DISCUSSION PAPER SERIES

56690-1701357918

AN ANALYSIS OF MORAL MOTIVES IN ECONOMIC AND SOCIAL DECISIONS

Marc Fleurbaey, Ravi Kanbur and Dennis Snower

PUBLIC ECONOMICS

CEPR

AN ANALYSIS OF MORAL MOTIVES IN ECONOMIC AND SOCIAL DECISIONS

Marc Fleurbaey, Ravi Kanbur and Dennis Snower

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Public Economics

AN ANALYSIS OF MORAL MOTIVES IN ECONOMIC AND SOCIAL DECISIONS

Abstract

We present an analysis of moral motives underlying economic and social decision making. The underlying idea is that moral motives are psychological instruments that induce people to cooperate in pursuit of collective goals and to suppress destructive competition. Different social contexts are associated with different collective action problems, which call for different cooperative relationships. These different cooperative relationships are associated with different moral motives. Cooperative contexts, associated with positive externalities, call for moral motives that reward people for promoting the common good, thereby enabling them to internalize the positive externalities. Competitive contexts, associated with negative externalities, call for moral motives that discourage people from harming one another, thereby promoting the internalization of the negative externalities. In this framework of analysis, people are subject to multiple, context-dependent moral motives. The diverse moral principles underlying the diverse moral motives are to be understood not as mutually exclusive alternatives, but as components to be applied in combination with one another.

Marc Fleurbaey - marc.fleurbaey@psemail.edu
Paris School of Economics, CNRS and ENS-PSL

Ravi Kanbur - sk145@cornell.edu
Cornell University and CEPR

Dennis Snower - dennis.snower@gmail.com
Hertie School, Berlin

An Analysis of Moral Motives in Economic and Social Decisions[*]

Marc Fleurbaey,[†]Ravi Kanbur[‡]and Dennis J. Snower[§]

November 2023

Abstract

We present an analysis of moral motives underlying economic and social decision making. The underlying idea is that moral motives are psychological instruments that induce people to cooperate in pursuit of collective goals and to suppress destructive competition. Different social contexts are associated with different collective action problems, which call for different cooperative relationships. These different cooperative relationships are associated with different moral motives. Cooperative contexts, associated with positive externalities, call for moral motives that reward people for promoting the common good, thereby enabling them to internalize the positive externalities. Competitive contexts, associated with negative externalities, call for moral motives that discourage people from harming one another, thereby promoting the internalization of the negative externalities. In this framework of analysis, people are subject to multiple, context-dependent moral motives. The diverse moral principles underlying the diverse moral motives are to be understood not as mutually exclusive alternatives, but as components to be applied in combination with one another.

JEL: D10, D62, D63, D64, D71.

Keywords: moral norms, moral behavior, public good, social competition, group formation.

1   Introduction

This paper analyzes the role of moral motives in economic and social decisions. Our point of departure is a well-known insight from anthropology, sociology and evolutionary science: one of the main drivers of human evolutionary success has been our capacity to cooperate beyond the bounds of enlightened self-interest.[1] This capacity has been shaped by moral motives throughout human development.[2] Our analysis studies how moral motives shape economic decision-making and can also be applied to social decision-making (social exchanges that do not involve economic transactions).

The inclusion of moral motives extends our analysis of economic decisions beyond individualistic purposes (described by payoffs to the individual) to include collective purposes (described by collective payoffs, including but extending beyond payoffs to the individual) as well. There is a large body of evidence that humans are driven by both sets of purposes,[3] whose relative salience differs across

---

[†]Paris School of Economics, CNRS and ENS-PSL. marc.fleurbaey@psemail.eu

[‡]Cornell University. sk145@cornell.edu

[§]Hertie School. snower@hertie-school.org
[1]For example, Henrich (2016, 2020) and Sober and Wilson (1998), and Wilson (2015).

[2]For example, Tomasello (2016).

[3]See, for example, Shweder and Bourne (1984), distinguishing between societies where individuals are meant primarily to serve the needs of society and ones where society is meant to serve the individual.

individualist versus collectivist cultures[4] and across competitive versus cooperative economic and social settings.[5] We rely on a functionalist explanation of moral motives: Moral motives induce people to cooperate in the pursuit of common goals and to suppress destructive competition.

That economic behaviour is driven by collective, as well as individualistic, goals is not surprising, since homo sapiens and their ancestors have been living in social groups for at least seven million years, and thus may be expected to have developed a variety of psychological[6] and cultural[7] mechanisms for promoting group cohesion, suppressing selfishness, and organizing collective action. Thereby humans have found ways of overcoming the free rider problem, with varying degrees of success. Our analysis views moral motives as psychological devices that induce people to cooperate when self-interest stands in the way of achieving collective goals. These motives provide intrinsic rewards and punishments, working alongside the extrinsic mechanisms[8] in the formation and maintenance of cooperation.

Given that moral motives are recognized as playing an important role in people's private lives, as well as their political allegiances and activities, it would be very surprising to find that they were absent in their economic decisions. After all, economic decisions are associated with many of the same moral problems that are present in the private and political domains. Our analysis indicates that moral motives can make a difference to our economic decisions and thus that there is no Occam's Rasor case for omitting them from economic analysis.

There is a large body of evidence that people are motivated to pursue multiple moral values in practice, with different values becoming salient under different circumstances.[9] The context-specificity of multiple moral values follows from the recognition that different collective goals call for different forms of interpersonal relationships and multiple values drive such multiple relationships. This aspect of morality is particularly obvious with regard to virtues, both religious and secular. These virtues—such as those described in the Judeo-Christian Bible, the Mahabarata, Homer's Illiad, Confucius' Classics, or the Sunna of Muhammad—are exemplars that shape the perceptions, beliefs, emotions and habits required for leading a good life in particular social roles and particular historical and cultural contexts. These virtues cannot be reduced to a single principle, but are rather multiple directions that often hold one another in check (such as Aristotle's Golden mean).[10]

Our analysis aims to capture the multiplicity of context-dependent moral motives by focusing on collective action problems involving positive and negative externalities. Moral motives are portrayed as generating moral incentives, which are psychic benefits and costs for internalizing these externalities. We examine how different economic and social contexts generate different externalities and how the internalization of different externalities calls for different moral motives, resting on different moral principles.

A *motive* is conceived in psychological terms as a force that gives direction and energy to one's behavior, influencing the objective of the behavior.[11] *Moral motives* are understood as motives associated with right or wrong actions or with good or bad outcomes, primarily within a social context. We conceive of moral motives as forces supporting our collective interests. *Moral incentives* are the

---

[4]In individualistic cultures, people's individual goals are salient and one is expected to achieve these goals primarily on one's own; whereas in collectivist cultures, collective goals are salient and people support one another in pursuing these goals (e.g, Hofstede 2001). This divide affects people's self-concept and personality: people in individualistic cultures define themselves in terms of their own preferences and abilities, whereas those in collectivist cultures define themselves as aspects of their social groups (e.g., Triandis 2002). Furthermore, the divide affects behavior patterns in the presence of social dilemmas (e.g., Parks and Vu 1994).

[5]For example, McGuire et al. (2018).

[6]For example, Kurzban, Tooby and Cosmides (2001).

[7]For example, Henrich (2016), Richerson and Boyd (2005), Turchin (2006).

[8]For example, Fehr and Gï¿œchter (2002).

[9]For example, Fiske (1991), Schwartz (1994) and Shweder et al. (1997).

[10]See also MacIntyre (1981).

[11]See Elliot and Covington (2001), following Atkinson (1964).

psychic benefits and costs arising from the moral motives.

To illustrate how different contexts can give rise to different externalities, we focus on two types of context: (i) a *cooperative context* characterized by positive externalities from a public good and (ii) a *positionally competitive context* characterized by negative externalities from status seeking.

In practice, each of these contexts can take a variety of forms. Let us define a *first-order cooperative context* as one that generates a positive externality without strategic effects. Such an externality arises through individual contributions to a public good whose benefit to each individual depends on the sum of the individual contributions. We show that this externality can be internalized through *individual moral incentives*, i.e., incentives for individual decisions, independent of other people's decisions. The individual moral incentives can be interpreted as "self-imposed Pigouvian subsidies," the psychic equivalent to the monetary Pigouvian subsidies in the conventional analysis of externalities.

We define a *second-order cooperative context* as one that generates an externality through a strategic complementarity. For example, one individual's contribution to a public good may raise another individual's productivity in contributing to the public good. We show that this externality can be internalized through *interpersonally reinforcing moral incentives*, specifying a positive feedback between the individuals' contributions.

A positionally competitive context can be analyzed along analogous lines. A *first-order positionally competitive context* creates a negative externality without strategic effects. Such an externality can arise through individual contributions to status competition, whereby an individual's gains depend on the difference between the individualistic payoffs. In this context, the externality can be internalized through *individual moral disincentives* that suppress competitiveness. These may be viewed as "self-imposed Pigouvian taxes."

A *second-order positionally competitive context* operates through strategic substitutabilities, whereby for example one individual's contribution to a collective good may reduce another individual's productivity in contributing to the good. In this case, the socially desirable moral incentives can be achieved through *interpersonally compensating moral incentives*, specifying a negative feedback between the individuals' contributions.

In short, the contributions to the public good are generated by moral motives and associated incentives in the form of psychic benefits—some operating individually and others operating interpersonally. Similarly, contributions to the status contest generate moral motives linked to moral incentives in the form of psychic costs. The nature of these psychic benefits and costs depends on the nature of the externality generated by a particular context.

Our analysis shows how these moral motives may be interpreted in terms of the application of multiple moral values. To illustrate the multiplicity of moral approaches, we initially focus on three separate moral principles: the "principle of merit" (one's moral rewards related to one's merit), the "principle of need" (one's moral rewards related to one's contribution to the satisfaction of others' needs) and the "principle of distributive justice" (one's moral rewards related to some principle of distributive equity). Later in our analysis, we consider two further illustrative principles: the principle of moral reciprocity (the Silver Rule) and the principle of moral redress (redressing the mistakes of others).

Our analysis suggests that different collective contexts give rise to different collective action problems which call for different sets of moral principles. The moral principles relevant in a cooperative context are different from those relevant in a positionally competitive context. Similarly, the principles required for an appropriate response to a first-order cooperative or competitive context are different from those required for an appropriate response to a second-order cooperative or competitive context.

From one perspective, our analysis is normative: it tells us what moral incentives should drive people's decision making in order for socially desirable outcomes to be achieved. From another perspective, it is also descriptive in an cultural evolutionary sense: Assuming that people tend to develop moral incentives that enable them to thrive, their decision making will come to reflect the moral incentives that our analysis derives, at least in a stable environment.

In a stable social, technological and physical environment, socially desirable collective practices may be adopted gradually through time in the process of cultural evolution and moral motives develop in tandem, enabling coordination for the sake of the common good.[12] In such an environment, people also have more opportunity to engage in deliberate moral cognition (using practical reasoning to address moral problems), rather than relying on automatically triggered moral emotions, and this also tends to promote coordination for the common good.[13] Under these circumstances, our analysis may be expected to offer descriptive insights into people's moral behavior.

However, in times of profound social, technological and physical change, it is of course possible that the moral motives underlying people's behavior patterns are no longer appropriate for new collective action problems they face. This is an example of evolutionary mismatch, with important implications for psychology and public policy.[14] Under these conditions, if there is broad agreement on social objectives (summarized in a social welfare function), our analysis offers prescriptive insights, indicating the moral motives that should be active to tackling the collective action problems. But when there is disagreement about moral norms, the population may split into subgroups following the different norms, and it is interesting to study how different types of individuals will be attracted by different moral norms in an equilibrium for group formation. Moral norms implying a greater degree of solidarity may be attractive to individuals who can benefit from it, or on the contrary to individuals who have a greater capacity to fund it. Our results suggest that the former is more likely.

In line with the evidence referenced below, moral motives can be primed and are highly context-specific. The role of politicians, business and civil society in priming motives, as well as their role in creating cooperative contexts that activate moral motives to tackle collective action problems, remains under-investigated in economic analysis.

The rest of the paper is organized as follows. Section 2 reviews related literature. Section 3 presents a general framework for the analysis that follows. Section 4 explores moral motives in a cooperative context and Section 5 does so in a competitive context, assuming away strategic complementarity or substitutability. Section 6 extends the analysis to second-order contexts, in which individual action alters the others' capacities to act. Section 7 applies our model to the moral implications of knowledge-biased technological progress. Section 8 examines group formation when individuals differ in capacity and needs, and groups may adopt different moral norms involving unequal levels of solidarity. Finally Section 9 concludes.

2 Related literature

To place our analysis in context of mainstream economic and social theory, we examine the role of morality in economics and in social relationship theory.

---

[12]For example, Christakis (2019) and Henrich (2016, 2020).

[13]On the distinction between moral reasoning and moral emotions, see for example Greene (2013).

[14]For example, Li, van Vugt and Colarelli (2018).

2.1 Morality in economics

Neoclassical economics, initially, was strictly individualistic, based on the view that economic decisions are made by autonomous individuals pursuing their selfish materialistic interests. Recent developments in behavioral economics have departed considerably from this simplistic and unrealistic approach. In particular, an important literature has studied moral behavior and the interplay between various sorts of motives and incentives. Representative of this literature are the works of Brekke et al. (2003) and Benabou and Tirole (2006). In the former paper, it is assumed that individuals are identical and adopt a common moral norm based on a Kantian reasoning applied to the maximization of social welfare ("what would be best for social welfare if everyone did the same?"). Individuals incur a disutility if their behavior differs from the norm. This paper then studies how public intervention in the form of enhanced productivity in the production of the public good, or subsidies to contributions to the public good, may affect the equilibrium, and finds that crowding out is possible, highlighting the role that an endogenous shift in the moral ideal may have on behavior at the equilibrium.

In the latter paper, individuals are heterogeneous in their motives. Moral behavior is then used by individuals to convey signals about their motives, in a context of imperfect information. Individuals have a mix of three types of motivation: intrinsic (moral), external (the moral action is subsidized), and reputational. The last term involves individuals computing how other people will infer, via Bayesian learning, their intrinsic and external degree of motivation from observing their action. The analysis focuses on how the presence of external incentives may crowd out intrinsic incentives by blurring the significance of observed levels of moral action. Similar analysis of signaling in contributions to public goods can also be found in Corneo (1997) and Seabright (2002). Daugherty and Reinganum (2010) study the trade-off generated by the fact that publicity of contributions may enhance reputation-seeking behavior and induce a deadweight loss, even if it increases the production of the public good.[15] Social image effects have also been studied in dictator games to explain fair behavior (Andreoni and Bernheim 2009).

In a related vein, the interplay between norms, laws and material incentives under various informational conditions has been investigated in particular by Benabou and Tirole (2012) and Fluet and Mungan (2022), in work that sheds light in particular on the power of policies that alter people's beliefs about the distribution of behaviors. Benabou et al. (2018) offer a comprehensive study of various aspects of moral behavior in a model with a simple moral-immoral dichotomy, and in which individuals care about being seen (by others or themselves) as moral. They examine behavior seeking or avoiding moral choice situations, the viral transmission of narratives about the consequences of moral or immoral behavior, the search for information about these consequences, and the mixed use of moral injunctions ("imperatives") and arguments by a principal seeking to influence agents.

The relationship between trust and "moral" behavior has been examined in a related literature. Tadelis (2007) describes "the power of shame," namely, the motivation to be perceived as trustworthy and the rational use of this effect by people granting trust to others. Elligsten and Johannesson (2008) analyse motivations for reciprocal trusting-trustworthy behavior among principals and agents caring about being esteemed by prosocial people. Sliwka (2007) develops a model in which a principal may choose to display trust in agents in order to signal a belief that a high proportion of agents are trustworthy, thereby inducing conformist agents to adopt the trustworthy behavior as well. Herold (2010) studies how a principal signaling trust in the agent by leaving a contract incomplete can enhance the incentives of an intrinsically motivated agent when the success of the project depends jointly on

---

[15]An earlier literature on charitable behavior includes Andreoni (1993, 2006), Glazer and Konrad (1996), Harbaugh (1998), Buraschi and Cornelli (2002), and Prendergast and Stole (2001).

unobserved efforts by the principal and the agent—and therefore on their mutual trust.

In all of this literature, the nature of morality is not questioned and the analysis bears on how actual behavior is influenced by various types of intrinsic or external incentives, such as monetary rewards or reputational benefits, and by various informational contexts and signaling strategies. Our own contribution differs from this literature by examining the content of morality itself rather than the strategic and informational context in which morality is displayed: how can moral motives alter individual behavior in a way that reflects various moral principles and is adapted to relevant contexts of cooperation and competition? We focus on the case in which there are no other external incentives, either of an economic or of a social-reputational sort, that may enhance or crowd out genuine moral motives, in order to decipher the relation between standard canons of morality and the quest for solutions to social problems of cooperation and competition of various types.

2.2   Social relationships

One particularly important way whereby people coordinate their behavior to address externalities and inequities in practice is to engage in cooperative social relationships (such as conversations, contests, group activities). The public good in our analysis can be interpreted in terms of such relationships. After all, cooperative relationships involve positively valenced psychic exchanges among individuals that often do not receive monetary compensation. They are, in other words, a type of public good. These cooperative relationships may also mitigate the negative externalities arising from status competition in our analysis.

In neoclassical economics, externalities and inequities are defined with reference to behavior patterns that would arise from purely self-interested behavior. Social relationships affect people's preferences and when these relationships are cooperative, they enable people to internalize some externalities and overcome some inequities. Social relationships may also mitigate negative externalities arising from status competition in our analysis (as shown below). Conversely, adversarial social relationships can create negative externalities and inequities.

Morality may be understood functionally as having two mutually reinforcing roles with regard to social relationships. First, moral motives promote cooperative social relationships, with different motives promoting different relationships. Second, these relationships, in turn, promote the salience of their associated moral motives (since these motives are dependent on the social context).

This connection between moralities and social relationships helps explain the strong emotional force that moral principles often have. People's survival commonly depends on their ability to cooperate with one another; social relationships enable them to do so; and moral motives are the drivers and regulators of these relationships. The connection between moralities and social relationships also elucidates the reasons for moral conflict that may occur across cultural lines. Different cultures often face different collective action problems, calling for different moral motives, supported by different moralities and social relationships. An example is the contrast between the "culture of honor" in the American South and the liberal culture in the American North. According to Nisbet and Cohen (2004), the former was a response to the coordination problems of herders in the Southern states, while the latter was adapted to the coordination problems of growers in the Northern states. When these cultures meet, they clash. Since their distinctive moral motives arose from a process of cultural evolution in response to collective action in different social contexts, Southerners and Northerners need not be aware of the functions their moralities serve. Instead, they view themselves as adhering to moral principles that are intrinsically worthwhile.

A clash of moralities can also arise when different cultures choose to solve the same collective action

problem through different moral motives, which generate different social relationships. Fiske (1991) provides an illustrative example of a small town considering how to organize firefighting. One option is to discuss the issue at meetings of the inhabitants, until a community consensus emerges. A second possibility is to let the leader of the community decide by fiat. A third option is to decide the issue by referendum. A final possibility is to let the market decide by letting people purchase fire insurance. Each of these coordination mechanisms is associated with distinctive moral motives.

To understand the connection between these coordination mechanisms, their associated social relationships and their underlying moral motives, it is useful to classify cooperative relationships in terms of the social-relational approach to moral psychology,[16] which posits four basic types of social relationship: communal sharing, authority ranking, equality matching and market pricing.

In "communal sharing," people perceive themselves to be undifferentiated components of a social entity, as may occur in families, friendship groups, military," units, teams, nations, ethnicities and religions. In small-scale social relationships where group members have an intimate understanding of each other's wellbeing, communal sharing can be driven by the morality of care. In large-scale communities where this intimate understanding is lacking, people provide aid to one another in accordance with the morality of need. Both of these moralities serve to support the integrity of in-groups through a sense of collective responsibility.

"Authority ranking" is a hierarchical relationship, in which superiors have a sense of pastoral responsibility toward their subordinates, providing aid in accordance with their abilities, and the subordinates are obligated to respect their superiors. This coordination mechanism is appropriate when there are significant asymmetries of ability, information and power. Along the lines above, small-scale communities can motivate this social relationships through the morality of care, while large-scale communities can do so through the morality of merit.

"Equality matching" aims at achieving an even balance and in-kind reciprocity in social relationships. It is governed by the morality of equality-based forms of distributive justice.[17] Finally, under "market pricing," people are compensated in proportion to the benefits and costs they bestow on others. When this compensation is monetary, the social coordination is performed by the market, so that self-interest is sufficient,[18] though support for choosing this coordination mechanism may have to arise from the morality of merit.

Each of these social relationships entails its own obligations and transgressions. The psychological function of morality in this context is to facilitate the creation and maintenance of these cooperative relationships. Different social relationships entail different moralities, which in turn support their distinctive social relationships. When one culture uses its own moral standards to judge the morality of another culture, the result may be bitter conflict. Recognizing the sources of divergent moral motives may help the parties understand the context-dependence of their moral disagreements and thereby come a step closer to mutual tolerance and respect.

In this context, it is clear how moral motives promote the internalization of positive externalities associated with relationship-driven public goods. The moralities of need and merit, as well as the morality of care, all support preferences that reward contributions to such public goods. They also serve to reduce the negative externalities associated with positional competition through psychic costs of anti-social behavior. Furthermore, the morality of distributive justice serves to bring individual preferences into line with social concerns regarding inequities. All these moralities limit the capacity

---

[16]In particular, the relational models theory of Fiske (1991, 1992), Rai and Fiske (2011), Fiske and Haslam (2005), Haslam (1994, 2004) and Haslam and Fiske (1992).

[17]For example, Deutsch (1975) and Folger et al. (1995).

[18]In this case, the conventional neoclassical rationale for the Invisible Hand is applicable.

7

for self-interest to subvert collective goals.

Our analysis below also illustrates how disagreements concerning distributive justice lead to reduced provision of the public good, represented in terms of cooperative relationships.

3 Framework of analysis

There are n individuals indexed $i = 1, 2, 3, \ldots n$. Let the action of individual i be denoted y_i. In order to keep the analysis simple, y_i is assumed to be one-dimensional (a real number). Let the actions of individuals other than i be denoted y_{-i}. Express the self-interested well-being of individual i, dependent on actions of all individuals, as $V_i(y_i, y_{-i})$, which is concave in y_i and increasing in y_{-i} for the case of a positive externality of an individual's action on the wellbeing of others, but decreasing in y_{-i} for the case of negative externality.

When pursuing narrow self-interest, each individual i maximizes $V_i(y_i, y_{-i})$ with respect to y_i. In the case of an interior solution, the following first-order conditions are satisfied:

$$V_{iy_i}(y_i, y_{-i}) = 0, \ \ i = 1, 2, ..., n \tag{1}$$

where $V_{iy_i}(y_i, y_{-i})$ denotes the partial derivative of $V_i(y_i, y_{-i})$ with respect to y_i. Equations (1) solve for the Nash Equilibrium outcomes in actions.

Let group well-being, the social welfare function, be given by a weighted sum of individual well-beings:

$$\sum_j \alpha_j V_j(y_j, y_{-j}). \tag{2}$$

There can be many different moral theories underlying this social welfare function. Our analysis, however, does not cover moral theories that would recommend Pareto-suboptimal allocations.

Maximizing group wellbeing with respect to each y_i requires actions defined by the following first order conditions:

$$V_{iy_i}(y_i, y_{-i}) + \frac{1}{\alpha_i} \sum_{j \neq i} \alpha_j V_{jy_i}(y_i, y_{-i}) = 0, \ \ i = 1, 2, ..., n \tag{3}$$

The difference between the sets of first order conditions in (1) and (3) captures the misalignment of private incentives in the advancement of group wellbeing. Now consider the objective of a "moral individual" by adding a moral incentive to the individual's self-interested welfare such that the first order conditions for maximizing the augmented individual objectives are altered as follows:

$$V_{iy_i}(y_i, y_{-i}) + r_i = 0, \ \ i = 1, 2, ..., n \tag{4}$$

The adjustment factor r_i for individual i, which in general could depend on all the actions and on the parameters of the problem, will change individual behavior from narrow advancement of self-interest (given by (1)) towards the advancement of group wellbeing (given by (2)). We will discuss alternative interpretations and motivations of the adjustments r_i below, but formally the necessary conditions for the first order conditions in (4) to coincide with the first order conditions for group maximization in (3) are:

$$r_i^* = \frac{1}{\alpha_i} \sum_{j \neq i} \alpha_j V_{jy_i}(y_i, y_{-i}), \ i = 1, 2, ..., n \tag{5}$$

Under this condition, individual welfare maximization is consistent with maximization of group well-being. We will explore the implications of (5) in detail in this paper.

The form of (4) suggests thinking of r_i as a self-imposed Pigouvian tax or subsidy. It is not of course a monetary cost or compensation for the action, from the state or from anyone else. It is a "self-imposed" psychic cost or benefit arising from the individual's moral motives. It is essential to understand that the role of r_i is first and foremost to impose a *marginal* shift to the first-order condition. Whether it also affects i's *level* of well-being in any specified way depends on the type of moral motives at play in the considered context and on the prevailing norms.

A few typical examples can illustrate this important point. First, consider the case in which i's overall well-being level is represented by the function $V_i(y_i, y_{-i}) + \gamma_i(y_i - \hat{y}_i)$, where the additional term represents the psychic reward or penalty coming from a deviation between the action y_i and a benchmark $\hat{y}_i$. In this case, when $\hat{y}_i$ is equal to the socially optimal level for y_i and γ_i is the optimal marginal parameter r_i^*, i will incur no additional utility or disutility from "simply doing the right thing". In contrast, if $\hat{y}_i$ is suboptimal, i will reap a psychic benefit from behaving morally, whereas if $\hat{y}_i$ is above the optimum, a negative psychic effect will remain from a form of guilt for not doing enough. The important point is that, in this particular configuration, $\hat{y}_i$ has no impact on behavior and only affects the level of reward or penalty felt by the agent from moral considerations. In other words, with this type of overall wellbeing function, norms of behavior have no behavioral impact but only impose a pure psychic benefit or cost on moral individuals.

Second, consider the different case in which overall well-being is afflicted by a penalty for deviating from the norm in any direction, as with the function[19] $V_i(y_i, y_{-i}) - \gamma_i(y_i - \hat{y}_i)^2$. In this case, the moral shift term in (4) is equal to

$$r_i = \gamma_i(\hat{y}_i - y_i).$$

In this formula it is clear that the shift term r_i is increasing in the norm $\hat{y}_i$, other things equal. The overall impact on action y_i is less obvious, as y_i now appears twice in the first-order condition. Differentiating (4) yields:

$$\frac{\partial y_i}{\partial \hat{y}_i} = \frac{\gamma_i}{\gamma_i - V_{iy_iy_i}(y_i, y_{-i})}$$

which is positive but lower than 1 when $\gamma_i > 0$ and $V_i(y_i, y_{-i})$ is concave in y_i. So, a greater norm leads to greater contributions in normal circumstances. But the interesting feature of this case is that $r_i > 0$ only when $\hat{y}_i - y_i > 0$, i.e., when the agent fails to do the right thing. This is natural, as the term $\gamma_i(y_i - \hat{y}_i)^2$ is incorporated by the agent in a trade-off with self-interested wellbeing. This implies that with this type of overall wellbeing function, the norm of behavior needs to be above the social optimum in order to produce the socially optimal situation.

It is straightforward to modify this wellbeing function in order to transform the cost of guilt into a pride of virtue, simply by introducing a benchmark term as in the previous example: $V_i(y_i, y_{-i}) + \gamma_i\left(\zeta_i - (y_i - \hat{y}_i)^2\right)$.

These examples show that certain questions about wellbeing effects of morality are hard to answer in absence of precise insights about whether morality enters positively or negatively through various feelings and psychic mechanisms. For instance, one can show that, with the last function, the influence

---

[19]A similar formulation is found in Brekke et al. (2003).

of γ_i on wellbeing is negative if ζ_i is small and positive if it is large enough.[20] In light of this observation, we focus our analysis of wellbeing impacts of morality on the self-centered term $V_i(y_i, y_{-i})$ in the remainder of this paper, with an exception in section 8. To preclude any misunderstanding, let it be emphasized here that our model does involve a positive *marginal* effect on well-being of conforming better to the moral norm (this is essential for morality to incentivize moral behavior), but we prefer to take no stance on the question of whether morality as such also has a *level* effect on well-being, since it could go either way, triggering shame or pride overall.

This analytical framework is useful to set the scene for our analysis and we will return to it later. But we now proceed to various instructive special cases that allow for tractable solutions and more detailed investigation of moral incentives. Specifically, we are interested in how constellations of moral incentives are composed of constituent parts that appeal to different strands of moral reasoning.

4   Moral motives in a first-order cooperative context

In this section we consider two individuals, indexed 1 and 2, who interact in a first-order cooperative context, where their contributions to a public good generate a positive externality without strategic effects.[21]

4.1   The setup

Individual i makes effort Y_i as a contribution to a public good. The impact on the public good is mediated by a productivity coefficient A_i so that the effective contribution of each individual to the public good is $A_i Y_i$. The overall public good is given by

$$G = A_1 Y_1 + A_2 Y_2. \tag{6}$$

The value of G to individual i is $B_i G$. The cost of effort Y_i to individual i is $\frac{1}{2} Y_i^2$. So the net benefit to individual i, in terms of narrowly defined self-interest, is given by

$$V_i = B_i (A_1 Y_1 + A_2 Y_2) - \frac{1}{2} Y_i^2. \tag{7}$$

If the two individuals maximize their self-interested wellbeing, their efforts are given by

$$Y_1^P = B_1 A_1 \tag{8}$$

---

[20] Note that a greater γ_i increases the weight of guilt for a fixed behavior, but also induces a reduction of the gap to the norm $\hat{y}_i$. The impact is computed as

$$\left[V_{iy_i}(y_i, y_{-i}) + 2\gamma_i (\hat{y}_i - y_i) \right] \frac{\partial y_i}{\partial \gamma_i} - (y_i - \hat{y}_i)^2 + \zeta_i.$$

One computes

$$\frac{\partial y_i}{\partial \gamma_i} = \frac{\hat{y}_i - y_i}{\gamma_i - V_{iy_i y_i}(y_i, y_{-i})}.$$

Plugging this into the welfare impact, and taking account of (4), one obtains that the change in overall wellbeing is equal to

$$(y_i - \hat{y}_i)^2 \left[\frac{V_{iy_i y_i}(y_i, y_{-i})}{\gamma_i - V_{iy_i y_i}(y_i, y_{-i})} \right] + \zeta_i,$$

and the first term is negative, meaning that the direct disutility of a greater feeling of guilt outweighs the indirect reduction of guilt due to more moral behavior, and this is related to the fact that the adjustment of behavior to the norm $\partial y_i / \partial \hat{y}_i$ is lower than 1, because the negative term in brackets, in the above expression, is equal to $\partial y_i / \partial \hat{y}_i - 1$. With a large enough ζ_i, this negative term is counterbalanced.

[21] A second-order cooperative context, where their public good contributions generate a strategic complementary, will be considered later.

$$Y_2^P = B_2 A_2 \tag{9}$$

where the superscript P indicates private optimization. Note that each individual's self-interested effort level depends only on their productivity and preference, with no reference to the other individuals.

Now let group wellbeing be defined by a weighted sum of the individual net benefits:

$$V_1 + \alpha V_2$$

where α is the weight of individual 2's wellbeing relative to that of individual 1.

We begin with the assumption that there is agreement on this relative weight. (The issue of disagreement and its implications is taken up later in this section.) Then the wellbeing of the social group is maximized with respect to the efforts of both individuals when

$$
\begin{align}
Y_1^* &= A_1 \left(B_1 + \alpha B_2 \right) \tag{10}\\
Y_2^* &= A_2 \left(B_2 + \frac{1}{\alpha} B_1 \right) \tag{11}
\end{align}
$$

We refer to these as socially optimal contributions to the public good. Each individual's contribution depends on both individuals' preferences and on their weights in the group. It should be clear that, comparing (8) with (10), and (9) with (11), group optimal contributions are greater than self-interested contributions, because they take into account the positive externality of one individual's contribution on the wellbeing of the other through the public good.

4.2 Moral incentives

Finally, suppose that the individuals are motivated both by self-interest and moral values, which induce them to take the public interest into account. Let us represent the moral incentives of individuals 1 and 2 through marginal terms R_1 and R_2, respectively, so that the individuals behave as if they maximized:

$$
\begin{align}
M_1 &= V_1 + R_1 Y_1 \tag{12}\\
M_2 &= V_2 + R_2 Y_2 \tag{13}
\end{align}
$$

The moral incentives R_1 and R_2 can be interpreted as the perceived marginal reward from contributing effort towards the public good, or "doing the right thing". In the context of a model containing only linear relations—e.g., the benefit from public goods is linearly related to the contributions, social welfare is linearly related to individual utilities, and so on—R_1 and R_2 can be interpreted as "self-imposed Pigouvian subsidies" on effort towards producing the public good. These give rise to the morality-adjusted solutions:

$$
\begin{align}
Y_1^M &= B_1 A_1 + R_1 \tag{14}\\
Y_2^M &= B_2 A_2 + R_2 \tag{15}
\end{align}
$$

Compared to (12) and (13), each individual's effort is adjusted by the "moral incentives" R_1 and R_2.

Starting from zero, increases in the moral weights R_1 and R_2 given to public-good contributions will naturally increase individual effort from the pure self-interested level. In order for these effort

levels to coincide with the group optimal levels (10) and (11), the moral incentives will have to be set at:

$$R_1^* = \alpha B_2 A_1 \tag{16}$$

$$R_2^* = \frac{1}{\alpha} B_1 A_2 \tag{17}$$

In effect, if each individual applies a self-imposed Pigouvian subsidy on effort at the rates given above, the effort levels will be as in (10) and (11), the socially optimal outcome.

We proceed to interpret these moral incentives in terms of two alternative approaches: justice and care. The justice approach is commonly articulated deontologically, in terms of principles of justice. They are the outcome of rational deliberation, applied to "thin" social relations, appropriate for social groups in which people have limited personal experience of one another and thus limited knowledge of the wellbeing of other group members. Under these circumstances, they may achieve their common purpose through an appropriate application of moral principles that have normative force independently of the people to whom they are applied. The associated moral incentives can be conceived as arising through a process of cultural evolution of a population with stable traits in a stable environment.

By contrast, for social groups in which members have "thick" social relations—from wide-ranging personal relationships[22] (inducing them to engage in effective perspective-taking and profound empathetic concern) to cultural affinities (enabling more modest perspective-taking and concern)—a different normative approach may be more effective, namely, one that rests on one's direct participation in the wellbeing of others. These thick social relations are associated with a "thick," particularist caring morality, which is based on different evaluative lines from the "thin," universalist morality that we apply to strangers in other cultures.[23]

In short, the justice approach is particularly appropriate to civil societies, whereas the caring approach pertains primarily to communities.

4.3 The justice approach

Interpreting the moral incentives through principles of justice, we will focus on three principles, which we define in restrictive terms: the principles of merit, need and distributive justice.

According to the *"principle of merit*," everyone should be rewarded in accordance with one's merit.[24] In general, the rewards could be external (such as monetary remuneration) or internal (such as the "warm glow" of philanthropy) and merit can encompass a wide variety of forms (talents, efforts and accomplishments). However, the conventional economic interpretation of this principle is that one's productivity (a particular manifestation of merit) should receive a corresponding monetary reward (directly through remuneration or indirectly through signals that lead to remuneration).[25]

Our analysis points to a different facet of the idea of merit, namely, that people owe to society in proportion to their productivity. This perspective on the principle of merit is consonant with the principle of "noblesse oblige" and the moral precept "from each according to his ability." It is also in

---

[22]For example, Selman (2008).

[23]These notions of thick and thin morality are taken from Walzer (1994).

[24]See, for example, Sandel (2020).

[25]External rewards arising for moral reasons—employers compensating employees of high ability on moral grounds rather than profitability, universities offering places to the most able students, welfare policy measures designed to promote equality of opportunity—play potentially important roles in the functioning of economies. An analysis of these rewards lies beyond the scope of this paper.

the spirit of Luke 12:48: "From everyone who has been given much, much will be demanded; and from the one who has Ben entrusted with much, much more will be asked." All religious faiths have some such principle. Higher productivity generates higher income, which obliges one to contribute more to the common good. When one does so, one should receive an internal (psychic) reward arising from fulfilling one's moral obligation. In this sense, the principle of merit becomes a statement about the satisfaction one is justified in feeling from enabling society to benefit from one's abilities.

In cooperative settings, the psychic benefits can arise from the intrinsic satisfaction of contributing to the common good in accordance with one's ability and from social prestige.[26] Alternatively, there may be psychic costs of shame and guilt[27] from failing to contribute in accordance with one's merit. In competitive settings, the psychic benefits may rise from feelings of "noblesse oblige" or pride that one has not exploited one's advantage over others even though it was in one's power to do so.

Psychologically, the principle of merit is associated with the "achievement motive." People thus motivated show preferences for challenges or competitive situations[28] and tend to engage in persistent, ambitious, or dominant actions[29]. When achievement-motivated individuals receive positive feedback on their productivity in contributing to a public good, they tend to solve more tasks in a given time than individuals who lack this motive[30].

The *"principle of need"* prescribes that everyone should be rewarded in accordance with their need.[31] It provides the moral basis for means-tested unemployment and pension benefits, whereas contribution-based unemployment and pension benefits rely on the principle of merit. Our analysis focuses on the morally motivated psychic benefits from contributing to the common good or suppressing competitiveness in accordance with the needs of others.

Psychologically, the principle of need is associated with the compassion motive.[32] People primed with the compassion motive with the name of a secure attachment figure are more willing to help people in distress.[33]

The *"principle of moral reciprocity"* requires that we repay others in accordance with what they have done to us. Under pure reciprocity, the repayment is equivalent; under "moral reciprocity," by contrast, the degree of repayment depends on morally relevant characteristics of the individuals involved. (For example, needier people may be required to repay a benefit in accordance with their means.) The principle of moral reciprocity has been called the Silver Rule,[34] to distinguish it from the Golden Rule (which prescribes benevolence, rather than reciprocity). An example of this rule is generous tit-for-tat.[35] In many religious traditions, the requisite degree of generosity is related to one's merit.

Psychologically, the principle of moral reciprocity—or fairness—is associated with several motives, particularly the affiliation motive to reward fairness, the anger (threat approach) motive to punish fairness violations and the fear (threat avoidance) motive to avoid such punishment. Individuals driven by the affiliation motive experience the need to belong, to be accepted, and the urge for relatedness.[36] When these needs are not met, they experience self-reported separation distress.[37] The affiliation

---

[26]See, for example, Henrich and Gil-White (2001).

[27]For example, Tangney and Dearling (2002).

[28]Winter (1973) and Schultheiss & Brunstein (2010).

[29]Winter (1973); (Pang, 2010); for a review see Hall, Stanton & Schultheiss (2010).

[30]Lowell (1952); French (1956)

[31]For example, Gilligan (1982).

[32]This is to be distinguished from the affiliation motive, dealing with the need to belong to a social group (Weinberger, Cotler, & Fishman, 2010) and Reiss (2004).

[33]Mikulincer, Shaver, Gillath & Nitzberg (2005).

[34]For example, Hirshleifer (1985).

[35]For example, Wedekind and Milinski (1996).

[36]Atkinson et al. (1954); Boyatzis (1973); Baumeister & Leary (1995).

[37]Eisenberger, Lieberman & Williams (2003).

motive induces individuals to initiate social interactions and maintain the cohesion of social groups.[38] It is also responsible for anxiety-driven efforts to meet affiliative needs.[39]

Fairness violations are commonly perceived as social threats by third parties, who commonly respond with anger.[40] The anger motive involves aggressive behavior and increased proclivity towards risky decisions[41] that are often destructive, intimidating or antisocial.[42] The fear motive induces individuals to flee, to freeze, or to appease.[43] Emotionally, fear-motivated people experience anxiety or panic.[44]

The *"principle of distributive justice"* creates the obligation to treat people's needs in accordance with some principles of distributive equity.[45] Our analysis focuses on a particular aspect of distributive equity: interpersonal comparisons of wellbeing with regard to the moral drivers of economic decisions. For example, under Benthamite utilitarianism, the material wellbeing of the poor should receive more weight than the material wellbeing of the rich (in line with the declining marginal utility of wealth); under Rawls' difference principle, inequalities in material wellbeing should be designed to benefit the least advantaged members of society; and so on. In this sense, the principle of distributive justice may be considered separately from the principles of merit, need or moral reciprocity, since the former is concerned with interpersonal comparisons of utility, whereas the latter are concerned with rewards for particular activities (contributions to public goods and suppression of status seeking).

Psychologically, the principle of distributive justice is associated with the same motives as the principle of moral reciprocity, but these motives are applied toward different ends (distributive justice rather than moral reciprocity). The combination of the anger and fear motives helps promote the principle of distributive justice by eliciting punishments (through the anger motive) and simultaneously generating the sensitivity to these punishments (through the fear motive).

In this context, our analysis makes two significant claims: (1) To coordinate economic decisions in the public interest, the various principles of justice need to be combined in well-defined ways, not viewed as mutually exclusive alternatives. (2) The socially desirable combination of justice principles is context-specific, i.e. it depends on the nature of the collective action problem to be addressed.

The justice principles underlying the moral incentives described by equations (16) and (17) may be summarized as follows:

- The greater is an individual's productivity (A_i), the greater is the individual's moral incentive (R_i^*). Thus people of greater ability should receive a greater moral reward for their contribution to the public good and, correspondingly, feel a greater obligation to contribute to the common good. This is the "morality of merit."

- The greater is an individual's need (B_i), the greater is the other individual's moral incentive (R_j^*, for $i \neq j$). Thus a person should recognize a greater sense of social responsibility, the greater are the needs of others in his or her social group. The greater the needs of others, the greater should be the moral reward for satisfying these needs by contributing to the common good. This is the "morality of need."

---

[38] Atkinson et al. (1954); Lansing & Heyns (1959); Exline (1962); Walker & Heyns (1962); McAdams & Constantian (1983); McClelland (1985); Weinberger, Cotler, & Fishman (2010).

[39] Weinberger, Cotler, & Fishman (2010).

[40] Fessler (2010).

[41] Kornadt, (1984); Fessler, Pillsworth, & Flamson (2004); Leith & Baumeister (1996); for a review see Carver & Harmon-Jones (2009).

[42] Tedeschi, Smith & Brown (1974); Small & Lerner (2008); Eimontaite, Nicolle, Schindler, & Goel (2013).

[43] For reviews see Adolphs (2013); LeDoux, (1998).

[44] Avram et al., (2010); Schaefer et al. (2010); Stemmler et al. (2001).

[45] Summarized, for example, in Fleurbaey (2004).

- The greater is the relative weight of an individual's wellbeing in social wellbeing (such as α for individual 2), the lower is that individual's moral incentive (R_2^*) and the greater is the other individual's moral incentive (R_1^*). Thus the importance of an individual's wellbeing in society justifies lower contributions to the common good. This is the "morality of distributive justice." The weights of individuals' wellbeing in social wellbeing can be underpinned by a variety of moral values. For example, Benthamite utilitarianism under declining marginal utility of wealth implies that poorer individuals should be given greater weight in the social utility function.

In sum: In order to achieve socially optimal contributions to the provision of a standard public good (for which one's benefit depends on the sum of the individuals' contributions), an individual's moral rewards must be proportionate to (1) the individual's merit, (2) the needs of others and (3) the social importance of others.

These results are striking, since they stand in contrast to much conventional moral reasoning. For instance, it is common to depict the morality of merit (reward people in accordance with their abilities) as opposed to the principle of need (reward people in accordance with their needs). Moreover, the principle of distributive justice is commonly viewed as quite separate from the principles of need or merit. The results above, however, suggest that these various moral principles should be combined in the appropriate proportions in order for socially optimal outcomes to be achieved.

The implications are profound and far-ranging. For example, in *The Tyranny of Merit,* Sandel argues that giving people rewards, through status and income, in accordance with their abilities leads to a neglect of the public good, since it induces the winners to regard their winnings as a just reward for their merit and the losers to blame themselves for their lack of success. Our analysis generates no principle whereby higher abilities should receive monetary rewards flowing exclusively to the individual.[46] On the contrary, higher abilities should receive psychic rewards only insofar as they lead to higher contributions to the public good. Furthermore, receiving such psychic rewards solely on the basis of ability (i.e., $R_i^* = A_i$) does not yield the social optimum. Rather, the psychic rewards from contributing to the public good in proportion to one's abilities should be supplemented by rewards for the satisfaction of needs and for deference to social importance.

4.4 Moral disagreement and the limits of social concern for others

Now suppose that people disagree with regard to their relative importance in social welfare. One common form of such disagreement arises from the influence of self-interest.[47] Insofar as each individual accords herself greater importance in her subjective social welfare function than other individuals do so in their subjective social welfare functions, this moral disagreement has important implications for the size of a social group contributing to a common public good.

Let α_1 and α_2 represent the beliefs of individuals 1 and 2, respectively, concerning the weight of individual 2 relative to individual 1 in social welfare, and let both individuals be driven by moral concerns, as depicted by their moral incentives. It is worth noting at the outset that since the self-interested contributions (equations (8) and (9)) are less than the socially optimal contributions ((10) and (11)), some moral incentive (i.e. some positive R_i) is better for society than no moral incentive,

---

[46]Other authors have shown that recognition of prior possession mitigates conflicts over resources (e.g. Gintis 2007, Hare, Reeve and Blossey 2016, Maynard Smith 1982). But with regard to our analysis, this insight begs the question concerning why one's productivity should be considered one's prior possession. After all, an individual's productivity invariably depends in large part on the knowledge gained from others, the experience shared with others, as well as more direct support within teams.

[47]Another form of common moral disagreement - not analyzed in this paper - arises from adherence to conflicting moral principles.

regardless of the degree of moral disagreement. In other words, starting from $R_i = 0$, a small increase in R_i will raise social welfare for any α_i.

Since the behaviors of individuals 1 and 2 are driven by their beliefs α_1 and α_2, respectively, their moral incentives become, by analogy with (16)-(17):

$$
\begin{align}
R_1^\beta &= \alpha_1 B_2 A_1 \tag{18}\\
R_2^\beta &= \frac{1}{\alpha_2} B_1 A_2 \tag{19}
\end{align}
$$

where the superscript β denotes that the moral incentives are belief-driven.

The associated contributions to the public good are:

$$
\begin{align}
Y_1^\beta &= A_1 \left(B_1 + \alpha_1 B_2 \right) \tag{20}\\
Y_2^\beta &= A_2 \left(B_2 + \frac{1}{\alpha_2} B_1 \right) \tag{21}
\end{align}
$$

The resulting provision of the public good is

$$
G^\beta = A_1^2 \left(B_1 + \alpha_1 B_2 \right) + A_2^2 \left(B_2 + \frac{1}{\alpha_2} B_1 \right) \tag{22}
$$

The associated self-centered level of wellbeing for individuals i $(i = 1, 2)$:

$$
V_i^\beta = B_i G^\beta - \frac{1}{2} \left(Y_i^\beta \right)^2 \tag{23}
$$

We assume that both individuals are self-interested, $\alpha_1 < 1$ and $\alpha_2 > 1$. As the individuals become more self-interested, α_1 falls and α_2 rises. The fall in the subjective weight α_1 leads to a fall in individual 1's contribution to the public good, as $\frac{\partial Y_1^\beta}{\partial \alpha_1} = A_1 B_2 > 0$. This boosts individual 1's wellbeing and the cost of a reduced public good is not sufficient to outweigh this gain, as one computes:

$$
\begin{align}
\frac{\partial V_1^{\beta*}}{\partial \alpha_1} &= B_1 \frac{\partial G^\beta}{\partial \alpha_1} - Y_1^\beta \frac{\partial Y_1^\beta}{\partial \alpha_1}\\
&= B_1 A_1^2 B_2 - A_1^2 \left(B_1 + \alpha_1 B_2 \right) B_2\\
&= -\alpha_1 A_1^2 B_2^2 < 0
\end{align}
$$

A symmetrical result is obtained for individual 2. However, when $\eta = \alpha_1 = 1/\alpha_2$ and this common η parameter moves for both individuals simultaneously, one finds that the impact on self-centered wellbeing may have a different sign:

$$
\begin{align}
\frac{\partial V_1^{\beta*}}{\partial \eta} &= B_1 \frac{\partial G^\beta}{\partial \eta} - Y_1^\beta \frac{\partial Y_1^\beta}{\partial \eta}\\
&= B_1 \left(A_1^2 B_2 + A_2^2 B_1 \right) - A_1^2 \left(B_1 + \alpha_1 B_2 \right) B_2\\
&= A_2^2 B_1^2 - \eta A_1^2 B_2^2,
\end{align}
$$

which is positive when

$$
\eta < \left(\frac{A_2/B_2}{A_1/B_1} \right)^2.
$$

16

Likewise, for individual 2, one computes

$$
\begin{aligned}
\frac{\partial V_2^{\beta *}}{\partial \eta} &= B_2 \frac{\partial G^\beta}{\partial \eta} - Y_2^\beta \frac{\partial Y_2^\beta}{\partial \eta} \\
&= B_2 \left(A_1^2 B_2 + A_2^2 B_1 \right) - A_2^2 B_1 \left(B_2 + \eta B_1 \right) \\
&= A_1^2 B_2^2 - \eta A_2^2 B_1^2,
\end{aligned}
$$

which is positive when

$$
\eta < \left(\frac{A_1/B_1}{A_2/B_2} \right)^2 .
$$

This means that when $\eta < \min \left\{ \left(\frac{A_2/B_2}{A_1/B_1} \right)^2, \left(\frac{A_1/B_1}{A_2/B_2} \right)^2 \right\}$, both individuals' self-centered wellbeing rises with a common increase in η. When $\eta > \max \left\{ \left(\frac{A_2/B_2}{A_1/B_1} \right)^2, \left(\frac{A_1/B_1}{A_2/B_2} \right)^2 \right\}$, both individuals' self-centered wellbeing decreases. When η lies in the interval between these two levels, the two wellbeing levels move in opposite directions.

Observe that the two bounds of the interval are on opposite sides of 1. If one takes the benchmark $\eta = 1$ as the maximum level that one can reasonably expect from morally motivated individuals, as it represents full impartiality (and full moral agreement between the two individuals), the second bound is never reached. The picture is then quite clear. For low values of η, both individuals gain from an increase in this altruism parameter. When η crosses the lower bound, the individual with the greater value of the capacity-need ratio A_i/B_i loses from a further increase in η, whereas the other individual continues to benefit. As Rawls would put it, the individual with the greater capacity-need ratio feels "strains of moral commitment" when an increased concern for the other, even shared throughout society, implies a reduction in her own self-centered wellbeing.

4.5 The caring approach

As noted, the caring approach is appropriate to communities with "thick" social ties, enabling their members to participate in each other's wellbeing. Such participation is articulated in the *"principle of care,"* which prescribes altruism[48] and is articulated in the Golden Rule (treating others as one wants to be treated oneself), espoused by most of the world's religions. This morality is supported psychologically by the care motive, involving emotions of warmth, friendship, affection, or love.[49] Care-motivated individuals show positive concern with the well-being of others.[50] Such individuals tend to have a reduced cognitive focus on their own needs in favor of others' needs.[51] They tend to engage in non-instrumental interpersonal sharing and to make choices that benefit another individual or group.[52]

Among thick social groups, pursuing the morality of care has important advantages over deontological moralities, particularly in times of change. In such times, the danger of evolutionary mismatch in the pursuit of abstract deontological principles is relatively large, whereas direct participation in the wellbeing of other people whom one knows well permits flexible adjustment to such contextual change.

In the analysis below, we portray individual i's participation in the wellbeing of individual j by including j's self-interested utility function, as perceived by i, in i's utility function. In general, j's

---

[48] For an excellent overview, see Ricard (2015).
[49] McAdams & Powers (1981); McAdams, (1989); Weinberger, Cotler, & Fishman (2010).
[50] McAdams, Healy, & Krause (1984).
[51] Batson et al. (1987).
[52] French, (1956, 1958); McAdams (1989); Klimecki et al. (2013, 2014).

self-interested utility is weighted relative to i's self-interested utility: equally weighted utilities stand for perfect altruism; when i weights her own self-interested utility more than j's self-interested utility, the care motive is muted. This approach can be articulated through a consequentialist approach to the morality of care.

The caring utility function of individual i may be expressed as

$$U_i = \left(B_i \left(A_i Y_i + A_j Y_j \right) - \frac{1}{2} Y_i^2 \right) + \kappa_i \widetilde{V}_{ji}$$

where the parameter κ_i $(0 \leq \kappa_i \leq 1)$ denotes the degree of care, $\widetilde{V}_{ji}$ is the utility attributed to individual j by i:

$$\widetilde{V}_{ji} = \widetilde{B}_{ji} \left(A_i Y_i + \widetilde{A}_{ji} Y_j \right) - \frac{1}{2} Y_j^2$$

where $\widetilde{B}_{ji}$ and $\widetilde{A}_{ji}$ are attributions to j by i.

Regarding the degree of care, at one extreme, $\kappa_i = 0$ represents no care (pure self-interest), while at the other extreme, $\kappa_i = 1$ represents perfect care.

With this formulation, the corresponding contribution levels are given by:

$$Y_i^C = A_i \left(B_i + \kappa_i \widetilde{B}_{ji} \right) \tag{24}$$

Obviously, the greater is the degree of care and the more accurate are the attributions $\widetilde{B}_{ji}$, the closer these effort levels come to their social optimal ones. When $\kappa_1 = \kappa_2 = 1$, $\widetilde{B}_{21} = B_2$ and $\widetilde{B}_{12} = B_1$, then these effort levels are identical to their socially optimal counterparts (for the social welfare objective giving equal weights to both individuals): $Y_1^C = Y_1^*$ and $Y_2^C = Y_2^*$.

A comparison of (24) with (20)-(21) shows the mathematical similarity between these two cases. Nevertheless, it is important to emphasize that the psycho-social mechanisms underlying the morality of care are quite distinct from those underlying the deontological moral principles above. Whereas the deontological principles result from the rational application of moral precepts (e.g. reward merit, reward need, defer to accept standards of social standing), the morality of care involves empathetic concern and the imaginative attribution of mental states to other people. These abilities can rest on distinct capacities, such as to simulate other people's mental states, i.e. putting ourselves in other people's shoes (along the lines of "simulation theory"),[53] or to infer others' mental states from token behavior patterns (along the lines of "theory theory")[54], possibly along Bayesian lines.[55]

5 Moral motives in a first-order competitive context

Up to now we have considered only positive externalities from individual effort contributions to the public good. Now consider the case in which these expenses only serve to advance one's status in society.

These effects of status competition can be modeled by modifying the private utility function as

---

[53]See, for example, Batson (1991), de Vignnemont and Singer (2006),
Gallese, Keysers and Rizzolatti (2004) and Preston and de Wall (2002).
[54]See, for example, Bonawitz , van Schijndel, and Schulz (2012).
[55]For example, Gopnik (2011) and Ullman and Tenembaum (2020).

follows:

$$V_i = b_i \left(a_i Y_i - a_j Y_j \right) - \frac{1}{2} Y_i^2,$$

where the parameter a_i measures the status creation efficiency of effort Y_i and b_i converts status disparity into private utility. Thus, from the private first-order conditions we get effort levels as:

$$Y_i^P = b_i a_i. \tag{25}$$

With a social welfare function $\alpha_1 V_1 + \alpha_2 V_2$, the first-order conditions for group optimization give the following socially optimal levels of effort:

$$Y_i^* = b_i a_i - \frac{\alpha_j}{\alpha_i} b_j a_i. \tag{26}$$

Comparing (25) with (26), we see that in the competitive setting, self-interested effort on the part of each individual exceeds socially optimal effort. The reason is that status seeking behavior is associated with a negative externality: more status gained by one individual means less status for the other.

As before, let the moral utility functions be ordinally equivalent to

$$M_i = V_i + R_i Y_i,$$

the maximization of which yields effort levels

$$Y_i^M = b_i a_i + R_i. \tag{27}$$

From (26) and (27) we get the optimal moral incentive factors:

$$R_i^* = -\frac{\alpha_j}{\alpha_i} b_j a_i. \tag{28}$$

As seen from the negative term in (28), to mitigate the damage from this negative externality, the moral incentive now represses competition among individuals. It has been argued that repression of competition is one of the major forces in the evolution of human cooperation, alongside kin selection and reciprocal altruism.[56] Repression of competition aligns individual interests with group interests, because when competition has been repressed, an individual can make herself better off primarily by promoting the interests of her group. Thus, repression of within-group competition promotes group selection. Alexander (1979, 1987) argued that repression of within-group competition shaped social structures over the past millenia, favoring successful groups and limiting the opportunities for reproductive dominance. Maynard Smith (1988) provided a general formulation of the principle that group-level selection requires suppression of within-group competition.[57]

Moral principles underlying the repression of competition are well-known. Adam Smith's "impartial spectator" represents one prominent approach, namely, if a person "If he would act so as that the impartial spectator may enter into the principles of his conduct, which is what of all things he has the greatest desire to do, he must, upon this, as upon all other occasions, humble the arrogance of his self-love, and bring it down to something which other men can go along with." (Smith 1996, p. 120). Another major approach is the "veil of ignorance" of Rawls' (1971), whereby a just society follows

---

[56]See, for example, Frank (2003).

[57]"One can recognize in the evolution of life several revolutions in the way in which genetic information is organized. In each of these revolutions, there has been a conflict between selection at several levels. The achievement of individuality at the higher level has required that the disruptive effects of selection at the lower level be suppressed. (Maynard Smith 1988, pp. 229–230).

rules that individuals regard as fair behind a veil of ignorance concerning their position in society. This reasoning also suppresses selfish competition among individuals.

In this context, it is noteworthy that the composition of the socially optimal moral incentives is the same as for the cooperative context, except that now the moral incentive promotes the suppression of competition—the control of greed—rather than the expansion of cooperation in public good provision:

- The greater is an individual's ability (a_i), the greater should be the moral reward for suppressing greed. In other words, merit obligates: greater ability calls for more forbearance in pursuit of self-interest.

- The greater is an individual's need (b_i), the more the other individual is obligated to suppress greed. The morality of benevolence promotes voluntary restraint in exploiting one's position of status.

- The greater is the relative weight of an individual's wellbeing in social wellbeing (such as α for individual 2), the more the other individual needs to suppress greed and the less the individual in question needs to do so. For example, if poorer individuals receive relatively more weight in social wellbeing, then the poor individuals have less obligation to suppress their greed and the rich individuals have more obligation to do so.

Let us now put together the public good contribution game and status competition, considering the case in which status can be acquired through the contribution to the public good—one can think of philanthropy as a leading example of this phenomenon. The effects of status competition of this particular sort can be represented with the following utility function, which combines the two previous models:

$$V_i = b_i \left(a_i Y_i - a_j Y_j \right) + B_i \left(A_i Y_i + A_j Y_j \right) - \frac{1}{2} Y_i^2,$$

From the private first-order conditions the effort levels are as follows:

$$Y_i^P = b_i a_i + B_i A_i. \tag{29}$$

With the social welfare function $\alpha_1 V_1 + \alpha_2 V_2$, the first-order conditions imply the following socially optimal levels of effort:

$$Y_i^* = b_i a_i + B_i A_i - \frac{\alpha_j}{\alpha_i} \left(b_j a_i - B_j A_i \right). \tag{30}$$

This socially optimal level would be obtained with moral incentive factors equal to:

$$R_i^* = \frac{\alpha_j}{\alpha_i} \left(B_j A_i - b_j a_i \right). \tag{31}$$

This formula retrieves Benabou and Tirole's (2006) observation that the Pigou subsidy for public goods must be reduced by a tax on reputation-seeking. In absence of external Pigouvian intervention, it is clear that status competition contributes to alleviating the insufficient private incentives to contribute to the public good. If $B_j A_i = b_j a_i$, the moral factor vanishes and is no longer needed to yield a socially optimal allocation. It may even happen that status competition could induce an excessive production of public good. In the case in which there are multiple public goods, status competition may also fail to match social optimality in terms of the mix of public goods—a well-known problem with private philanthropy.

6 Second-order cooperative and competitive contexts

Thus far, moral principles have been interpreted as self-imposed Pigouvian taxes and subsidies, levied on contributions to the public good (or on positional expenses). But this interpretation holds only for models in which there is no interaction among individuals in public good provision. We now consider second-order cooperative and competitive contexts, where the two individuals' contributions to the public good are strategic complements and substitutes in production, respectively. In a second-order cooperative context, the productivity of one individual's contribution to the public good depends positively on the contribution of the other individual, where as in a second-order competitive context, the productivities of the two individuals are negatively related.

In particular, let the production function for the public good be

$$G = Y_1 + Y_2 + EY_1Y_2 \tag{32}$$

where E is a positive parameter representing the magnitude of the strategic complementarity or a negative parameter representing the magnitude of the strategic substitutability. The strategic complementarity can be interpreted in terms of a public good that takes the form of a mutually beneficial social relationship, such as in a relationship of care. The strategic substitutability, on the other hand, can be interpreted in terms of jobs involving congestion in production.

Assuming again that individual i's benefit from G is B_iG and that the cost of effort Y_i for individual i is $\frac{1}{2}Y_i^2$, then individual i's self-interested utility is

$$V_i = B_i \left(Y_1 + Y_2 + EY_1Y_2 \right) - \frac{1}{2}Y_i^2. \tag{33}$$

For simplicity, we have now assumed that the two individuals have equal productivity, allowing us to focus on the strategic complementarity ($E > 0$) or substitutability ($E < 0$).[58]

Simplifying for the purpose of exposition, let us focus on a form of social welfare that is symmetric between the two self-interested utilities:

$$\Omega = V_1 + V_2 = (B_1 + B_2)(Y_1 + Y_2 + EY_1Y_2) - \frac{1}{2}Y_1^2 - \frac{1}{2}Y_2^2$$

The first-order conditions for social optimality yield the individuals' optimal contributions to the public good:

$$Y_i^* = (B_1 + B_2)(1 + EY_j) \tag{34}$$

and imply that the optimal contributions are

$$Y_1^* = Y_2^* = \frac{B_1 + B_2}{1 - E(B_1 + B_2)}. \tag{35}$$

We now represent the objective of each morally-driven individual in terms of psychic rewards from contributing to the public good, both individually (R_i per unit of contribution to Y_i) and in interaction with the other individual (the value of R_i may depend on Y_j):

---

[58]Allowing for productivity differences clearly makes no difference to our qualitative results.

$$M_i \quad = \quad B_i \left(Y_1 + Y_2 + E Y_1 Y_2\right) + R_i Y_i - \frac{1}{2} Y_i^2 \tag{36}$$

The first-order conditions for moral behavior imply that

$$Y_i^M \quad = \quad B_i \left(1 + E Y_j\right) + R_i \tag{37}$$

And social optimality is obtained when each individual adopts a moral factor equal to:

$$R_i = B_j \left(1 + E Y_j\right) \tag{38}$$

This means that each individual i's contribution to the public good should be determined by

- the other individual's need (B_j) and

- a principle of moral reciprocity, in which the degree of reciprocation depends on two morally relevant characteristics: the other individual's need (B_j) and the complementarity or substitutability $(E Y_j)$.

The interpersonal moral incentives explicitly call for coordination of the individuals' contributions. This requirement reflects a common feature of moral judgments, namely, prescribing behavior not just for oneself individually, but also for one's interactions with others. The interpersonal moral incentives may be characterized as "moral reciprocity," whereby the degree of reciprocity depends on particular morally relevant characteristics (such as need or merit). Under strategic complementarity $(E > 0)$, such as when the public good is a mutually beneficial social relationship, these conditions include the principle of moral reciprocity (the Silver Rule): each individual has a moral incentive to reciprocate the other individual's contribution to the relationship, weighted in accordance with morally relevant factors.

By contrast, under strategic substitutability $(E < 0)$, the opposite principle applies. This may be called the *"principle of moral redress:"* the less one individual contributes to the public good, the more another individual is obligated to contribute. This principle is not surprising, since one individual's contribution reduces the other individual's productivity in contributing to the public good. Under these circumstances, one individual's failure to contribute adequately to the public good should be "redressed" by the other individual. The degree of redress should depend on the magnitude of the strategic substitutability (Y_j) and the other's need (B_j).

To recap: Strategic interdependence in the social context calls for strategic interdependence of moral incentives. (a) A strategic complementarity in the provision of the public good calls for the inclusion of the principle of moral reciprocity into individuals' moral incentives. The degree of reciprocation depends positively on another's need and on the magnitude of the strategic complementarity. (b) When people's contributions to the public good are strategic substitutes, individuals are enjoined to pursue the principle of moral redress, i.e. compensating for others' miserliness and taking advantage of others' generosity. The degree of redress should depend positively on others' needs and the magnitude of strategic substitutability.

It is important to recognize the significance of this principle. Given the behavior of individual j, individual i has a moral incentive to respond reciprocally in the case of strategic complements (or through redress in the case of strategic substitutes). This holds regardless of whether individual j is

22

behaving in the socially optimal way.[59]

Naturally, under the assumption that both individuals are behaving optimally, it is possible to specify each individual's moral incentives in purely individualistic terms. It is easy to derive from (35) and (37) that the socially optimal moral incentives for the two individuals are:

$$R_i = B_j \left(1 + E \frac{(B_1 + B_2)}{1 - E(B_1 + B_2)} \right) \tag{39}$$

But this specification of moral incentives is not just impracticably complicated; it also misses an essential point: The socially desirable moral incentive calls for moral reciprocity (or redress) regardless of whether the other individual is behaving optimally. In the real world, optimal behavior of others cannot be taken for granted. When the other individual is not behaving optimally, then equation (39) yields less social welfare than equation (38). In short, the act of reciprocity is itself in the social interest.

7 Application: Moral implications of technological market bias

On the basis of the analysis above, we now inquire how morality should evolve in response to the phenomenon of technological market bias, whereby the productivity growth generated by technological advances falls more on market activities than on prosocial relationships.[60]

The rationale for this phenomenon is analogous to the "Baumol effect:" The amount of time required for prosocial relationships (such as relations with family and friends) has changed much less than the amount of time to produce many commodities. For example, while the productivity in producing computing power has increased more than a million-fold over the past half century, the time required for successfully raising one's children has not fallen to anything approaching one millionth.

Just as technological change is often more closely associated with market activities than with prosocial relationships, so is positional competition. The reason is that positional competition generally requires a measuring rod whereby status may be assessed. Market activities provide such a measuring rod: prices. Prosocial activities are not only harder to measure; measurement is not relevant to their performance. Consequently, people tend to compete in terms of income and conspicuous consumption, not in terms of the affection they receive from family and friends.[61]

Consider an economy that produces two outputs: a public good (which may be interpreted as a prosocial relationship) and a private good. For simplicity, let the private good be devoted entirely to positional competition.[62] Each individual has a fixed endowment W_i, which is divided between her contribution to the public good and her expenditure on the private good. For individual i, let the contribution to the public good be Y_i and the consumption of the private good be $(W_i - Y_i)$.

As above, we can restrict ourselves to consideration of two individuals. The magnitude of the public good is

$$G = A_1 Y_1 + A_2 Y_2 \tag{40}$$

where A_i is the productivity of individual i.

---

[59]Furthermore, note that the principle of need plays a dual role here, directly and via the degree of interdependence. Since we assumed that the individuals have identical productivity and receive the same weight in the social welfare function, the principles of merit and distributive justice have no role to play.

[60]For example, Snower and Bosworth (2016).

[61]However, in the last decade, social media have introduced measures of social success that exacerbate competition in the realm of social relations. But these measures of social success do not capture some essential qualitative aspects of family relations and friendship.

[62]Provided that positional competition does not decline with technological advance, the inclusion of non-positional goods does not change the qualitative conclusions of our analysis. (See Snower and Bosworth (2016).)

Individual i's utility from the public good is $B_i G$. Let her direct utility from the consumption good be $\frac{1}{1-\rho} \left(C_i \left(W_i - Y_i \right) \right)^{1-\rho}$, where $C_i > 0$ is the productivity of agent i with respect to the private good and ρ is the elasticity of marginal utility. In accordance with the technological market bias and the positional competition bias hypotheses, technological advance can be captured primarily by a rise in the parameter C_i, representing the productivity in producing the positional good.[63]

Furthermore, let individual 1 have a greater endowment than individual 2: $W_1 > W_2$. We consider two costs of inequality, one falling on the poorer individual (ε) and the other on the richer individual (γ). We may think of these costs as "envy" and "guilt" in the context of inequality aversion context of Fehr and Schmidt (1999), psychological and social costs of social fragmentation,[64] and so on.[65] Individual 1 experiences "guilt" with utility $\gamma \left(C_1 \left(W_1 - Y_1 \right) - C_2 \left(W_2 - Y_2 \right) \right)$, and Individual 2 experiences "envy," with utility $\varepsilon \left(C_1 \left(W_1 - Y_1 \right) - C_2 \left(W_2 - Y_2 \right) \right)$, where γ, ε are positive parameters.

Then the utility of individual i is

$$V_i = B_i \left(A_1 Y_1 + A_2 Y_2 \right) + \frac{1}{1-\rho} \left(C_i \left(W_i - Y_i \right) \right)^{1-\rho} - \theta_i \left(C_1 \left(W_1 - Y_1 \right) - C_2 \left(W_2 - Y_2 \right) \right),$$

with $\theta_1 = \gamma$, $\theta_2 = \varepsilon$.

Under self-interested behavior, each individual maximizes her utility with respect to her activities, so that

$$Y_1 = W_1 - \frac{C_1^{\frac{1-\rho}{\rho}}}{\left(B_1 A_1 + \gamma C_1 \right)^{\frac{1}{\rho}}}$$

$$Y_2 = W_2 - \frac{C_2^{\frac{1-\rho}{\rho}}}{\left(B_2 A_2 - \varepsilon C_2 \right)^{\frac{1}{\rho}}}$$

It is worth noting that the Baumol effect, represented by the $C_i^{\frac{1-\rho}{\rho}}$ term in these formulas, is absent in the salient case of a logarithmic utility ($\rho = 1$). When $\rho > 1$, as is claimed in various empirical approaches to the estimation of the elasticity of marginal utility, the Baumol effect is actually reversed, in the sense that a greater productivity of consumption induces, through this term, a greater investment in the public good. The "guilt" term for the rich reinforces this phenomenon, whereas the "envy" term for the poor attenuates it.

Let us now focus on the case of logarithmic utility ($\rho = 1$). It can be shown that the social optimum maximizing social welfare ($\alpha_1 V_1 + \alpha_2 V_2$) with respect to Y_1 and Y_2 is achieved by:

$$Y_1^* = W_1 - \frac{1}{A_1 \left(B_1 + \frac{\alpha_2}{\alpha_1} B_2 \right) + C_1 \left(\gamma + \frac{\alpha_2}{\alpha_1} \varepsilon \right)}$$

$$Y_2^* = W_2 - \frac{1}{A_2 \left(\frac{\alpha_1}{\alpha_2} B_1 + B_2 \right) - C_2 \left(\frac{\alpha_1}{\alpha_2} \gamma + \varepsilon \right)}$$

Moral individuals, once again, would have the following objective: $M_i = V_i + R_i Y_i$. These individuals choose the following activity levels:

---

[63]Insofar as technological advance falls on the prosocial relationships as well, it is captured by a rise in the parameter A_i.

[64]Examples include rising crime, falling trust and political gridlock. See, for instance, Atkinson (2015).

[65]Alternatively, we may think of the parameter γ as representing "pride," a psychic benefit gained by the winner of the positional competition. This parameter then has a negative value in our analysis.

$$Y_1^M = W_1 - \frac{1}{B_1 A_1 + C_1 \gamma + R_1}$$
$$y_2^m = W_2 - \frac{1}{B_2 A_2 - C_2 \varepsilon + R_2}$$

Thus, it is easy to see that the socially optimal moral incentives are:

$$R_1 = \frac{\alpha_2}{\alpha_1} (A_1 B_2 + C_1 \varepsilon)$$

$$R_2 = \frac{\alpha_1}{\alpha_2} (A_2 B_1 - C_2 \gamma)$$

Interpreting the parameters γ and ε in terms of the social costs of inequality, we find that:

- the greater the social cost of inequality falling on the advantaged agent (i.e., the greater γ is), the more the disadvantaged agent is relieved of moral responsibility to contribute to the public good, and

- the greater the social cost of inequality falling on the disadvantaged agent (i.e., the greater ε is), the greater is the moral responsibility of the advantaged agent to contribute to the public good.

Regarding the role of technological progress (raising the C_i parameters), a few insights stand out:

- technological progress implies a rise in the moral incentive R_1 and a fall in the moral incentive R_2.

- insofar as technological progress falls primarily on the better-endowed individual (through C_1), the rise in R_1 will exceed the fall in R_2 in absolute magnitude, meaning that technological progress should be accompanied by an ongoing rise in the richer individual's moral responsibility for contributing to the public good and an ongoing fall in the poorer individual's moral responsibility for such contributions.

It is also worth noting that if technological advance falls on the prosocial relationships as well, this development is captured by a rise in the parameter A_i, putting upward pressure on the moral responsibility of both agents, in proportion to their productivity in generating these prosocial relationships.

8 Group formation by moral affinity

Generalizing the model of subsection 4.4 to many individuals, one can build a theory of group formation along the following lines. Suppose that the public good is only produced and consumed within each group, and that when a group forms, a common norm of moral concern for others in the group takes the form of a coefficient η in the individuals' objective function:

$$M_i = V_i + \eta \sum_{j \neq i} V_j - M_i^*.$$

This coefficient may be specific to the group. The variable M_i^* serves the calibrate the specific effect of morality on well-being, and is further discussed below. Similarly as before, the public good in a group (which is now, therefore, a local public good) is equal to $G = \sum_i A_i Y_i$, and self-centered wellbeing is

equal to

$$V_i = B_i G - \frac{1}{2} Y_i^2,$$

where A_i is the individual's "capacity" and B_i is the individual's "need". When maximizing M_i, individual i, adhering to the moral norm embodied in the "moral concern coefficient" η, contributes

$$Y_i = A_i \left(B_i + \eta \sum_{j \neq i} B_j \right),$$

so that the quantity of public good produced within the group is

$$G = \sum_i A_i^2 \left(B_i + \eta \sum_{j \neq i} B_j \right) = \sum_i A_i^2 B_i + \eta \sum_{i,j \neq i} A_i^2 B_j.$$

The consequences for self-centered well-being of an increase in the moral concern coefficient combine the positive effect of a greater quantity of public good, and the negative effect of a greater contribution:

$$\frac{\partial V_i}{\partial \eta} = B_i \sum_{k,j \neq k} A_k^2 B_j - A_i^2 \left(B_i + \eta \sum_{j \neq i} B_j \right) \sum_{j \neq i} B_j.$$

This expression is positive, i.e., well-being is advanced through an increase in the moral concern coefficient, when

$$\eta < B_i \frac{\sum_{k,j \neq k} A_k^2 B_j - A_i^2 \sum_{j \neq i} B_j}{A_i^2 \left(\sum_{j \neq i} B_j \right)^2} = \frac{B_i}{A_i^2} \sum_{j \neq i} \frac{A_j^2}{B_j} \frac{\sum_{k \neq j} B_j B_k}{\sum_{l \neq i, k \neq i} B_l B_k}.$$

In this expression, one sees an individual with a high capacity-need ratio A_i^2/B_i will perceive greater moral concern as a cost for self-centered well-being over a wider range of values for η. In other words, moral fatigue (or "strains of commitment", the expression coined by Rawls 1971) occurs for lower values of η among individuals with relatively high capacity-need ratio.

In order to study group formation through individuals migrating among groups, a notion of equilibrium must be defined. Let us define an equilibrium by the condition that every i belongs to a group that gives i the greatest value of utility, among the existing groups (and the singleton formed by i alone). This Nash stability (Bogomolnaia and Jackson 2002) is admittedly a liberal notion of equilibrium, as it ignores the possible breaking up of groups by subgroups coordinating to form more advantageous coalitions. It allows for multiple equilibria very easily. In particular, the whole population generally forms a single-group equilibrium because no other group competes with it to attract migrants and singletons cannot generate much public good. One advantage of this approach is that it involves minimal informational and rationality assumptions, since individuals only consider individually migrating across groups and do not have to coordinate their strategies.

Another key ingredient of the analysis is how migrating to another group with a different level of coefficient η affects utility M_i. The way it affects V_i is easy to compute. The way in which the "moral component" $\eta \sum_{j \neq i} V_j - M_i^*$ varies can depend on the moral psychology of the individual. In section 3, various possible cases have been introduced, with a positive or negative effect of a greater moral concern depending on whether this increases self-esteem or guilt. The first model suggested in section

26

3, adapted here, would have

$$M_i^* = \eta \sum_{j \neq i} V_j^*,$$

where V_j^* is the target level for j's self-centered utility when everyone in the group conforms to the prevailing moral norm. Thus, M_i^* represents the moral goal of individual i—unlike section 3, the moral goal is expressed not in terms of own action but in terms of consequences for others. At the equilibrium, this target level will be achieved, and therefore the "moral component" of utility $\eta \sum_{j \neq i} V_j - M_i^*$ is null (because $V_j = V_j^*$ for all j).

This type of moral psychology corresponds to the individual who feels "just fine" when doing the right thing. When moving to another group with greater morality, several effects will modify the components. The greater moral coefficient η may, under certain conditions discussed above, raise the moral target $\sum_{j \neq i} V_j^*$, which has a depressing effect on the moral component of utility because this particular morality is more demanding. On the other hand, the equilibrium $\sum_{j \neq i} V_j$ also becomes greater which makes the individual feel better, and the two effects exactly compensate one another at the equilibrium. In addition, the greater value of η makes the individual more sensitive to both effects, more vulnerable to guilt and readier to feel good with good deeds. At the end of the day, the moral component of utility always remains null at the equilibrium, which simplifies the analysis of migrations because only the component V_i of M_i may possibly vary when the individual moves to another group and conforms to that group's moral norm. Later in this section we shall consider an alternative case in which a greater η coefficient may have a net (positive or negative) effect on the moral component of utility at the equilibrium.

In the first case of a null moral component of utility at the equilibrium, individual i's utility from joining group g is equal to

$$M_i = V_i = B_i G_g - \frac{1}{2} Y_{ig}^2,$$

where

$$G_g = A_i^2 B_i + \sum_{j \in g} A_j^2 B_j + \eta_g \left(A_i^2 \sum_{j \in g} B_j + B_i \sum_{j \in g} A_j^2 \right) + \eta_g \sum_{j \in g, k \neq j} A_j^2 B_k,$$

$$Y_{ig} = A_i \left(B_i + \eta_g \sum_{j \in g} B_j \right).$$

Then, an equilibrium is characterized by the fact that the possible groups making up the population $N \setminus \{i\}$ and that i could join,[66] have parameters $(\alpha_g, \beta_g, \gamma_g, \eta_g)$ such that i actually joins a group g yielding the greatest value for the expression

$$\alpha_g B_i - \beta_g \frac{A_i^2}{B_i} + \gamma_g, \tag{41}$$

where $\alpha_g = \eta_g \sum_{j \in g} A_j^2$, $\beta_g = \frac{1}{2} \left(\eta_g \sum_{j \in g} B_j \right)^2$, $\gamma_g = \sum_{j \in g} A_j^2 \left(B_j + \eta_g \sum_{k \in g, k \neq j} B_k \right)$. Expression (41) comes from eliminating terms that are redundant or do not depend on g in the developed formula for V_i.

When $\eta_g = 0$, (41) boils down to $\sum_{j \in g} A_j^2 B_j$, which implies that selfish groups are appreciated in the same way by all prospective members. In particular, when all groups are selfish, all individuals want to move to the group with the largest $\sum_{j \in g} A_j^2 B_j$, and no partition of the population is possible unless it is composed of groups with exactly the same value for this magnitude. The most likely

---

[66]We neglect the possibility for i to form a singleton.

outcome is therefore that the whole population forms a single group.

Let us now turn to the more interesting case of groups with varying levels of moral concern. Some partitions of the population are impossible in an equilibrium, and this gives a good sense of the association between group morality, size, capacity and needs in a typical equilibrium. Indeed, consider two homogeneous groups g, g' in which all members are identical within each group. Stability of the equilibrium requires that no one from either group wants to migrate to the other group. This holds true if

$$(\alpha_g - \alpha_{g'}) B_g - (\beta_g - \beta_{g'}) \frac{A_g^2}{B_g} + (\gamma_g - \gamma_{g'}) \geq 0 \geq (\alpha_g - \alpha_{g'}) B_{g'} - (\beta_g - \beta_{g'}) \frac{A_{g'}^2}{B_{g'}} + (\gamma_g - \gamma_{g'}), \quad (42)$$

where A_g, B_g denote the common characteristics of members of g, and likewise for $A_{g'}, B_{g'}$ in group g'.

A necesssary condition for (42) to hold is that

$$(\alpha_g - \alpha_{g'}) (B_g - B_{g'}) - (\beta_g - \beta_{g'}) \left(\frac{A_g^2}{B_g} - \frac{A_{g'}^2}{B_{g'}} \right) \geq 0.$$

This can be rewritten as

$$\left(\eta_g n_g A_g^2 - \eta_{g'} n_{g'} A_{g'}^2 \right) (B_g - B_{g'}) \geq \frac{1}{2} \left[(\eta_g n_g B_g)^2 - (\eta_{g'} n_{g'} B_{g'})^2 \right] \left(\frac{A_g^2}{B_g} - \frac{A_{g'}^2}{B_{g'}} \right), \quad (43)$$

where $n_g, n_{g'}$ denote the population sizes of the two groups. Focusing on the sign of the four terms making up this inequality, one obtains one salient pattern that is impossible to find in an equilibrium, when $B_g \geq B_{g'}$.

This impossible pattern is defined by

$$\frac{\eta_{g'} n_{g'}}{\eta_g n_g} > \frac{B_g}{B_{g'}} \geq \sqrt{\frac{B_g}{B_{g'}}} > \frac{A_g}{A_{g'}}. \quad (44)$$

To check that this pattern violates the necessary condition (43), observe that $\frac{\eta_{g'} n_{g'}}{\eta_g n_g} > \frac{B_g}{B_{g'}}$ implies $(\eta_g n_g B_g)^2 - (\eta_{g'} n_{g'} B_{g'})^2 < 0$ and $\sqrt{\frac{B_g}{B_{g'}}} > \frac{A_g}{A_{g'}}$ implies $\frac{A_g^2}{B_g} - \frac{A_{g'}^2}{B_{g'}} < 0$ so that the right-hand side is positive because both terms are negative.[67] The left-hand side is non-positive because $B_g - B_{g'} \geq 0$ while $\eta_g n_g A_g^2 - \eta_{g'} n_{g'} A_{g'}^2 < 0$ since

$$\frac{\eta_{g'} n_{g'} A_{g'}^2}{\eta_g n_g A_g^2} > \frac{B_g}{B_{g'}} \frac{A_{g'}^2}{A_g^2} > 1.$$

The lesson from (44) is clear: it is impossible to have a group that is substantially needier, compared to the capacity ratio, and to have this group embrace a relatively low degree of moral concern or have a relatively small size. In particular, comparing two groups of equal size, a group with greater needs

---

[67]Another sign configuration would have both terms positive on the right-hand side and one negative sign on the left-hand side. But one can check that this impossible. Indeed, the inequalities $\frac{B_g}{B_{g'}} > \frac{\eta_{g'} n_{g'}}{\eta_g n_g}$ and $\frac{A_g}{A_{g'}} > \sqrt{\frac{B_g}{B_{g'}}}$ are necessary for both terms on the right-hand side of (43) to be positive, but this also implies $\eta_g n_g A_g^2 - \eta_{g'} n_{g'} A_{g'}^2 > 0$ because $\frac{\eta_{g'} n_{g'} A_{g'}^2}{\eta_g n_g A_g^2} < \frac{\eta_{g'} n_{g'}}{\eta_g n_g} \frac{B_{g'}}{B_g} < 1$.

($\frac{B_g}{B_{g'}} > 1$) and lower capacity ($\frac{A_g}{A_{g'}} < 1$) must necessarily have

$$\frac{\eta_g}{\eta_{g'}} \geq \frac{B_{g'}}{B_g}.$$

When $\frac{B_{g'}}{B_g} < 1$, this is compatible with $\eta_g < \eta_{g'}$, but this puts a lower bound on the ratio $\frac{\eta_g}{\eta_{g'}}$. When needs are identical, ($\frac{B_g}{B_{g'}} = 1$), one cannot have

$$\frac{\eta_{g'} n_{g'}}{\eta_g n_g} > 1 > \frac{A_g}{A_{g'}},$$

which means that if a group has the same size but lower capacity, it must necessarily have at least as great a value for η_g.

Thus, the typical partition that would emerge in an equilibrium would have groups with higher needs and lower capacity tend to adopt higher moral concern or be larger. This makes sense, since greater needs correspond to a greater demand for the public good, the production of which is enhanced by the level of moral concern. As for capacity, we have seen that it tends to reduce the tolerance for high moral concern because the demands on able individuals to contribute to the public good increase with moral concern. Population size plays a similar role as moral concern, as, for a given positive level of moral concern, it increases the moral requirements on public good contributions, which is attractive for needy individuals and less so for high-capacity individuals.

Now consider the case in which the coefficient η_g may have a direct utility effect on M_i at the equilibrium, and allowing for heterogeneity among individuals regarding this effect. This can be studied by slightly modifying the previous analysis, introducing an additional term in the equilibrium value of M_i as follows:

$$M_i = V_i + \eta_g S_i,$$

implying that individual i seeks a group that maximizes

$$\alpha_g B_i - \beta_g \frac{A_i^2}{B_i} + \gamma_g + \eta_g \frac{S_i}{B_i},$$

where S_i is a personal sensitivity (which can be positive or negative) to η_g. In this case, the previous analysis is modified by the fact that, other things equal, high-morality groups will tend to attract individuals with greater positive S_i and repulse individuals with negative S_i. This opens new possibilities, depending on whether this effect is minor or, on the contrary, dominates the high-ability-low-need ranking identified previously as central to group formation. In the latter case, the population can split into "cultural groups" along affinities for moral concern rather than socioeconomic capacity-needs characteristics. Interestingly, the above formula shows that individuals with high needs will be, other things equal, less sensitive to the direct utility effect of moral concern than less needy individuals. The formation of "cultural groups" is thus more likely among individuals with low needs, while individuals with high needs will tend to the display the same pattern of group formation as described previously.

To summarize the main results from this section: In absence of moral concern in society, there can be no splitting in separate groups. Under the prevalence of moral parameters of concern shared by individuals in every group, separate groups can form and be stable when individuals have heterogenous capacities and needs. In particular, equilibria separating selfish high-capacity-low-need groups from solidaristic low-capacity-high-need groups tend to arise, unless a direct effect of the society-level degree of moral concern on well-being compensates the strains of moral commitment for high-capacity-low-

need individuals.

9   Concluding thoughts

To gain perspective, it is useful to place our contribution in the broad sweep of moral thinking. Morality appears to have been embedded in human thinking and decision making since the beginnings of humanity. Around 400,000 years ago Homo Heidelbergensis began collaborative hunting and foraging.[68] At this time, collaboration must have become compulsory, since it was vital for survival. Socially cooperative individuals had better chances of finding a mate and thereby cooperation emerged through social selection. This social cooperation, built on joint intentionality, was the source of morality. Around 150,000 years ago, growing tribal groups began splitting into smaller units. Individuals must have felt strong affiliation with their sub-group and somewhat weaker affiliation with their super-group, which needed to be distinguished from other tribes with which one might be in competition. These multi-level affiliations were the beginnings of "culture," from which elaborate moral systems grew. With the invention of agriculture, people formed larger, sedentary groups that had sophisticated divisions of labor and responsibility, reinforced by "objective moralities" whereby people shared common conceptions of being both judges and judged.[69] These moral systems could be understood in terms of highly context-dependent behavioral rules.

During the Axial Age (from around 800 to 200 BCE), many societies developed systems of virtue ethics,[70] which encompassed not just rules of behavior, but also worldviews that shaped perceptions, intuitions and emotions in terms of the relevant physical and social contexts. Within these contexts, people aspired to multiple virtues, to be achieved through ongoing practice and the development of the relevant skills.[71] Virtue-based ethical systems remained dominant throughout the Middle Ages and are enjoying a revival in philosophy nowadays.[72]

In the 18th century, philosophers of the European Enlightenment began constructing ethical systems from secular first principles. The two leading approaches that emerged were deontology (judging actions in terms of whether they are right or wrong) and consequentialism (judging actions in terms of their consequences). Unlike virtue ethics, the underlying principles here were abstract and universal, based on logical reasoning or calculation. Morality became focused on "What should I do?" rather than "Who should I become?"

Moral psychology was narrowed along analogous lines, focusing primarily on quandaries concerning deontological versus consequentialist choices (e.g. whether or not to throw a switch that would divert a trolley to kill one person in order to save five others)[73] and justice versus care. For example, Kohlberg (1969) claimed that moral development in children was about the development of reasoning concerning justice, whereas Gilligan (1982) claimed that such development included the "ethic of care."

Meanwhile anthropologists and cultural psychologists focused on the distinction between individualism and collectivism in the formation of identity.[74] With regard to social affiliations, it is useful to distinguish between what Tönnies (2001/1887) called "Gemeinschaft" (community) and "Gesellschaft" (civil society). The former involves thick ties in communities of place and belief, whereas the latter refers to the thin ties among people who live in close proximity but make their own choices, provided

---

[68]Stiner and Kuhn (2006).
[69]Tomasello (2018).
[70]Aristotle (1941).
[71]MacIntyre (1981), Churchland (1998), McDowell (1979).
[72]For example, Chappell (1996) and Crisp (1996).
[73]For example, Greene (2008), Hauser (2006).
[74]For example, Shweder and Bourne (1984) and Triandis (1995).

that they don't harm others. Moral psychology has been mainly about the obligations that create Gesellschaft, such as respecting the rights of others, not harming others and helping people in need.

More recently, moral and evolutionary psychologists have focused on the function of morality, namely, to promote cooperation among people (particularly in-group) and suppress destructive self-interest.[75] This is the approach pursued in this paper. For this purpose, we began our analysis by considering a public good that is intrinsically worthwhile (without specifying the rationale for this assessment) and a social welfare function that specifies how the individual objectives of different people are to be weighted in assessing social wellbeing. On this basis, we investigated how people need to cooperate for the optimal provision of the public good. Finally, we derived the moral incentives that lead to this socially desirable outcome.

We then conducted an analogous line of inquiry with regard to positional competition. The underlying insight is that social cooperation is often undermined through positional competition (favoring the individual rather than the social group)—except when social status is acquired through cooperative behavior itself. All in all, many of our moral successes can be understood in terms of our capacity to promote social cooperation while suppressing our drive to seek status.

It is commonly observed that economic and social decisions in our daily lives are riddled with moral dilemmas and it is common to feel the tug of conflicting moral demands. For example, should people be supported in accordance with their need or rewarded in accordance with their merit? How should we weigh the rights of individuals and communities in decisions on public infrastructure? To what degree should income be redistributed when redistribution hurts economic efficiency? Is theft condonable if the needs of the perpetrators are substantially greater than those of the victims? What are the moral limits of markets? Should we be able to sell our organs, pay mercenaries to fight our wars, transact pollution rights, sell citizenship to immigrants, pay for basic health care, and so on?

A reason why we perceive these and other issues to be moral dilemmas is that we generally assume moral principles as being mutually exclusive when they pull in contrary directions. In other words, we adopt an "all or nothing" approach to moral principles, e.g. we follow either the ethic of care or the ethic of justice. Another reason why we are so vulnerable to moral dilemmas is our Enlightenment-driven assumption that moral principles are to be abstract and universal, not context-dependent.

Our analysis calls both these assumptions into question. The socially desirable moral motives that we derive can be interpreted in terms of context-dependent combinations of deontological moral principles. Different social contexts generate different collective action problems, which call for different combinations of moral principles. These principles are particularly applicable in the context of Gesellschaft, wherein the members of civil society have little intimate knowledge of each other's sources of wellbeing.

In the context of Gemeinschaft, where people can engage in the requisite perspective-taking concerning the wellbeing of other group members, the optimal provision of the public good and the optimal suppression of positional competitiveness can be achieved through the requisite degree of perspective-taking combined with the morality of care (conceived in consequentialist terms).

In short, our analysis builds bridges between the major approaches to morality—deontology, consequentialism and virtue ethics—and between various deontological principles:

- Deontological principles need to be combined in context-dependent ways to achieve socially desirable outcomes for collective action problems.

- The deontological and consequentialist approaches to morality are alternative routes to achieving

---

[75]For example, Haidt (2007).

31

desirable social outcomes, but under different informational conditions.

- The context-dependence of socially desirable moral principles harkens back to the essential features of virtue ethics: One is meant to pursue multiple virtues and the relevant virtues depend on our physical and cultural setting as well as our social roles within our social hierarchies. In practice, the appropriate combination of moral principles is something that can be learnt only through practice, following in the footsteps of moral exemplars.

The main contribution of our analysis is to analyze the link between social contexts, socially desirable moral motives and decision making. The exploration of a wider range of social contexts and their moral implications is a subject for future research.

Our analysis suggests that including moral motives in the analysis of economic decisions serves several important purposes: it gives useful insights for understanding how people behave in addressing their collective action problems, it provides guidelines for how people ought to apply moral principles in response to new collective action problems, and it provides a starting point for rethinking the division of responsibility between government and civil society.

References

[1] Adolphs, R. (2013). The biology of fear. Current Biology, 23(2): R79–93.

[2] Alexander, R. D. (1979). Darwinism and Human Affairs. Seattle, WA: University of Washington Press.

[3] Alexander, R. D. (1987). The biology of moral systems. Aldine de Gruyter, New York.

[4] Andreoni, J. (1993). An Experimental Test of the Public-Goods Crowding-Out Hypothesis. The American Economic Review, 83(5): 1317–1327.

[5] Andreoni, J. (2006). Philanthropy. Handbook of the economics of giving, altruism and reciprocity, 2: 1201-1269.

[6] Andreoni, J., Bernheim, B. D. (2009). "Social image and the 50–50 norm. A theoretical and experimental analysis of audience effects" . Econometrica, 77: 1607–1636.

[7] Aristotle (1941). Nichomachean Ethics (trans. By W.D. Ross). New York: Random House.

[8] Atkinson, A. (2015). Inequality: What Can Be Done? Cambridge, MA: Harvard University Press.

[9] Atkinson, J. W. (1964). An Introduction to Motivation. Princeton: Van Nostrand.

[10] Atkinson, J. W., Heyns, R. W., & Veroff, J. (1954). The effect of experimental arousal of the affiliation motive on thematic apperception. The Journal of Abnormal and Social Psychology, 49(3): 405.

[11] Avram, J., Balteş, F. R., Miclea, M., & Miu, A. C. (2010). Frontal EEG activation asymmetry reflects cognitive biases in anxiety: Evidence from an emotional face Stroop task. Applied psychophysiology and biofeedback, 35: 285–292.

[12] Batson, C. D. (1991). The altruism question: Toward a social-psychological answer. Lawrence Erlbaum Associates, Inc.

[13] Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. Journal of Personality, 55(1): 19–39.

[14] Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychological bulletin, 117(3): 497.

[15] Benabou, R., Falk, A., and Tirole, J. (2018). Narratives, Imperatives and Moral Reasoning, IZA DP No. 11665.

[16] Benabou, R., Tirole, J. (2006). Incentives and prosocial behavior. American economic review, 96(5): 1652-1678.

[17] Benabou, R., Tirole, J. (2012). Laws and norms, IZA Discussion Paper No. 6290.

[18] Bogomolnaia A., M.O. Jackson (2002). The stability of hedonic coalition structures. Games and Economic Behavior, 38(2): 201–230.

[19] Bonawitz EB, van Schijndel TJ, Friel D, Schulz L (June 2012). Children balance theories and evidence in exploration, explanation, and learning. Cognitive Psychology. 64 (4): 215–234.

[20] Boyatzis, R. E. (1973). Affiliation motivation. Human motivation: A book of readings, 252-276.

[21] Brekke, K.A., S. Kverndokk, K. Nyborg (2003). An economic model of moral motivation. Journal of Public Economics, 87(9–10): 1967-1983.

[22] Buraschi, A., & Cornelli, F. (2002). Donations. Available at SSRN 331485.

[23] Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. Psychological bulletin, 135(2): 183.

[24] Chappell, T.D.J. (1996). Values and Virtues: Aristotelianism in Contemporary Ethics, Oxford: Oxford University Press.

[25] Christakis, N. (2019). Blueprint: The Evolutionary Origins of a Good Society. N.Y.: Little, Brown Spark.

[26] Churchland, P.M. (1998). Toward a Cognitive Neurobiology of the Moral Virtues. Topoi, 17: 83–96.

[27] Corneo, G., and Jeanne, O. (1997). Conspicuous Consumption, Snobbism and Conformism. Journal of Public Economics, 66(1): 55–71.

[28] Crisp, R. (ed) (1996). How Should One Live? Essays on the Virtues. Oxford: Oxford University Press.

[29] Daugherty, A. F. and Reinganum, J. F. (2010). Public goods, social pressure, and the choice between privacy and publicity. American Economic Journal: Microeconomics, 2(2): 191–221.

[30] de Vignemont, F. & Singer, T. (2006). The empathic brain: How, when and why? Trends in Cognitive Sciences. 10 (10): 435–441.

[31] Deutsch, M. (1975). Equity, Equality and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice? Journal of Social Issues, 31: 137–149.

[32] Eimontaite, I., Nicolle, A., Schindler, I., & Goel, V. (2013). The effect of partner-directed emotion in social exchange decision-making. Frontiers in Psychology, 4: 469.

[33] Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. Science, 302(5643): 290–292.

[34] Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. Educational Psychology Review, 13(2): 73–92.

[35] Exline, R. V. (1962). Effects of need for affiliation, sex, and the sight of others upon initial communications in problem solving groups. Journal of Personality, 30(4): 541–556.

[36] Fehr, E., Gächter, S. (2002). Altruistic punishment in humans. Nature 415: 137–140.

[37] Fehr, E. and K.M. Schmidt (1999). A Theory of Fairness, Competition and Cooperation. Quarterly Journal of Economics, 114(3): 817-868.

[38] Fessler, D. M. T. (2010). Madmen: An evolutionary perspective on anger and men's violent responses to transgression. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), International handbook of anger: Constituent and concomitant biological, psychological, and social processes (pp. 361–381).

[39] Fessler, D. M. T., Pillsworth, E. G., & Flamson, T. J. (2004). Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking. Organizational Behavior and Human Decision Processes, 95(1): 107–123.

[40] Fiske, A.P. (1991). Structure of Social Life. New York: Free Press.

[41] Fiske, A.P. (1992). The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations. Psychological Review, 99: 689–723.

[42] Fiske, A. P., & Haslam, N. (2005). The Four Basic Social Bonds: Structures for Coordinating Interaction.

[43] Fleurbaey, M. (2004). Normative economics and theories of distributive justice. In John Bryan Davis & Alain Marciano (eds.), The Elgar Companion to Economics and Philosophy. Edward Elgar. pp. 132.

[44] Fluet C., M.R. Mungan (2022). Laws and norms with (un)observable actions. European Economic Review 145: 104–129.

[45] Folger, R., B.H. Sheppard, and R.T. Buttram (1995). Equity, Equality and Need: Three Faces of Social Justice. In B. Bunker and J.Z. Zucker (eds), Conflict, Cooperation and Justice, San Francisco, CA: Jossey-Bass, 261–289.

[46] Frank, Steven A. (2003). Repression of Competition and the Evolution of Cooperation. Evolution, 57(4): 693–705.

[47] French, E. G. (1956). Motivation as a variable in work-partner selection. The Journal of Abnormal and Social Psychology, 53(1): 96.

[48] Gallese, V.; Keysers, C. & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. Trends in Cognitive Sciences. 8 (9): 396–403.

[49] Gilligan, C. (1982). In a Different Voice: Psychological Theory and Women's Development. Cambridge: Harvard University Press.

[50] Gintis, H. (2007). The Evolution of Private Property. Journal of Economic Behavior and Organization, 64(1), 1-16.

[51] Glazer, A., & Konrad, K. A. (1996). A signaling explanation for charity. The American Economic Review, 86(4): 1019-1028.

[52] Gopnik, A. (2011). The Theory Theory 2.0: Probabilistic Models and Cognitive Development. Child Development Perspectives, 5: 161-163.

[53] Greene, J. (2013). Moral Tribes: Emotion, Reason, and the Gap Between Us and Them. London: Atlantic Books.

[54] Haidt, J. (2007). The New Synthesis in Moral Psychology. Science, 316: 998–1002.

[55] Hall, J. L., Stanton, S. J., & Schultheiss, O. C. (2010). Biopsychological and neural processes of implicit motivation. In O. Schultheiss & J. Brunstein (Eds.), Implicit motives (pp. 279–307). Oxford: Oxford University Press.

[56] Harbaugh, W. T. (1998). The prestige motive for making charitable transfers. The American Economic Review, 88(2): 277–282.

[57] Hare, D., H.K. Reeve and B. Blossey (2016). Evolutionary Routes to Stable Ownership. Journal of Evolutionary Biology, 29(6): 1178–1188.

[58] Haslam, N. (1994). Categories of Social Relationship. Cognition, 53: 59–90.

[59] Haslam, N. (2004). Relational Models Theory: A Contemporary Overview. Mahwah, NJ: Erlbaum.

[60] Haslam, N., and A.P. Fiske (1992). Implicit Relational Prototypes: Investigating Five Theories of Cognitive Organization of Social Relationships. Journal of Experimental Social Psychology, 28: 441–474.

[61] Hauser, M. (2006). Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong. New York: Harper Collins.

[62] Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. Journal of Economic Behavior & Organization, 53(1): 3–35.

[63] Henrich, J. (2016). The Secret of Our Success. Princeton, N.J.: Princeton University Press.

[64] Henrich, J. (2020). The Weirdest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous. London: Penguin.

[65] Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior, 22(3): 165–196.

[66] Herold, F. (2010). Contractual Incompleteness as a Signal of Trust. Games and Economic Behavior 68: 180–191.

[67] Hofstede, G. 2001. Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. Thousand Oaks, CA: Sage.

[68] Hirshleifer, J. (1985). The Expanding Domain of Economics. American Economic Review, 75 (6): 53–69.

[69] Klimecki, O. M., Leiberg, S., Lamm, C., & Singer, T. (2013). Functional Neural Plasticity and Associated Changes in Positive Affect After Compassion Training. Cerebral Cortex, 23(7): 1552–1561.

[70] Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. Social Cognitive and Affective Neuroscience, 9(6): 873–879.

[71] Kohlberg (1969). Stage and Sequence: The Cognitive-Developmental Approach to Socialization. In D.A. Goslin (ed), Handbook of Socialization Theory and Research, Chicago: Rand McNally.

[72] Kornadt, H. J. (1984). Motivation theory of aggression and its relation to social psychological approaches. In A. Mummendey (Ed.), Social psychology of aggression (pp. 21–31). Berlin/Heidelberg: Springer.

[73] Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. PNAS 98 (26): 15387–92.

[74] Lansing, J. B., & Heyns, R. W. (1959). Need affiliation and frequency of four types of communication. The Journal of Abnormal and Social Psychology, 58(3): 365.

[75] LeDoux, J. (1998). The emotional brain: The mysterious underpinnings of emotional life. Simon and Schuster.

[76] Leith, K. P., & Baumeister, R. F. (1996). Why Do Bad Moods Increase Self-Defeating Behavior? Emotion, risk taking, and self-regulation, Journal of Personality and Social Psychology, 71(6): 1250–1267.

[77] Li, N.P., M. van Vugt, and S.M. Colarelli (2018). The Evolutionary Mismatch Hypothesis: Implications for Psychological Science. Current Directions in Psychological Science, 27(1): 38-44.

[78] Lowell, E. L. (1952). The effect of need for achievement on learning and speed of performance. The Journal of Psychology, 33(1): 31–40.

[79] MacIntyre, A. (1981). After Virtue. Notre Dame: University of Notre Dame Press.

[80] Maynard Smith (1982). Evolution and the Theory of Games. Cambridge: Cambridge University Press.

[81] Maynard Smith, J. (1988). Evolutionary progress and levels of selection. Pp. 219–230 in M. H. Nitecki, ed. Evolutionary progress. Univ. of Chicago Press, Chicago, IL.

[82] McAdams, D. P. (1989). Intimacy: The need to be close. Doubleday & Co.

[83] McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. Journal of personality and social psychology, 45(4): 851.

[84] McAdams, D. P., & Powers, J. (1981). Themes of intimacy in behavior and thought. Journal of Personality and Social Psychology, 40(3): 573.

[85] McAdams, D. P., Healy, S., & Krause, S. (1984). Social motives and patterns of friendship. Journal of Personality and Social Psychology, 47(4): 828.

[86] McClelland, D. C. (1985). How motives, skills, and values determine what people do. American Psychologist, 40(7): 812.

[87] McDowell, J. (1979). Virtue and Reason. Monist, 62: 331–350.

[88] McGuire, L., Rizzo, M. T., Killen, M., & Rutland, A. (2018). The development of intergroup resource allocation: The role of cooperative and competitive in-group norms. Developmental Psychology, 54(8): 1499–1506.

[89] Mikulincer, M., Shaver, P. R., Gillath, O., & Nitzberg, R. A. (2005). Attachment, caregiving, and altruism: boosting attachment security increases compassion and helping. Journal of personality and social psychology, 89(5): 817.

[90] Nisbett, R.E., and D. Cohen (2004). Culture of Honor: The Psychology of Violence in the South. Boulder, CO: Westview Press.

[91] Pang, J. S. (2010). The achievement motive: a review of theory and assessment of n Achievement, hope of success, and fear of failure. In O. Schultheiss & J. Brunstein (Eds.), Implicit motives (pp. 30–71). Oxford: Oxford University Press

[92] Parks CD, and AD Vu (1994). Social Dilemma Behavior of Individuals from Highly Individualist and Collectivist Cultures. Journal of Conflict Resolution, 38(4):708-718.

[93] Prendergast, C., & Stole, L. (2001). The non-monetary nature of gifts. European Economic Review, 45(10): 1793–1810.

[94] Preston, S.D. & de Wall, F.B.M. (2002). Empathy: its ultimate and proximate bases. Behavioral and Brain Sciences. 25 (1): 1–20, discussion 20–71.

[95] Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. Psychological review, 118(1): 57.

[96] Rawls, J. 1971. A Theory of Justice. Harvard Univ. Press, Cambridge, MA.

[97] Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. Review of General Psychology, 8(3): 179.

[98] Ricard, M. (2015), Altruism (trans. By C. Mandell and S. Gordon), New York: Little, Brown and Company.

[99] Richerson, P. J., & Boyd, R. (2005). Not by genes alone: How culture transformed human evolution. University of Chicago Press.

[100] Sandel, M. (2020). The Tyranny of Merit, London, Allen Lane.

[101] Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition and Emotion, 24(7): 1153–1172.

[102] Schultheiss, O. C., & Brunstein, J. C. (Eds.). (2010). Implicit motives. Oxford University Press.

[103] Schwartz, S.H. (1994). Are there universal aspects in the structure and contents of human values? Journal of Social Sciences, 50(4): 19–45.

[104] Selman (2008). Through Thick and Thin. Human Development, 51: 318–325.

[105] Shweder, R.A., and E. Bourne (1984). Does the Concept of the Person Vary Cross-Culturally? In R. Shweder and R. LeVine (eds), Culture Theory, Cambridge: Cambridge University Press, 158–199.

[106] Shweder, R. A., Much, N. C., Mahapatra, M., Park, L., Brandt, A., & Rozin, P. (1997). Morality and health. A. Brandt & P. Rozin (Eds.), 119–169.

[107] Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. American Economic Review, 97: 999–1012.

[108] Small, D. A., & Lerner, J. S. (2008). Emotional Policy: Personal Sadness and Anger Shape Judgments about a Welfare Case. Political Psychology, 29(2): 149–168.

[109] Smith, Adam (1759/1996). The Theory of Moral Sentiments. Sentry Press, New York. Reprint of 1853 ed., published by Henry G. Bohn, London, first published in 1759.

[110] Snower, D.J., and S. Bosworth (2016). Identity-Driven Cooperation versus Competition. American Economic Review, Papers and Proceedings, 2016. http://hdl.handle.net/10419/130402

[111] Sober, E., and D.S. Wilson (1998). Unto Others: The Evolution and Psychology of Unselfish Behavior, Cambridge, MA: Harvard University Press.

[112] Stemmler, G., Heldmann, M., Pauls, C. A., & Scherer, T. (2001). Constraints for emotion specificity in fear and anger: The context counts. Psychophysiology, 38(2): 275–291.

[113] Stiner, M.C., and S. Kuhn (2006). Changes in the 'Connectedness' and Resilience of Paleolithic Societies in Mediterranean Ecosystems. Human Ecology, 34(5): 693–712.

[114] Tadelis, S. (2007). The Power of Shame and the Rationality of Trust. Mimeo, Haas School of Business, U.C. Berkeley.

[115] Tangney, J.P, and R.L. Dearling (2002). Shame and Guilt, London: Guilford Press.

[116] Tedeschi, J., Smith, R., & Brown, R (1974). A reinterpretation of research on aggression. Psychological Bulletin, 81(9), 540–562.

[117] Tomasello, M. (2016). A Natural History of Human Morality. Cambridge, MA: Harvard University Press.

[118] Tomasello, M. (2018). Precis of a natural history of human morality. Philosophical Psychology, 31(5): 661–668.

[119] Tönnies, F. (2001/1887). Community and Civil Society (trans. By J. Harris and M. Hollis). Cambridge: Cambridge University Press.

[120] Triandis, H. (2002). Individualism-Collectivism and Personality. Journal of Personality, 69(6): 907–924.

[121] Triandis, H.C. (1995). Individualism and Collectivism. Boulder, CO: Westview.

[122] Turchin, P. (2006). War and Peace and War: The Life Cycles of Imperial Nations. Karlsruhe: Pi Press.

[123] Ullman, Timer D; Tenembaum, Joshua B (2020). Bayesian Models of Conceptual Development: Learning as Building Models of the World. Annual Review of Developmental Psychology. 2 (1): 553–558.

[124] Walker, E. L., & Heyns, R. W. (1962). An anatomy for conformity. Englewood Cliffs.

[125] Walzer, M. (1994). Thick and Thin: Moral Argument at Home and Abroad, Notre Dame, IN: University of Notre Dame Press.

[126] Wedekind C, Milinski M. (1996). Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus Generous Tit-for-Tat. Proc Natl Acad Sci U S A. 93(7): 2686–2689.

[127] Weinberger, J. Cotler, T. & Fishman, D. (2010). The duality of affiliative motivation. In O. Schultheiss & J. Brunstein (Eds.), Implicit motives. Oxford: Oxford University Press.

[128] Wilson, D. S. (2002). Evolution, morality and human potential. In Evolutionary psychology: Alternative approaches (pp. 55–70). Boston, MA: Springer US.

[129] Wilson, D.S. (2015), Does Altruism Exist? New Haven: Templeton Press.

[130] Winter, D. G. (1973). The power motive. New York: Free Press.

[131] Wollbrant, C.E., Knutsson, M. & Martinsson, P. (2022). Extrinsic rewards and crowding-out of prosocial behaviour. Nat Hum Behav 6: 774–781.