgments up, for all to be positive.

Let $w_{ij} \in \mathbb{R}_{\geq 0}$ be the **(moral) weight** of Judger j to Judged i. Let all w_{ij} be fixed and given. Let $cV_i = \sum_{j \in N} w_{ij} v_{ij}$, or $cV_i = \sum_{j \in N} w_{ij} aV_{ij}$, if the *option* is used, be the **combined value judgment** received by i. I make an implicit assumption that combined value judgments are linear in weights. This may not be entirely true in real life, as perhaps the smaller the interest or conspiratorial group that adopts the same morality, the lower the organization cost, and thus the more powerful their combined value judgment.

Let $NU_i = (x_1, x_2, \ldots, x_n)$ denote the **NU**, or non-moral utility, of Player i, and $MU_i = cV_i$ denote the **MU** of Player i. When talking about some Player i's weight, if without a prefix, let it refer to the *projecting weights*, or $w_{ji}\ \forall j \in N$; else, let *receiving weights* refer to $w_{ij}\ \forall j \in N$. Let the NU function be continuous, differentiable, and positive monotonic in all variables. Additionally, assume NU_i is **convex** or, for any consumption bundles, $x, y, z \in \mathbb{R}^t$, where $NU_i(y) \geq NU_i(x), NU_i(z) \geq NU_i(x)$, and $y \neq z$, then, for every $\theta \in (0,1)$, $NU_i(\theta y + (1-\theta)z) \geq NU_i(x)\ \forall i \in N$.

Let $U_i = U_i(NU_i, MU_i)$ be the **utility** of Player i, and U_i be continuous, differentiable, and positive monotonic in the two variables. Here, I assume that NU and MU are separable. The separability assumption mainly serves interpretation of the efficiency of NU and MU.

Let an **allocation**, $A = \{(m_i)_{\forall i \in N}, (x_i)_{\forall i \in N}\}$, refer to a set of moralities and goods consumed for all i and j. Let a **moral allocation** refer to $\{(m_i)_{\forall i \in N}\}$. Let an **action allocation** refer to $\{(x_i)_{\forall i \in N}\}$.

Let an action allocation be **feasible** if $c_i(x_i) \leq I_i\ \forall i \in N$. Let a moral allocation be **feasible** if $m_i \in M\ \forall i \in N$. Let an allocation be **feasible** if both the moral allocation is feasible, and the

action allocation is feasible and constitutes a subgame perfect Nash equilibrium, given the moral allocation.

An allocation, $\{(m_i)_{\forall i \in N}, (x_i)_{\forall i \in N}\}$, is ***NU*-efficient**, or **efficient**, if there exist no other feasible allocation, $\{(m_i')_{\forall i \in N}, (x_i')_{\forall i \in N}\}$, such that $NU_i(x_1', x_2', \dots, x_n') \geq NU_i(x_1, x_2, \dots, x_n)\ \forall i \in N$, with $NU_i(x_1', x_2', \dots, x_n') > NU_i(x_1, x_2, \dots, x_n)$ for some i.

An allocation, $\{(m_i)_{\forall i \in N}, (x_i)_{\forall i \in N}\}$, is **true *NU*-efficient** if there exist no other feasible action allocation, $\{(x_i')_{\forall i \in N}\}$, such that $NU_i(x_1', x_2', \dots, x_n') \geq NU_i(x_1, x_2, \dots, x_n)\ \forall i \in N$, with $NU_i(x_1', x_2', \dots, x_n') > NU_i(x_1, x_2, \dots, x_n)$ for some i. True efficiency ignores the constraints by the moral choice set. This is provided to account for the limited influence of morality on others.

Similarly, an allocation, $\{(m_i)_{\forall i \in N}, (x_i)_{\forall i \in N}\}$, is ***MU*-efficient** if there exist no other feasible allocation, $\{(m_i')_{\forall i \in N}, (x_i')_{\forall i \in N}\}$, such that $MU_i((m_i')_{\forall i \in N}, (x_i')_{\forall i \in N}) \geq MU_i((m_i)_{\forall i \in N}, (x_i)_{\forall i \in N})\ \forall i \in N$, with $MU_i((m_i')_{\forall i \in N}, (x_i')_{\forall i \in N}) > MU_i((m_i)_{\forall i \in N}, (x_i)_{\forall i \in N})$ for some i.

Let a **moral game** have two stages, wherein each player maximizes U_i. In the first stage, each player, i, decides m_i. In the second, each decides x_i.

Let a **NU moral game**, or ***NU*-moral game**, have two stages, wherein in the first stage, each only intends to maximize NU_i. Nevertheless, in the second stage, when deciding their actions and information set, each aims to maximize U_i.

Let a **status-only moral game**, or ***MU*-moral game**, have a single stage, wherein each player maximizes MU_i, given the actions and information set for all players.

Let a **moral-absent game** have a single stage, wherein each player maximizes NU_i by choosing x_i under budget constraints.

An **equilibrium** refers to a subgame perfect Nash equilibrium.

Let **simple moralities** be those that are not based on information dependent on the weight or moral choices of any players. That is, for any information piece, $\theta_i^{\alpha},\ \alpha \in \{1,2,\dots,p\}$, that m_j is based on, θ_i^{α} is independent of m_k or $w_{lk}\ \forall i,j,k,l$. Moralities that can be based on such information are **complex moralities**. **MoM** is a complex morality that this based on a piece of information, $\theta_i^{\alpha}(m_i), \alpha \in \{1,2,\dots,p\}$, dependent on the moral choice of oneself. An example of MoM is anti-racism: racism can discriminate based on color, while anti-racism discriminates based on the moral choice of racism.

Assume that, in all types of moral games, if a player is indifferent between moral choices, among those moral choices, they will choose a NU-efficient one. This is so because, if the other player exerts a tiny amount of pressure, the former player will do so.

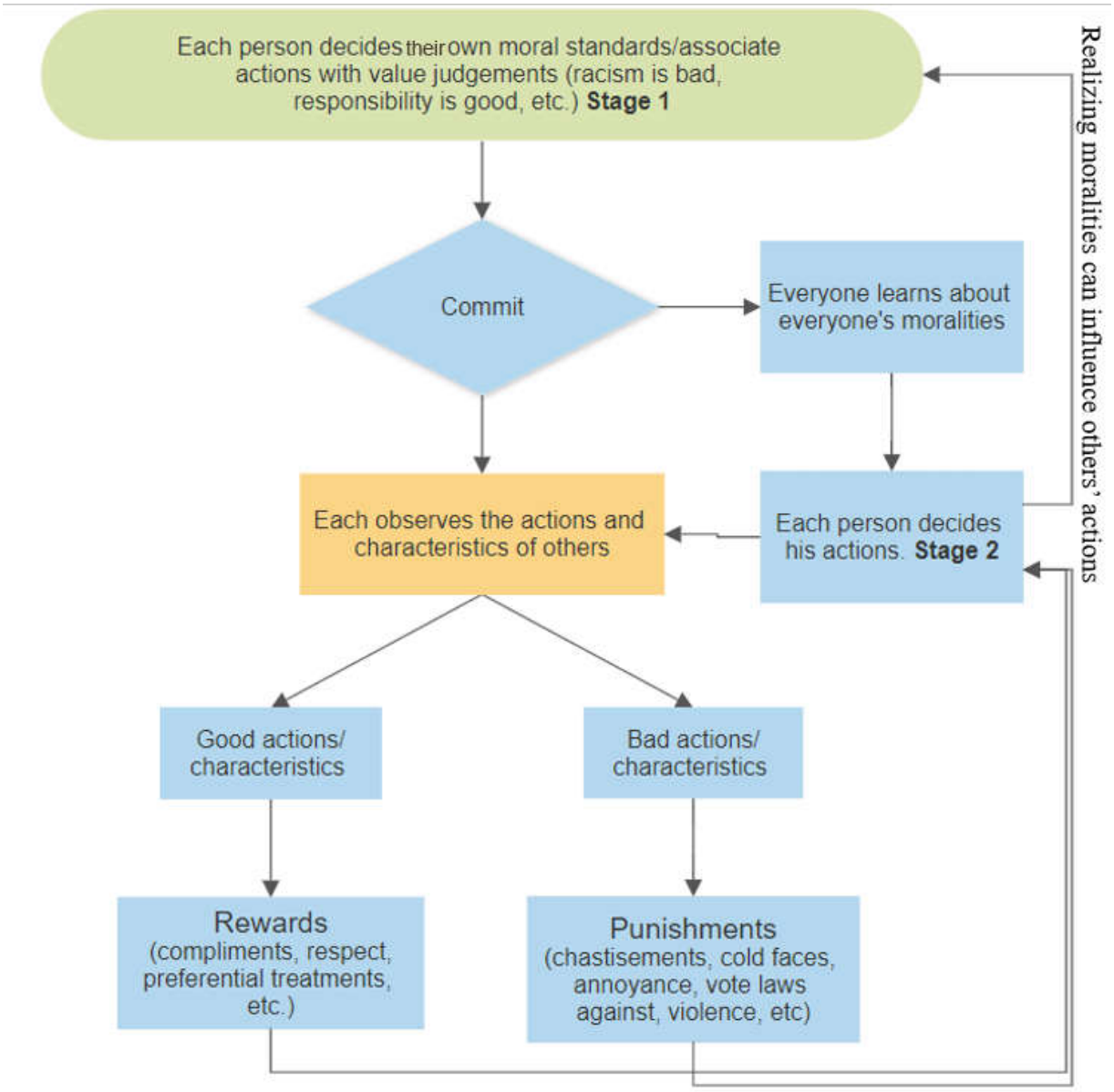

Figure 1: Flowchart of the Moral Game

Justifications

Morality is a controversial topic. The greatest obstacle to a model on morality is to justify itself as moral. Some may not agree that people, indeed, on the individual level, do make moral

choices endogenously in real life, as in this model. Others may deem morality based on private selfish interest heretical. Below, I provide nine justifications from different aspects to claim that this model, indeed, reasonably captures how morality functions.

First, I define morality as a function that assigns actions and information value judgments. I believe this definition captures what people consider to be morality. Most moralities, even nihilism, can be described as a set of moral and value judgments on things. Examples of value judgments include good, bad, right, wrong, kind, evil, superior, inferior, shall, shall not, meaningful, meaningless, and so forth. This model accounts for this portion of morality.

Second, I believe the greatest impact of morality on the physical world and people's actions is through encouraging and discouraging certain actions. Imagine that, suddenly, all moral notions disappeared: how would society behave differently? I conceive there would be fewer boundaries for people to conduct actions that were previously associated with bad value judgments, and to conduct fewer actions associated with good value judgments. This portion of morality is captured in how moralities function as additional incentives for actions associated with good value judgments and, vice versa, as disincentives for bad ones.

Third, people's moral notions not only guide the actions of self, but also those of others. Many people, especially under the influence of liberalism, reject this. This is utterly false. People naturally impose their own moral belief on the actions of others, rewarding good people with more favor, such as more encouragement, approval, recognition, praise, smiles, nepotism, support, vote laws rewarding, and so on, and treating bad people with less favor, such as mean behavior, annoyance, cold faces, verbal assault, physical assault, vote laws punishing, and so on. Even an extreme individualist who opposes interfering in the lives of others may yield to the

temptation of punishing those who meddle in their affairs, or even worse, may choose not to punish a person who has committed atrocious crimes, such as rape, murder, and cannibalism.

Fourth, peoples' actions are indeed influenced by the moralities of others, although some to a greater and others to a lesser extent; to what extent is captured through moral weights. Family, friends, direct supervisors, media magnates, those who control the law and police, etc. have larger projecting moral weights than others. As a matter of existence, of course, there exist persons whose morality has the ability to influence others' actions. Existence is out of the question. The question that remains is how and to what extent people's morality influences the actions of others.

Fifth, I provide some details on how, exactly, personal moral notions translate into action changes or real-life incentives for others. Generally, there are psychological and material paths, and non-institutional and institutional paths, through which such transformation is achieved. For the psychological path, I refer to the psychological need of humans to seek recognition, praise, and superiority from others, as opposed to humiliation, denial, or isolation. Psychological effects occur as long as the value judgments enter into the ears or eyes of the judged. In this sense, the moral weight structure depends on the frequency of exposure of one's value judgment to another. The more frequent the exposure, for example, among those within a family, between friends and colleagues, or in a media broadcast, the more powerful the effect occurs. The cost of such form of moral effect is very low, as only communication and related infrastructure are used. For the material path, I refer to the consequent real-life actions one may take against those they consider good or evil. For example, a good slave is less beaten, or someone receives preferential treatment because of their color. Compared to psychological methods, material methods are more costly. The difference between non-institutional and institutional paths are quite self-explanatory. Non-

institutional paths are more organic, such as friend relationships, families, etc.; the costs are relatively low. Institutional channels include the legal system, government, the police, the military, media and the internet, religious institutions, the education system, company policies or culture, private social organizations, voting mechanisms, and so on. Such type of moral execution is relatively costly, as there are additional organization costs. Such a form of moralities also exhibits the characteristic of projecting moral weight that is concentrated in the hands of a few powerful individuals, and a uniform projecting of moral weights, or $w_{ij} = w_{kj}\ \forall i, k$, where every member of the affected population receives relatively similar treatments and, thus, incentives regarding such actions. This can become a favorable trait in deriving efficient equilibrium action allocation, as will be shown later. This study does not include the cost of transforming one's value judgments into incentives, as it further complicates the math, and many of the moral weights are endowed and cannot be converted to money or traded, for example, the guardianship authority of parents granted by law on their children; instead, it assumes fixed moral weights, partially justified by how they are short-term fixed in real life.

Sixth, I include three different incentives for individuals when making their moral decision. I prove, by example, that in real life, each of these incentives exists and makes sense. The first is NU_i or private interest from the externalities of others. For private interest, slave owners definitely have private interest in maintaining moral systems that support slavery, especially when they have no better alternative ways of investment. A person has an incentive against criminal behavior that may one day hurt themselves. The second is v_{ii} in MU_i, or self-righteousness. This self-righteousness can stem from both the morality itself, as a MoM, $m_i(\theta(m_i))$, that is, one is trusting right because it is right to trust right, or from possible superior feelings against others. One example of the former is missionaries preaching because what they

preach associates the act of preaching with a high value judgment. One example of the latter is how one may have an added incentive to be a racist, elevating the status of their own color, because having others to look down upon may bring psychological pleasure and relief. The last is $v_{ij}\ \forall j \neq i$ in MU_i, or pressure from others. In medieval Europe, many converted their religion not because of some divine inspiration, but due to real-life threats from the state, the church, and their neighbors, who might well have burnt them alive, collected more tax, or imposed discriminatory treatment if they did not. Besides the three in this model, there may be other incentives for moral decision. A notable one is some form of "human nature." Almost every morality claims that it is part of human nature. Yet, first, as Alexander (1987) argues, human nature cannot be the deterministic cause of the social choice of morality: else, either the moral allocation does not change since human nature does not change, which is against what history shows, or such nature is not so "nature" of itself, and requires some other exogenous element, such as revelation, which changes through time. In this case, that changing element then becomes the deterministic factor of morality, while the human nature part is not as important. Second, from an economist's standpoint, even if there is such human nature, as long as there exist other factors of influence, such as the three points described above, such human nature but only partially influences the moral outcome. I did not incorporate natural preferences for morality into the model; however, a simple shift of all individuals' morality to the direction of human nature will stimulate the effect. Since human nature is quite arbitrary and not dependent on either NU or MU, the effects are arbitrary as well.

Seventh, I believe individuals adopt moralities based on a certain moral choice set that they are aware of, instead of inventing them on the spot. It would be unthinkable to assume that people drafted up their own unique moralities out of nowhere, just to satisfy their private interest. This

model solves this question through the notion of a moral choice set, or the set of moralities one is exposed to and can choose from. I separate drafting and picking moralities into two different tasks. Professionals, such as theologians, political theorists, social movement leaders, and others, create new moralities and spread them. Ordinary people choose among these those that they feel are best for them. Note that most moralities that professionals create only provide a general direction of value judgments that are associated with actions. People are free to choose the extent to which they carry them out. Some may agree with a morality so superficially that they will not carry out any actual responses when they see injustice or justice. In this model, the moral choice set also serves the technical purpose of limiting the magnitude of execution of a morality. Fanatics may very well sacrifice their lives for their moral doctrines; however, for others, this may not be the case. There is a limit to how much moral influence one can exert on another. Fixing either the moral weights or the maximum value judgment bounds the other. As reasoned previously how individuals choose freely between the maximum and minimum value judgments, the moral choice set shall be convex.

Eighth, I am convinced that individuals do, indeed, determine their own morality. The idea behind the moral mechanism is that regardless of whether there are exogenous variables, such as genetics or God, in determining morality, it remains up to individuals to execute their moral actions, or corresponding rewards and punishments according to their morality. Since individuals execute moral actions, they are afforded an opportunity to realize consciously or subconsciously that they have the ability to change them. As long as they realize this ability, even subconsciously, and have incentives and preferences among the different moralities they can take, they will do so. I claim that this is a truly economic stance: as long as people realize they can change something, and they have incentives for acting one way instead of another, they do.

In a large society with MoM, there are rewards for following the mainstream moralities, and punishments for going against them. This may contribute to people's notion that others, instead of themselves, are determining their own morality. Either way, the mainstream ideology part is endogenized and included in the model. If one must believe the other way around, that is, that society collectively votes on one's morality, indeed the whole model, along with its conclusions, can remain, as long as one normalizes w_{ij}s such that $p_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}$ represents the probability that Player i acts according to Player j's choice of morality, and all players play the game in interim terms, maximizing the expected utility, $E[U_i\big(NU_i, (\sum_k w_{ik}) \sum_j p_{ij} m_j(x_i)\big)]$.

Ninth, this model of morality allows for contradictory moralities to occur, as in history, and provides a passable story. Take slavery as an example: once, it represented order and good; now, it represents evil and oppression. A realistic model of morality shall at least be able to model both these situations, and attribute any difference to some reasonable cause. Alternatively, praising the current set of moralities as absolute will definitely fail both for future and past analysis. Fortunately, in my model, there are only three drivers of morality in people , and only the NU part is truly exogenous; therefore, I can address most of the changes to NU. For example, I can argue that slavery may represent the most profitable form of investment in a certain stage of development; however, as better alternatives emerge, such as industrial investment, which responds to a free labor market, the old morality is replaced by a new one. This is but an oversimplified story; however, it is a passable story on how this model of morality accommodates contradictory mainstream moralities. Additionally, this model allows contradictory moralities between different persons, by treating moralities as personal choices.

This model is fundamentally different from attempts to use direct or indirect reciprocity to explain morality. The individuals in this game are not looking forward to, nor do they have a belief in the reciprocal actions of others. Take slavery, racism, or colonialism as morals, for example. Neither a slave nor a slave master chooses, respectively, to be a slave or to enforce slavery expecting a reciprocal act. Racists or victims of racism do not expect direct or indirect reciprocity or cooperation of any form. Colonies do not have the choice of whom to trade with, or, from the beginning, the choice to trade or not trade. These moral outcomes are forced, contrary to the liberalist assumption in mainstream economics, whereby each individual often has a "standalone" or "to not be messed by others" noncooperative default option. This model realizes this aspect of morality and incorporates these artificial reward and punishment systems into the decision process of individuals acting and choosing moralities.

Interestingly, Andreoni (1988) wrote on how altruism failed to work in large populations. This model on morality can provide a good alternative explanation for acts that seem altruist but, instead, are NU motivated. Truly altruistic preferences enter appropriately into the NU externality structure, just as slavery or public goods do.

II. SIMPLE MORALITIES

Assume, for this chapter, that the moral choice set is limited to simple moralities.

Proposition 1 A moral game or NU-moral game equilibrium allocation that is NU-efficient is never NU-Pareto dominated by a moral-absent game equilibrium allocation with the same utility and cost functions and moral weights.

Proof The moral-absent game equilibrium allocation can be achieved in a moral game or NU-moral game by each player setting $m_i(x_i)$ equal to a constant, such that the consequent value judgments do not depend on each individual's actions. Since the moral game or NU-only moral game equilibrium allocation is NU-efficient, it cannot be Pareto dominated by such an allocation.

QED

This proposition partly points out that the moral mechanism is never a bad thing to have, given NU-efficiency. Only under very rare and extreme circumstances does the moral mechanism do bad. Afterall, the moral mechanism is a mechanism that internalizes externalities.

Public Good Moral Game

Let x be a public good and y be a private good, such that $U_i = U_i(a\sum_{j\in N} x_j + y_i, cV_i)\ \forall i \in N$, where a is a constant between 0 and 1. Let each have identical cost functions and budgets.

Let a NU-moral game, with the conditions described above, be a **public good moral game**.

Proposition 2 If $w_{ij} = w_{kj}\ \forall j \in N, \forall i, k \in N$, then a public good moral game equilibrium is NU-efficient.

Proof Since the utility and cost functions are identical for every player, and $w_{ij} = w_{kj}$, then, for any morality adopted by any player, each player consumes identical amounts of the public and private goods.

We prove NU-efficiency through contradiction. Suppose there exists an inefficient equilibrium. This implies that every player can be better off. Yet, any change will result in either an increase or a decrease in the consumption of the total public good, thus increasing or decreasing every

player's utility. Due to the convexity of NU_i, a local max is also a global max. Therefore, at least one player has an incentive to change their morality such that they are better off. The existence of an allocation that Pareto dominates it, and how information is public shows that at least one player can shift their morality such that everyone is better off. This shows that the inefficient equilibrium is not an equilibrium. Contradiction

QED

Although the action allocation in the equilibrium is unique, the moral allocation is not. All can equally support the public good, or it is possible that some support it more, with others advocating against it. People do not have to agree on the same morality to reach efficiency in a public good moral game. If a few become too morally pure and fanatic, others have an incentive to advocate against it on an absolute scale.

When the "adjust to average" *option* is applied, and MU is included in the decision process of the first period, individuals with larger budget constraints, or those who can afford the public good at a lower cost, knowing that they would consume more of the public good in Stage two, have an additional MU incentive to set the moral standards higher, to receive higher adjust-to-average value judgments. This can be one explanation why the rich may have higher moral standards, and may often be crowned more virtuous than the rest.

Corollary 2.1 With large enough moral constraints, and $w_{ij} = w_{kj}\ \forall j \in N, \forall i, k \in N\ s.t\ \forall \theta_i = \theta_j$, the public good moral game leads to true NU-efficiency.

Proof As long as we find that there exists a set of moralities for all players that yields true NU-efficiency, then, by Proposition 2, the proof completes. This is done by finding $m_j\ \forall j \in N$ such that the second stage condition would yield true NU-efficiency. By eyeballing, we arrive at

$\sum_{j\in N} w_{ij}\frac{\partial m_j}{\partial x} = a(n-1)$ and $\frac{\partial m_j}{\partial y} = 0$. This simplifies to $\frac{\partial m_j}{\partial x} = a(n-1)/\sum_{j\in N} w_{ij}$, if we set all m_j equal. By setting $m_j(0) = 0$, and $\frac{\partial m_j}{\partial x} = a(n-1)/\sum_{j\in N} w_{ij}$ for the domain of x, we find a set of $m_j\ \forall j \in N$ with Bounds $b_{up}^2 \geq (n-1)/\sum_{j\in N} w_{ij}$ and $b_{up}^1 > x_{max}a(n-1)/\sum_{j\in N} w_{ij}$, where x_{max} is the maximum amount of x each player can consume, and a true NU_i-efficienct equilibrium yields. By Proposition 2, the public good moral game yields true NU_i-efficiency with large enough bounds.

QED

This corollary gives a glimpse of the relationship between the moral choice set and true efficiency. There is an upper bound or limit of how much externality the moral mechanism can internalize. In the public good moral game, the larger the aggregate moral weight or number of players or moral constraint, the more likely true efficiency is achieved. Once true efficiency is reached, larger governments or other moral institutions and infrastructure do not help. This pattern is, generally, true.

Private Goods Moral Game

Let x and y both be private goods, such that $U_i = U_i(x_i, y_i, cV_i)\ \forall i \in N$.

Let a NU-moral game, with the conditions described above, be a **private good moral game**.

Proposition 3 There exists a NU-efficient and true NU-efficient equilibrium allocation in the private good moral game. Moreover, all such equilibrium allocations are NU-efficient.

Proof As an example for existence, the equilibrium wherein each player sets morality to a constant is such an allocation. The efficient and true efficient action allocation is where each

player only maximizes NU in the second stage. If not, in the first stage, whoever has a morality that pushes others to consume more of one of their private goods shall do so less, as their consumption of private goods does not affect this player's utility. This is because of the assumption that if a player is indifferent between moral choices, among those moral choices, they will choose a NU-efficient one.

QED

The private good moral game shows that the moral mechanism will really not have an effect on and interfere with non-social actions, if such moralities are NU-driven. In other words, moralities that affect truly private actions, with no externalities, most likely involve MU pursuits, and are less likely to exist. For example, there are status goods and conspicuous consumption (Veblen et al., 1899; Mason, 1980). This happens when the "adjust to average" *option* is applied, while those who know that they will consume more of such a luxury good often have more moral weight, and will push society in such a direction to elevate their status in the eyes of much of the population. Interestingly, when the poor have more moral weight, luxury goods may then serve as reverse status goods, and mean "evil," as in Maoist China (Chan et al. 1992). This adds validity to the theory that some seemingly innocuous and "natural" tastes for luxuries may not be inherent, but may depend on whether those with power, and who control the most moral weight benefit from luxury consumers being admired.

Personal Favor Moral Game

Let k be the special player receiving the personal favor. Let x_i be i's personal favor done for l, and y the private goods, such that $U_i = U(y_i, cV_i)\ \forall i \in N/\{k\}$ and $U_l = U_l(\sum_{j\in N} x_j, y_i, cV_i)$.

Let a NU-moral game, with the conditions described above, be a **personal favor moral game**.

Proposition 4 The personal favor moral game equilibriums are NU-efficient and true NU-efficient.

Proof There are two cases. In the one case, it is not profitable for Player k to set a morality that encourages consumption of the personal favor, due to the consequent loss of their own private goods consumption. If so, they will not, and everyone is happy. In the other case, wherein it is profitable for k to do so, they would try their best to set a morality that encouraged others to contribute to the player's personal favor. Others would try their best to counter such morality. The equilibrium is NU-efficient and true NU-efficient, because for k to be better off, others must be worse off, while, for others to be better off, k must be worse off.

QED

Note that, in the worthy case in which k prefers others to do more personal favors for them, the former must overcome the opposition from the rest of the population on whom they attempt to impose their morality. The special player receiving the personal favor must have an immense moral weight advantage over the aggregate population to achieve the goal of others doing them personal favors, or else none will do so. This may explain why personal favors are not broadly considered "moralities" or mandatory, unless the person is extremely powerful, such as a monarch, or the population that the morality affects is fairly small, such as a family or company. "Selfish" requests without the appropriate power to enforce them rebound quickly.

Interest Group Moral Game

Let $A_1, A_2, .. A_t$ be partitions of N, x_i be i's contribution to the interest of members in Group A_1, and y the private goods, such that $U_i = \left(\sum_{j \in N} x_i^{A_q}, y_i, cV_i\right) \forall i \in \{A_q\}\ q \in \{1,2, \dots, t\}$. Let the utility functions and constraints for all the players in Group A be identical.

Let a NU-moral game, with the conditions described above, be an **interest group moral game**.

Proposition 5 Under $w_{ij} = w_{kj}\ \forall j \in N\ \forall i, k \in A_q\ q \in \{1,2, \dots, t\}$, a group interest moral game equilibrium is NU-efficient.

Proof Since $w_{ij} = w_{kj}\ \forall j \in N\ \forall i, k \in A_q\ q \in \{1,2, \dots, t\}$, and the utility functions and constraints of any player in Group A are identical, members in the interest group consume identical x and y.

Proof by contradiction. Suppose an equilibrium allocation is not NU-efficient: then, there are two cases. In Case 1, a player in A can gain more NU, with others remaining at least indifferent. Since all the players in A have identical x and y, this implies that all the players in A can gain more NU, while the members that are not in the interest group, A, are at least indifferent. This can only occur when each member who is not in the group consumes equal or less x; however, all players then have an incentive to shift morality to consume less x. Since value judgments enter into cV linearly and independently of each other, at least one has an incentive and can change their morality. Hence, it cannot be an equilibrium. In Case 2, a player who is not in Group A can gain more NU, while others remain at least indifferent. Using similar reasoning, this is not an equilibrium. Contradiction.

QED

An interest group moral game resembles an expanded version of a personal favor moral game, wherein, instead of one player, many benefit from a specific interest group good. The interest group will ask others to serve them at their cost, while the non-interest group members resist. The final outcome depends on which group has more aggregate projecting moral weight, and will mostly likely result in a compromise, if only simple moralities are allowed. Yet, irrespective of who holds more weight, the outcome is always efficient. MoM reduces moral conflicts at some cost in efficiency.

All goods with externalities may be viewed as layers of group interest goods. Depending on utility functions and budget constraints, a good can benefit a group of individuals while, simultaneously, it benefits a member of the group more than others. For example, a monarchy benefits not only the monarch, but also the feudal lords, their relatives, bureaucrats, and even their head servants over others, to varying extents. For such a morality to be maintained, the aggregate moral weight of the participants who benefit more from this morality than from the "opportunity morality" must be greater than the aggregate moral weight of those who benefit more from the "opportunity morality." Or else the old morality will be overthrown and the new one established. In this sense, moralities are written by those with more aggregate moral weight.

Interest group goods, additionally, tie back to how actions and externalities are constructed. Take the example of slavery. In practice, a slave's duties only benefit their master. Yet, slavery requires considerable moral weight, or organized violence in this case, to enforce. A master may or may not have such violence on hand to keep slavery functioning. However, by creating an abstract idea of slavery across all slaves and masters, the masters signed a mutual support agreement in sharing violence. In this model, slave duties are then elevated to an abstract interest group good that benefits all masters, perhaps to varying extents. The shared violence can then be

expressed in a larger aggregate weight. Criminal events can also be abstracted similarly; thus, one does not only oppose crime against themselves, but abstractly, against all.

Linear Moralities

Definition By saying **only allowing linear moral functions** in any type of moral game, the moral choice set of the game is limited to linear functions in all its inputs.

Theorem 3 Only allowing linear moral functions, if $w_{ij} = w_{kj}\ \forall i,k \in A\ \forall j \in N$, then an interior NU-moral game equilibrium allocation is NU-efficient.

Proof Since the morality functions are linear in each variable, and value judgments are linear to each person's utility function, due to how the moral mechanism is constructed, any change in the morality of one or more persons must result in all players together consuming more/less/the same of a type of good.

Proof by contradiction. Suppose there exists an equilibrium allocation that is not NU-efficient. Let the allocation be $\{(m_i)_{\forall i\in N}, (x_i)_{\forall i\in N}\}$, and the allocation that NU-Pareto dominates it be $\{(m_i')_{\forall i\in N}, (x_i')_{\forall i\in N}\}$. I claim that there exists a positive integer, t, such that, in the original allocation, at least one player can change their morality, such that the final allocation becomes $\left(\frac{1}{t}x_i + \left(1-\frac{1}{t}\right)x_i'\right)_{\forall i\in N}$. The proof follows as below.

By the convexity assumption, the allocation is interior; we can apply the Karush–Kuhn–Tucker (KKT) theorem. Each player must satisfy the following condition in the subgame perfect Nash equilibrium:

$$\left(\frac{\partial U_i}{\partial NU_i}\frac{\partial NU_i}{\partial x_i^a} + \frac{\partial U_i}{\partial cV_i}\sum\nolimits_{j\in N} w_{ij}\frac{\partial m_j}{\partial x^a}\right)\Big/\frac{\partial c_i}{\partial x^a} = \left(\frac{\partial U_i}{\partial NU_i}\frac{\partial NU_i}{\partial x_i^b} + \frac{\partial U_i}{\partial cV_i}\sum\nolimits_{j\in N} w_{ij}\frac{\partial m_j}{\partial x^b}\right)\Big/\frac{\partial c_i}{\partial x^b}$$

, for any goods, a and b. Since moralities must be linear, $\frac{\partial m_j}{\partial x^a}$ must be a constant for any good, a. To have the two allocations different, either $\sum_{j\in N} w_{ij} \frac{\partial m'_j}{\partial x^a} > \sum_{j\in N} w_{ij} \frac{\partial m_j}{\partial x^a}$ or $\sum_{j\in N} w_{ij} \frac{\partial m'_j}{\partial x^a} < \sum_{j\in N} w_{ij} \frac{\partial m_j}{\partial x^a}$; either way, since $w_{ij} = w_{kj}\ \forall i,k \in A\ \forall j \in N$, due to the convexity of the subgame perfect Nash equilibrium to the morality of each player, there must exist a player, k, with feasible morality, m_k^*, such that $\sum_{j\in N} w_{ij} \frac{\partial m'_j}{\partial x^a} \geq \sum_{j\in N/\{k\}} w_{ij} \frac{\partial m_j}{\partial x^a} + w_{ik} \frac{\partial m_k^*}{\partial x^a} \geq \sum_{j\in N} w_{ij} \frac{\partial m_j}{\partial x^a}$ or $\sum_{j\in N} w_{ij} \frac{\partial m'_j}{\partial x^a} \leq \sum_{j\in \frac{N}{\{k\}}} w_{ij} \frac{\partial m_j}{\partial x^a} + w_{ik} \frac{\partial m_k^*}{\partial x^a} \leq \sum_{j\in N} w_{ij} \frac{\partial m_j}{\partial x^a}\ \forall i \in N$, since $w_{ij} = w_{kj}\ \forall i,k \in A\ \forall j \in N$; however, the resulting solution can then be described as $\left(\frac{1}{t}x_i + \left(1-\frac{1}{t}\right)x'_i\right)_{\forall i\in N}$.

Since the preferences for each player are convex, the allocation, $\{(m_i)_{\forall i\in\frac{N}{k}}, m_k^*, \left(\frac{1}{t}x_i + \left(1-\frac{1}{t}\right)x'_i\right)_{\forall i\in N}\}$, guarantees that at least one person is better off, with no persons worse off, which contradicts the fact that the original solution is an equilibrium.

QED

Linear moralities are perhaps the simplest and most common type of moralities. The type suggests that each moral issue is judged on its own, and not intermingled with others. Each action is either judged "good" or "bad," without much reference to the context or background. Such a primitive type of morality is easy to interpret and practice, and may seem impersonal; however, who would have thought it would create wonders! In real life, not all moralities appear linear. For example, in stable societies, radicalism in the direction of current morality may be punished. In this case, the monotonicity of moralities is violated, and thus cannot be linear.

The $w_{ij} = w_{kj}\ \forall i, k \in A\ \forall j \in N$ condition is one whereby each gets to be treated the same. It guarantees that if one player can shift the action allocation in a certain direction, any other player, as long as they are not on their moral boundaries, can do so too. This condition often relies on morality being translated through institutions, such as how the modern legal system treats most on an equal basis. However, only by having a democratic legal system wherein each can participate in it or, in math, exert a substantial amount of positive moral weight, does such a moral system internalize externalities. Else, it only internalizes the externalities of the few in power in relation to the rest, excluding those externalities among the disenfranchised population.

More ambitious researchers can construct and discover even more general conditions for the moral mechanism to serve NU-efficiency and true NU-efficiency. By finding such conditions, we can further eliminate the wastes of externalities in daily life, and build a more efficient society based on a scientific understanding of moralities.

III. MoM

MoM is a special type of morality that allows assigning value judgments based on the moral decision of others.

To begin, let us evaluate the situation in which individuals are only allowed to pass value judgments based on their own moral choice. If there is no constraint on MoM, then each player is best off setting their own morality at as high a level of value judgments as possible. With an incremental amount of "adjust to average" *option*, a player has an incentive to set any other moralities at as low a level of value judgments as possible. Such behavior will allow the original simple morality equilibrium action allocation to remain; however, it will also allow for more

equilibriums with slight deviations in terms of moral allocation and, consequently, action allocation from the original equilibrium. This is because each person has more of an incentive to do whatever they choose, creating this self-simultaneous effect, therefore expanding the action allocation equilibrium set centering simple morality-game, a game with all else the same by the moral choice set not allowing for MoMs, unique action allocation. This is how I describe traditions, or self-perpetuating moralities that enable a small deviation from the simple morality equilibrium action allocation. The extent of possible deviation is affected by the moral constraints and weights. The more power the moral mechanism has, the larger the possible deviation. At the expense of NU-efficiency, such deviation brings diversity of moral and action allocation, whereas previously, in simple morality NU games, the action allocation was deterministic. Diversity pertains to the possibility of different equilibrium action allocations under the same NU structure. It can be quite a precious gift, as humans may not have an accurate NU understanding of the externalities of different actions before trying them out. For example, at the emergence of capitalism and the industrial revolution, few realized or envisioned the huge productivity that diversity could unleash; it allows societies to try out different paths.

If, however, there is a linear constraint on MoMs, whereby MoMs must be linear on a real vector, $\theta(m_i) \in \mathbb{R}$, then individuals are best off by setting MoMs in the direction of their morality, with as high a slope as possible. Consequently, each player's morality shifts toward the extreme side slightly, along with their action allocation. This leads to extremism and polarization in terms of both moralities and actions. The original simple morality equilibrium allocation may no longer remain as an equilibrium. The set of equilibrium allocations deviates even further from the original equilibrium allocation, and possibly in both directions. The attributes of linear moralities were described previously.

Now, allowing for MoM based on others' morality yields similar results. Unconstrainted MoMs provide more entrenched traditions, expanding the equilibrium set to a larger extent. Linear MoMs create even larger extremism and polarization. However, there is an additional layer of simultaneous effect between MoMs of different players. I hypothesize that such an effect will lead to conformity between different persons' moralities, thus forming a mainstream ethics or ideology. This hypothesis derives from the other hyposthis that self-enforcing MoMs, or MoMs that give themselves a high value judgment, are more likely to survive and out-compete other MoMs. Thus, groups with high interconnected moral weights or similar NU interests will have identical MoMs. I do not presently have any mathematical proof.

Conformity, here, occurs at the level of MoM. For example, the conformity of freedom of speech is on a level where people all allow freedom of speech and punish those who do not, instead of on the level of what particular speeches are allowed. Conformity has its advantages: it reduces moral conflicts, which can be quite costly, especially when the military and wars over religion or ideology are involved.

Here is an example of a game with linear MoMs:

Let there be N players. Let $x_{ki} \in \mathbb{R}_{\geq 0}$, $k \in K$ denote Good k consumed by Player i. Let $c_{ki} \in [-k^*, +k^*] \subset \mathbb{R}$, $k \in K$ denote the pressure that Player i exerts on others to consume Good k. Let each player, $i \in N$, have the utility function,

$$U_i = \sum_{j \in N} \alpha_{ki} \ln\left(\sum_{j \in N} x_{kj}\right) + \ln\left(\sum_{j \in N} \sum_{k \in K} \left(w_{ij} c_{kj} (x_{ki} + c_{ki})\right)\right)$$

, where $\sum_{k \in K} \alpha_{ki} = 1$, and $\alpha_{ki} \geq 0$ for all $k \in K$. Let each player, $i \in N$, have a budget constraint, $\sum_{k \in K} p_k x_{ki} = I_i$, where $p_k > 0$ and $I > 0$ are constants for all $k \in K$.

Let this be a two-stage game wherein, in the first stage, each i chooses c_{ki} for all $k \in K$, while in the second stage, each player chooses x_{ki} for all $k \in K$.

The NU part of the utility function consists of a Cobb-Douglas function, where α_{ki} signals how much i benefits from having Good k. A strictly private good will have $\alpha_{ki} > 0$ for only one i, and $\alpha_{kj} = 0$ for all $j \neq i$. A strictly public good will have $\alpha_{ki} > 0$ for all $i \in N$. An interest group good will lie somewhere in-between. This is to stimulate the "layers of interest group goods" situation, wherein an interest group good benefits some more than others. An example is how a monarchy benefits the monarchs the most, the feudal lords less, the associated bureaucrats the least, and so on. MoM can be difficult to model, as $\theta_i(m_i) \in \mathbb{R}$ can take many forms. The linear form is one easy possibility.

To see extremism and tradition, take the following simplified game:

Let there be n players. Let all players have a utility function, $U_i = a\sum_{j\in N} x_j + y_i + cV_i$, where $0 < a < 1$, and a cost function, $x_i + y_i \leq 1$, and let all weights be 1. Let the moral choice set be limited, with moralities of the form, $m_j(x) = c_j(x_i + c_i)$, whereby each player determines the coefficient, c. Let the bounds be $b_{low}^2 = -n$ and $b_{up}^2 = n$.

We see that both cases in which all players set their moralities to $m_j(x) = n(x_i + c_i)$ or $m_j(x) = -n(x_i + c_i)$, and correspondingly consume all public or private goods, can be subgame perfect Nash equilibriums. Previously, the public good game always yielded efficiency. Yet now, MoM allows the possibility of both over-consumption and under-consumption of the public good. Previously, if, let us say, a changed, representing a change in externality that the public good, x, brought, the equilibrium allocation shifted correspondingly. Yet now, there is a

likelihood that the equilibrium will not change. Here is the concept of tradition and stickiness, given changes in the NU structure.

IV. A THEORY OF MORAL CHANGE

The model of morality can also be used to study moral change and evolution through history. Instead of constructing a dynamic model, I introduce the tools of **range of moral tolerance** and **range of moral allocations**. A range of moral tolerance refers to all the possible NU (including constraints) and w structures that allow for a certain moral allocation to remain an equilibrium. This generally translates into how long or under what "material" conditions a moral allocation will last. A range of moral allocation refers to all the possible moral allocations that can serve as an equilibrium under a certain NU and w structure. This translates into all the possible moral allocations in a specific historical stage, time, and place.

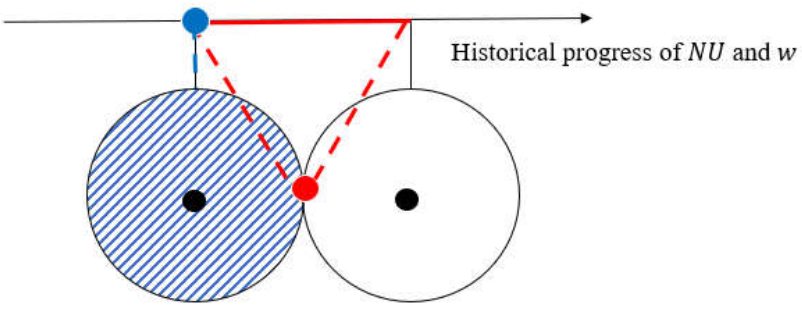

Figure 1: Possible moral allocations in a specific historical stage, time, and place

If one treats the earth as a closed society, then, at each point in time, there exists a certain material externality and moral weight structure largely determined by the natural laws of the universe. These historical conditions are illustrated in Figure 5 as the vector pointing right. Then, at each point in history, there exists a range of moral allocations, painted blue in Figure 5. It may or may not look convex. It can be an interesting endeavor to map out the simple game

equilibrium set of moral allocations, and to devise some sort of radius to measure how much deviance, induced by MoMs, is possible. As NU and w change, potentially caused by changes in technology, productivity, or geography, certain moral allocations cease to exist as possible equilibrium moral allocations. The range of moral tolerance is then the set of time periods wherein a certain moral allocation can serve as an equilibrium allocation. This corresponds to the red line in Figure 5. It does not have to be convex, either.

More on how exactly society chooses among the range of moral allocations remains both a theoretical and an empirical task. Does the old moral allocation remain the realized allocation throughout, until history can contain it no more, and move abruptly to a random new allocation? Or is the change smooth? I tend to favor the former pattern of change. If this is the case, then morality changes in step functions, staying the same for extended periods of time, then undergoing abrupt and violent changes, and repeating the cycle.

V. NATIVISM

This chapter investigates what happens in a MU-moral game in which each player adopts moralities that best enhance their combined value judgments. For this chapter, the "adjust to average" *option* is applied, as it is necessary to make value judgments scarce and the competition for higher status meaningful.

Given fixed actions and information, each player chooses their morality to increase their social status the most. When only simple moralities are allowed, every individual's rational choice is to find a morality that not only gives themselves the highest value judgments, but also gives others value judgments that are as low as possible. Therefore, status-driven moralities must be based on

actions or information that can be easily differentiated, where the distinction is large and clear, while the cost to achieve the actions or information is very low for some and very high for others. Examples of such information are race, gender, symbols of wealth, citizenship, and so on. Counter examples include owning non-luxuries, the shape of one's toes, etc., where either the cost of achieving them is low, or the cost of differentiation is very high.

With MoM, however, each now has the ability to influence the moral choices of others. Since individuals are more affected by the MoMs of those closest to them, they must base their moralities on information that those close to them share in common, but those further away do not have. Depending on the environment, examples can be race, wealth, and citizenship.

When a population can be divided into several relatively isolated subgroups, in terms of moral weights, then a stable status-driven morality arises where each within the subgroups praises information that is unique to most within the subgroup, and devalues those outside of the subgroup. When each group does this, nativism arises as a natural result of the moral mechanism.

Surprisingly, MU-wise, such type of nativism improves MU-efficiency. Imagine a scenario in which there are two planets: one planet is filled with red-eyes, while the other is filled with blue-eyes. If they both discriminate against each other based on their eye-color, because of how each hears the opinions of those closer to them more frequently, each will feel they are above the average, gaining more moral satisfaction compared to when each treats others as equals. In real life, family and nationalism are two great examples of how nativism leads to MU-efficiency. If all couples treat each other as the "most unique and important person in the world," then, because couples have more opportunity to demonstrate such affection, all will feel more loved and valued, compared to when there is no moral differentiation based on family belonging. Why bother to ask for a little recognition from all, when you can have a large stable dose of it from a

closed few? Friendship can also be one such MU-driven morality. Its content comprises treating one's friend with extra favor at the expense of a negligible drop of attention for non-friends.

To see how nativism leads to MU-efficiency, we prove a more general case limited to simple moralities.

Definition If j has a **weight correlated morality**, then $m_j(x_i) \geq m_j(x_k)\ if\ w_{ij} \geq w_{ik}$.

Proposition 6 Assume that it is possible for all players, i, to adopt weight correlated moralities; then, there always exists a MU-efficient allocation whereby all players adopt weight correlated moralities.

Proof Define a welfare function, $W = \sum_{i \in N} MU_i = \sum_{i \in N} \sum_{j \in N} w_{ij} f_i(\frac{JV_{ij}}{\sum_{l \in N} JV_{lj}}) =$

$\sum_{j \in N} \sum_{i \in N} w_{ij} f_i(\frac{JV_{ij}}{\sum_{l \in N} JV_{lj}})$. Make each i maximize W by changing $m_j(\theta_i)\ \forall i \in N$, given w_{ij}. Since weight correlated moralities are possible, and w_{ij} is positively correlated with $m_j(x_i)$, a solution to the maximization problem always exists and will distribute $m_j(x_i) \geq$ $m_j(x_k)\ if\ w_{ij} \geq w_{ik}$. Since the solution maximizes the welfare function, it must be an efficient allocation. Thus, there always exist weight correlated moralities that are MU-efficient.

QED

However, moralities that bring MU-efficiency do not necessarily bring NU-efficiency. The NU externality structure has no deterministic relationship with the moral weight structure. For example, nationalism can be a great political tool to unite popular support. However, such actions can gravely cost NU, as they can damage international relationships and economic efficiency.

VI. DISCUSSIONS

Positive Moral Weights for the Moral Mechanism to Work

We have already seen that the moral mechanism leads to efficiency under certain conditions. However, for the moral mechanism to function, each individual must have a non-zero moral weight toward others in the society. The model of morality assumes that each individual can convey their value judgments to others. If such a channel or medium does not exist, then there is neither a mechanism nor efficiency. Take the prisoner dilemma for example: since each cannot consistently punish or reward others for their actions, their moralities have no effect on each other. On earth, many countries are relatively isolated, while a country's citizens may never hear of or endure the effects of the morality of a person outside their borders. In this sense, some investments in moral channels/infrastructures need to made for the benefits of the mechanism to be applied to a larger society, whether it be better communications, media, unified language, unified government, law, interlinked economy, the military, or so on. Katz's (2000) claim that morality contributed much to the survival of humanity in the early stages suggests that the moral mechanism can be very effective in small, interconnected societies, such as tribes and villages.

Suggestions for Promoting Interest Group Goods

The key to having an upper hand in promoting your own interest group good is to have a larger aggregate moral weight than your opponents. This requires a large or powerful group sharing the benefits, and weak and isolated opponents.

By projecting more moral weight over your own interest group, you can better communicate and defend against opposing moralities; thus, it is necessary to conspire often with your interest

group. By projecting more moral weight over the opponents, you convince them of the justice and righteousness in your morality, and make them believe it, thus contributing to your group interest. By reducing the projecting moral weights of the opposition on your group, your group is less influenced and convinced of their way of thinking. By isolating and dividing your opponents, you suppress their ability to share their experience and the platform on which they convince one another against your group interest. Separate your potential opponents into groups, and have them center on a solid member of your interest group, while maintaining periodic contact with other members of your interest group. If this is impossible, divide your opponent group into equivalent-weight opposing interest groups, for them to spend most of their resources fighting each other instead of your interest group. Note that dividing the opposing interest group may also stifle their ability to promote public good moralities as their moral weights; thus, the aggregate moral weights of society decreases. If possible, try only to limit their weights on the specific interest group matter, and not on other actions. Additionally, individuals are known to betray their interest group; however, as numbers rise, their behaviors return to the statistical mean; therefore, practice caution, and exploit having a few individuals holding immense weight.

A Moral Data Base

A moral data-base can achieve many purposes. By collecting past moralities, one can empirically learn the pattern of how moralities change over time, and thus be able to build a dynamic model of morality, establish how the moral choice set looks, the coefficients linking externalities and moralities, their effects, possible alternatives, inter-period pattern, and so on. By collecting the moralities of groups of individuals now, and applying the model, one can predict future moralities and prepare for them. The interest group that has access to such a data-base and model has the upper hand in promoting their moralities. Society as a whole can benefit by further

exploring and enhancing the performance of the moral mechanism in dealing with social externalities, and eliminating the excess effects of MoM moralities.

Moral Engineering

Having converted moralities from abstract ideas to math, it is now possible to scientifically analyze and engineer them. Here are some guidelines and suggestions that may be useful in this process.

The goal of moral engineering can either be to Pareto improve the current action allocation, or to promote the utility of an interest group. These goals are not contradictory, as efficiency and group interest do not contradict each other. Indeed, those who fail to understand and use morality as a tool will most likely have their group interest taken away by those who do. This provides an additional incentive for whoever is able to invest in moral engineering.

The ideal way to achieve these goals is to build a moral data base through surveying each local population on their moral opinions, extent of moral execution, and their responses to the moral executions of others. Having converted the information into morality functions and NU estimates, one can then calculate the range of moral allocations. Then, whoever is in charge of or funding this moral engineering project can choose among the range the one allocation they prefer, and pay a one-time cost in attempting to shift society to that equilibrium. To lower the risk of deviation, those in charge can strategically pick a moral allocation with a relatively large range of tolerance. If more than one interest group have the moral data base and are adept at moral engineering, then a new type of equilibrium can be defined to include the one-time costs of shifting morality; thus, an equilibrium exists only if each interest group holding this tool finds the cost of shifting the moral allocation higher than their potential gain from shifting.

An alternative method to achieve the above goals is to find as-practical-as-possible conditions for the Nash equilibrium to be Pareto efficient. These conditions will mostly likely fall under either limiting the moral choice set or adjusting moral weights. Limiting the moral choice set can be achieved either by controlling the dissemination of these moral choices (forbidding undesirable moralities to spread), or using MoM against undesirable forms of moralities. Moral weights can be adjusted through building and alternating the mediums and channels through which morality travel, such as expanding the media, changing the political and legal process, mandatory language education, financing certain organizations, increasing military spending, and so on. After these conditions have been discovered and chosen, by engineering the moral choice set and moral weights to fit those conditions, true-efficiency can be achieved. This method, however, does not guarantee the promotion of the interests of a certain interest group.

Moral engineering can also include the creation and discovery of new moralities. Perhaps mathematicians can also participate in the task of designing new moralities and expanding the moral choice set. This requires considerable creativity in abstracting actions or information, such that externalities from those actions align with a significant interest group.

Non-Human Value Judgment Givers

Not all value judgment givers have to be humans. For example, pets, literature, music, games, other cultural products, religions, deities, etc., can also act as value judgment givers. Some cultural products are mere mediums for other humans to convey their moralities. However, some such non-human value judgment givers may violate the assumptions behind a standard moral game. For example, pets' obedience may substitute for social recognition from other humans. Yet, a pet is not a standard player, and has no projecting moral weight. Such substitution only serves to dilute the moral mechanism in internalizing externalities between humans.

Conclusion

This study builds a mathematical model on the moral mechanism, and derives several separate tendencies associated with it. Private interest-driven moralities serve to internalize externalities. MoMs lead to traditions, and expand the equilibrium allocation set. Status-driven moralities lead to nativism, along with MU-efficiency. Inherent moral preference has uncertain effects.

There is enormous scope for perfecting and analyzing this model theoretically, and testing, collecting, and forecasting moralities through it. With a scientific understanding and engineering of morality, we, humans, shall no doubt take over control of another powerful tool.

Acknowledgements

Thanks Maxim Sinitsyn for providing me with initial approval and encouragement. Without his kindness, I could not have carried on with this research. Thanks Ryan Fang for being my advisor. He also connected me with Roger Myerson, Philip Reny, and Benjamin Brooks. Benjamin Brooks suggested me to provide simple examples to facilitate understand. Thanks for Joel Sobel and Mark Machina, whom endured through the pain of reading my earlier drafts when I was an undergraduate student.

Works Cited

Alexander, R. (1987): The Biology of Moral Systems. Aldine de Gruyer, New York.

Andreoni, James. "Privately provided public goods in a large economy: the limits of altruism." *Journal of public Economics* 35.1 (1988): 57-73.

Bagwell, Laurie Simon, and B. Douglas Bernheim. "Veblen effects in a theory of conspicuous consumption." *The American Economic Review* (1996): 349-373.

Becker, Gary S. "A theory of social interactions." *Journal of political economy* 82.6 (1974): 1063-1093.

Becker, Gary S. *The economics of discrimination*. University of Chicago press, 2010.

Becker, Gary S. "The economics of crime." *Cross Sections* Fall (1995): 8-15.

Boyd, Robert, and Peter J. Richerson. "The evolution of indirect reciprocity." *Social networks* 11.3 (1989): 213-236.

Brekke, Kjell Arne, Snorre Kverndokk, and Karine Nyborg. "An economic model of moral motivation." *Journal of public economics* 87.9-10 (2003): 1967-1983.

Chan, Anita, Richard Madsen, and Jonathan Unger. Chen village under Mao and Deng. Univ of California Press, 1992.

Chao, Angela, and Juliet B. Schor. "Empirical tests of status consumption: Evidence from women's cosmetics." *Journal of Economic Psychology* 19.1 (1998): 107-131.

Coase, Ronald H. "The problem of social cost." *Classic papers in natural resource economics*. Palgrave Macmillan, London, 1960. 87-137.

Congleton, Roger D. "The economic role of a work ethic." *Journal of Economic Behavior & Organization* 15.3 (1991): 365-385.

Eastman, Jacqueline K., et al. "The relationship between status consumption and materialism: A cross-cultural comparison of Chinese, Mexican, and American student." *Journal of Marketing Theory and Practice* 5.1 (1997): 52-66.

Fershtman, Chaim, and Yoram Weiss. "Social rewards, externalities and stable preferences." *Journal of Public Economics* 70.1 (1998): 53-73.

Hardy, Sam A., and Gustavo Carlo. "Identity as a source of moral motivation." *Human development* 48.4 (2005): 232-256.

Kallbekken, Steffen, Hege Westskog, and Torben K. Mideksa. "Appeals to social norms as policy instruments to address consumption externalities." *The Journal of Socio-Economics* 39.4 (2010): 447-454.

Katz, Leonard D., ed. *Evolutionary origins of morality: Cross-disciplinary perspectives*. Vol. 1. Imprint Academic, 2000.

Manski, Charles F. "Identification of endogenous social effects: The reflection problem." *The review of economic studies* 60.3 (1993): 531-542.

Mason, R. (1984), "Buyer Behaviour and the Market for Status Goods", *Marketing Intelligence & Planning*, Vol. 2 No. 2, pp. 29-39.

Mason, Roger S. *Conspicuous consumption: A study of exceptional consumer behaviour*. Diss. Salford: University of Salford, 1980.

Mason, R. Ethics and the supply of status goods. *J Bus Ethics* **4,** 457–464 (1985).

Mason, Roger. "The economics of conspicuous consumption." *Books* (1998).

Nowak, Martin A., and Karl Sigmund. "The dynamics of indirect reciprocity." *Journal of theoretical Biology* 194.4 (1998): 561-574.

Nowak, Martin A., and Karl Sigmund. "Evolution of indirect reciprocity by image scoring." *Nature* 393.6685 (1998): 573-577.

Panchanathan, K., Boyd, R. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432,** 499–502 (2004).

Rushton, J. Philippe, Roland D. Chrisjohn, and G. Cynthia Fekken. "The altruistic personality and the self-report altruism scale." *Personality and individual differences* 2.4 (1981): 293-302.

Rushton, J. Philippe. "The altruistic personality." *Development and maintenance of prosocial behavior*. Springer, Boston, MA, 1984. 271-290.

Sigmund, Karl. "Moral assessment in indirect reciprocity." *Journal of theoretical biology* 299 (2012): 25-30.

Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature*, 43 (2): 392-436

Tajfel, Henri, et al. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56 (1979): 65.

Van den Berghe, Pierre L. "Human inbreeding avoidance: Culture in nature." *Behavioral and Brain Sciences* 6.1 (1983): 91-102.

Veblen, Thorstein, (1899), The Theory of the Leisure Class, McMaster University Archive for the History of Economic Thought.