DISCUSSION PAPER SERIES

DP11131

COOPERATION, MOTIVATION AND SOCIAL BALANCE

Dennis J. Snower, Steven Bosworth and Tania Singer

PUBLIC ECONOMICS

ISSN 0265-8003

COOPERATION, MOTIVATION AND SOCIAL BALANCE

Dennis J. Snower, Steven Bosworth and Tania Singer

Discussion Paper 11131
Published 22 February 2016
Submitted 22 February 2016

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **PUBLIC ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Dennis J. Snower, Steven Bosworth and Tania Singer

COOPERATION, MOTIVATION AND SOCIAL BALANCE

Abstract

This paper examines the reflexive interplay between individual decisions and social forces to analyze the evolution of cooperation in the presence of "multi-directedness", whereby people's preferences depend on their psychological motives. People have access to multiple, discrete motives. Different motives may be activated by different social settings. Inter-individual differences in dispositional types affect the responsiveness of people's motives to their social settings. The evolution of these dispositional types is driven by changes in the frequencies of social settings. In this context, economic policies can influence economic decisions not merely by modifying incentives operating through given preferences, but also by influencing people's motives (thereby changing their preferences) and by changing the distribution of dispositional types in the population (thereby changing their motivational responsiveness to social settings).

JEL Classification: A13, C72, D01, D03, D62, D64

Keywords: motivation, Reflexivity, Cooperation, Social dilemma, Endogenous preferences, Dispositions

Dennis J. Snower - dennis.snower@ifw-kiel.de
Kiel Institute for the World Economy and CEPR

Steven Bosworth - steven.bosworth@ifw-kiel.de
Kiel Institute for the World Economy

Tania Singer - singer@cbs.mpg.de
Max Planck Institute for Human Cognitive and Brain Sciences

1 Introduction

This paper takes a new approach in exploring the social foundations of human cooperation. Building on a vast literature in motivation psychology[1], the individual is understood to have access to multiple, discrete motives, each of which is associated with a distinct objective. In short, people are recognized to be "multi-directed".

Our analysis shows how different motives may be activated by different social settings. Some social settings encourage prosocial motives; other discourage them. Changes in social settings may lead to changes in motives. Thus preferences are not located exclusively in the individual, but rather become the outcome of the interplay between the individual and her social environment.

The social settings are understood as well-defined and structured in advance of individuals' entry into them. For instance, just as a tennis match structures the relations between the players, tournament wage contracts and team remuneration schemes exert different motivational influences on employees. In the model below, social settings are specified in terms of the strategic complementarities or substitutabilities between agents. Since social settings can affect agents' motives, they influence their behavior not just via their beliefs and constraints, but also in terms of their objectives.

Our analysis provides new insights into the role of "social balance" in economic decision making. The balance between meeting the needs of the individual and the community is shown to arise from the interplay between social settings and personal traits. The greater the relative frequency of cooperative settings (displaying strategic complementarities) relative to competitive settings (characterized by strategic substitutabilities), the greater is the degree to which pro-social traits will thrive relative to selfish ones. Due to plasticity of traits, changes in social settings influence the composition of traits in the population, thereby further changing the social balance between pro-social and selfish behavior patterns. Such changes in social balance can have significant welfare implications, since people with pro-social traits obviously internalize some of the externalities in social dilemma situations, whereas those with selfish traits do not.

In our analysis, cooperation among economic agents is not merely generated by economic synergies among self-interested agents with unique preferences (as in the gains from trade brought about by Adam Smith's invisible hand). Instead, it arises from people's motivated decisions in different social settings, some of which may be more conducive to prosociality than others. Since agents

The authors would like to acknowledge support from the Institute for New Economic Thinking under grant INO13-00036. We would also like to thank George Akerlof, Rachel Kranton, Paul Collier, Jean-Paul Carvalho and Robert Akerlof for feedback, as well as seminar participants at the Kiel Institute for the World Economy, the Max Planck Institute for Evolutionary Biology, the University of Pittsburgh, and attendees at the American Economic Association 2015 Annual Meeting, the SBRCR 2015 Workshop, the 3rd TILEC Economic Governance Workshop, and the WZB Interdisciplinary Perspectives on Decision Making Workshop.

[1] Heckhausen and Heckhausen (2010) provide an excellent survey.

are multi-directed, they do not have unique preferences and are thus not consistently self-interested or consistently altruistic. Instead, their objectives depend on the interplay between their individual traits and their social settings. In this context, creating a more cooperative society involves creating not just new economic synergies, but also social settings that elicit more cooperative motives and a greater frequency of altruistic dispositional types (by affecting the payoffs from the social settings).

Our analysis can help shed light on why individual behaviors in social dilemma situations often contradict the predictions of economic theory. In particular, people frequently cooperate in the absence of compensation, but to varying degrees depending on their individual traits and their social settings. We know for example that people engage in substantial philanthropic activity, voluntary work, and social activism. We also know that there is a wide heterogeneity of cooperative tendencies across people. Some people have a greater tendency to cooperate in social dilemmas than others, though individual-level behavior can, but need not fluctuate (Van Lange et al., 1997; Volk et al., 2012). In terms of our model, people's willingness to cooperate under some social settings, but not others, can be accounted for by changes in their motives in response to these settings. Under settings featuring strategic complementarities, they may be willing to cooperate even in the absence of compensation.

This analytical context provides a broader framework for policy analysis than conventional theory permits and thereby sheds light on new opportunities for policies to affect cooperation among economic agents – opportunities that are generally ignored in mainstream neoclassical analysis. Whereas neoclassical economic theory focuses on the effect of policies on economic incentives, specified with regard to given preferences, our model also shows how policies can influence people's social settings, which changes their motives and thereby alters their preferences. Over the longer run, policies can also affect the payoffs from these social settings, which may change the distribution of dispositional types and thereby alter the responsiveness of motives to social settings.

We claim that traditional economic theory overlooks significant sources of human economic cooperation by restricting its policy purview to incentives for self-interested agents with exogenous preferences. Beyond that, its policy prescriptions may even be counter-productive in some circumstances, since its proposed policies may crowd out pro-social motives (for example, Frey and Jegen, 2001; Gneezy et al., 2011; Bowles and Polanía-Reyes, 2012)[2]. Our analysis shows how policies may crowd out motivation to cooperate when they increase the degree to which situations discourage cooperation or increase the prevalence of such settings. By contrast, policies crowd in cooperation when they provide state- or trait-based support for cooperative motives.

This extension of policy analysis beyond traditional monetary incentives

[2]Two conflicting examples may be found in Gneezy and Rustichini (2000) and Gelcich et al. (2013). Gneezy and Rustichini find that volunteering goes down when small payments are introduced for collected donations, while Gelcich et al. find that the introduction of sanctions that should be small enough *not* to deter exploitation of a common pool resource do in fact deter exploitation.

has far-reaching implications. In mainstream economic theory, the purpose of economic policy is primarily to correct for externalities and inequalities. Externalities (social costs and benefits for which individuals receive no private compensation) are internalized by providing the appropriate compensation to self-regarding individuals. By contrast, in our analysis the degree to which externalities are internalized depends crucially on which motives are activated: what must be compensated under a selfish motive may become implicitly internalized under a pro-social motive. By examining how different social settings can activate different motives, our analysis paves the way toward an understanding of how economic decisions may be shaped not just be monetary incentives, but also by the social organization within which economic interactions take place. Thereby our analysis can provide a rationale for the influence of framing, choice architecture and nudging on economic activities, at least with regard to the social aspects of these phenomena.

Thus policies can affect economic behavior not just by influencing payoffs for a given set of preferences, but also by changing these preferences. These preference changes are associated with motivational changes, which arise from changes in policy-induced payoffs and policy-induced social settings, as well as changes in dispositional types arising from longer-term policy-induced payoffs to each type.

The rest of the paper is structured as follows. Section 2 summarizes the salient insights of the paper, providing the intuitions underlying the model that follows. Section 3 describes the motives and the social settings in which agents interact with one another. Section 4 analyzes how dispositional types are shaped by social settings. Section 5 examines the resulting "social topography," characterizing the equilibrium distribution of the dispositional types. Section 6 explores the effects of policies designed to increase cooperation. Finally Section 7 concludes.

2 Basic Insights

We begin by defining the three basic concepts of our analysis: motives, social settings and dispositional types. A *motive*, in the sense that the word is commonly used in motivation psychology, is a force that gives direction and energy to one's behavior, thereby determining the objective of the behavior, as well as its intensity and persistence[3]. The psychology literature has identified a number of different motives, such as the achievement,[4] affiliation[5] and power[6] motives. In the model here, however, we focus on two motives: a Self-interested Wanting[7]

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

[4]See for example Atkinson and Feather (1966); Pang (2010).

[5]McClelland (1967), H. Heckhausen (1989), or Heckhausen and Heckhausen (2010).

[6]For example, H. Heckhausen (1989); J. Heckhausen (2000); Heckhausen and Heckhausen (2010).

[7]This motivation system – the closest, though imperfect, match for the standard economic assumption of self-interest – does not receive much attention in the motivation psychology literature. See for example McDougall's (1932) propensity for foraging and ownership and

motive (aimed at maximizing one's own personal payoff) and a Caring[8] motive (aimed at maximizing joint payoffs).

In the model below, the Self-interested Wanting motive is represented by the utility function of the standard selfish, rational agent of neoclassical microeconomics. The Caring motive resembles the utility function of what is usually termed a 'pure altruist' by behavioral economists (see e.g. Andreoni, 1990). The paradigmatic example of this motive is that which drives a mother to care for her infant. In the evolutionary process, the Caring motive induces parents to support their offspring over the vulnerable childhood stage. It also carries over to helpful, compassionate behavior of more distant human kin and non-kin. Its evolutionary function is clearly to promote cooperation necessary for survival. Psychologically, this system is associated with feelings of affection, compassion, nurturance, friendliness and warmth related to prosocial goals,[9] particularly concern for the wellbeing of others.[10]

A *social setting* is a joint activity of several people producing a distribution of payoffs. In our model, there are just two social settings: a *Cooperative* setting (with complementary actions) and a *Competitive* setting (with substitutable actions). In a Cooperative setting, one person's contribution increases the productivity (payoff per unit of contribution) of another person involved in these activities. It includes activities such as household production, collaboration in teams, and common goods with increasing marginal benefits. In a *Competitive* setting, one participant's contribution reduces the productivity of another participant. The activities in this setting include tournament contracts in labor markets, striving for social position and other contests, as well as common goods with diminishing marginal benefits[11].

We will show that, from the perspective on an individual participant, the Self-interested Wanting motive is relatively well-suited to the Competitive setting, whereas the Caring motive is relatively well-suited to the Cooperative setting. In the Cooperative setting, one individual's cooperation implies that the best response of others is to become more Cooperative as well. On this account, people's actions are strategic complements. In the Competitive setting, one individual's cooperation induces more selfish behavior as a best response. Thereby people's actions become strategic substitutes.[12]

In the model below, social settings describe the contexts in which people make their decisions, and these contexts are shown to affect their motives. For this purpose, it is convenient to restrict our analysis to social settings that

Reiss' (2004) desire for eating, and Gilbert's (2013) seeking drive, an acquisition focused system.

[8]This motive is concerned with nurturance, compassion, and care-giving, e.g. Weinberger et al., (2010). The caring motive is often distinguished from the affiliation motive, e.g. McDougall (1932), Murray's (1938), McAdams (1980), H. Heckhausen (1989), and J. Heckhausen (2000).

[9]See for example McAdams and Powers (1981); Weinberger et al. (2010).

[10]McAdams et al. (1984).

[11]See Section 3.1 for an explanation and further specific examples.

[12]The characterization of social settings in terms of strategic complementarities or substitutabilities is in the spirit of (Bulow et al., 1985).

are exogenously given by existing institutions and customs, e.g. existing laws, cultures, norms, organizational structures, etc. In practice, social settings – in terms of opportunities for cooperation or competition – are shaped not only by existing institutions and customs, but also by the participants, e.g. people pursuing careers in social work or banking. In our model, a change in the available social settings may induce a change in people's motivations and thereby in their behavior patterns.

A disposition is a trait that can be interpreted as a crystalized motive. In the personality psychology literature, it is widely accepted that dispositions and traits are crystalized states (e.g. Fleeson, 2001). Specifically, if a particular state (a motive, in our model) is activated repeatedly and persistently, then this state becomes crystalized into a persistent disposition to act in accordance with this state. In practice, people are driven by various different motives, each which can be conceived as corresponding to different dispositional types. In the personality psychology literature, individual dispositional differences are usually represented in terms of distributions with respect to each disposition. For example, anxious personality traits are conceived as being normally distributed with the majority of people showing average degrees of the anxious trait and only minorities showing very high or very low levels of anxious trait.

For each of the two motives highlighted in our model – Self-interested Wanting and Caring – dispositions may crystalize, so that individuals may come to differ in terms of their propensity to Want (i.e. high on the Wanting scale are those with strong appetitive consumption-oriented drives, whereas depressed people, for example, are low on this scale) and their propensity to Care (i.e. high on the Caring scale are those who are altruistic, having a strong propensity to benefit other people; whereas narcissistic people are low in this scale, for example). Different individuals may be conceived as being distributed along the distributions for these two dispositions. In practice, virtually no one is positioned at the extreme ends of the Wanting and Caring scales, and if people show extremes on these dispositional scales, they are often classified according to pathological criteria. Furthermore, virtually everyone is able to modulate their degree of Wanting and Caring in response to their environment. In other words, virtually everyone is "multi-directed, in the sense that humans have multiple motives that can be differentially activated in different settings,[13] but they differ in terms of their positions in the dispositional distributions and thus also in their propensities to respond to particular external stimuli with particular motives.[14]

[13]Our conception of multi-directedness goes beyond framing effects, since the former links contextual stimuli to decision-making objectives, whereas the latter refers to context-driven cognitive biases. Multi-directed actions also differ from cue-driven ones (e.g. Laibson (2001)), since the former arises from multiple motives whereas the latter arise from shifts in perception and attention. In the seminal paper by Bernheim and Rangel (2004) on addiction, environmental cues affect behavior whenever individuals are in a "hot" mode, in which an individual chooses to consume an additive substance irrespective of her underlying preferences; whereas in our model the environment helps shape these preferences.

[14]In the psychology literature, social settings are known to affect the activation of motives both directly (a change in setting leading directly to a change in motive) and indirectly,

For the sake of analytic simplicity, however, we consider only three types of people, with the following special patterns of dispositions: (1) a Selfish type, representing the extreme high end of the Self-Interested Wanting dispositional distribution and the extreme low end of the Caring dispositional distribution, (2) a Caring type, representing the extreme high end of the Caring dispositional distribution and the extreme low end of the Self-Interested Wanting dispositional distribution, and (3) a Responsive type, representing the mean of the Caring dispositional distribution and the mean of the Self-Interested Wanting dispositional distribution.

By implication, the Self-interested type leads to the activation of the Self-interested Wanting motive in all social settings; the Caring type leads to the activation of the Caring motive in all settings; and the Responsive type leads to the activation of the Caring motive in the Cooperative settings and the Selfish motive in the Competitive settings, but at a cost. Our categorization serves merely as an extreme analytical simplification of individual dispositional differences, whereby some people are predominantly selfish in most social settings and others are predominantly caring in these settings, whereas others adjust their behavior more responsively to the settings they encounter.

The simplifications of our model are meant to build a bridge between economic theory and motivation and personality psychology. In neoclassical and behavioral economics, people are "single-directed" (possessing only one objective, given by their utility function). People of the Selfish type are functionally equivalent to the single-directed, selfish agents in standard neoclassical economic analysis. People of the Caring type are akin to the single-directed individuals with altruistic preferences in behavioral economics. We introduce multi-directedness into economic theory through people of the Responsive type, whose motives respond to their social settings (a phenomenon widely ignored in mainstream economics,[15] but accepted as ubiquitous in motivation and personality psychology).

The individuals in our model face stochastic social settings with idiosyncratic frequencies, in the sense that each individual encounters Competitive and Cooperative settings with the same probabilities, independently distributed across individuals and time. These probabilities determine the fitness of the Selfish, Caring and Responsive types, since the different types have different comparative advantages under different social settings. The greater the frequency of the Cooperative setting (relative to the Competitive one), the greater the comparative advantage (and thus the expected payoff) of the Caring type; whereas the greater the frequency of the Competitive setting, the greater the expected payoff of the Selfish type. Changes in the frequency of social settings will change the

since repeated exposure to a particular social setting, associated with repeated activation of a particular motive, will induce a change in dispositional type via plasticity of traits. In the model below, we focus on the direct channel.

[15] A limited exception is behavioral economists' explantation of framing effects and sensitivity to context. These are generally rationalized as arising from (consistently) conditionally cooperative preferences (Fischbacher et al., 2001) and exogenous shifts in beliefs (see e.g. Dufwenberg et al., 2011; Fosgaard et al., 2014).

comparative advantages of these types, which are assumed to lead to changes in the numbers of these types.

We define the *social topography* as the mapping of different agent types into different social settings. The social topography summarizes the overall context in which types are matched with social settings, thereby determining people's contributions and payoffs from their social settings.

Once again, these analytical simplifications regarding the interconnections among motives, dispositional types and social settings are meant to capture essential features of a more complicated reality. As the relevant literatures in motivational and social psychology, affective neuroscience and evolutionary biology show,[16] humans have access to far more than two motivations; there exist far more than two social settings and these settings may be distinguished through features other than just actions that are strategic substitutes and complements; and it is useful to distinguish among more than three types of people with different patterns of dispositional traits to understand individual differences in motivational responsiveness to changes in social settings.

Our analysis encompasses the following basic insights, each of which is well-known in other disciplines (particularly psychology, sociology, anthropology, biology and neuroscience), though commonly overlooked in economics thus far.

Insight 1: All behavior is motivated, driven by multiple, discrete motives.

Knowing a person's prior choices and constraints is not sufficient to determine the person's behavior, as it is under revealed preference theory. In line with the vast motivation psychology literature, we recognize that people generally have access to multiple, discrete motives. Each motive is associated with a different objective function for economic decisions.

Since our analysis explores the social opportunities for cooperative versus competitive decision making, it is useful to contrast this motivational approach with the notion of preferences. While behavioral economics has extended the standard neoclassical analysis through consideration of social preferences[17] grounded in experimental evidence, what behavioral and neoclassical economics have in common is the assumption that individual preferences are internally consistent, reasonably stable[18] and context-independent[19]. The novel contribution of our analysis lies in the recognition that the objectives underlying an individual's economic decisions depend on the individual's active motives and

[16] A partial overview is provided in Przyrembel et al. (mimeo).

[17] See, for example Loewenstein et al. (1989), Andreoni (1990), Rabin (1993), Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Battigalli and Dufwenberg (2007), or Cox et al. (2007).

[18] Exceptions are made for variations in behavior due to non-systematic random mistakes (e.g. McKelvey and Palfrey, 1995; 1998).

[19] Framing effects and sensitivity to context are generally rationalized by behavioral economists as arising from conditionally cooperative preferences (Fischbacher et al., 2001) and exogenous shifts in beliefs (see e.g. Dufwenberg et al., 2011; Fosgaard et al., 2014).

that these motives may be affected by the individual's social setting, which may change predictably and abruptly. In this context, preferences need not be internally consistent and temporally stable across motives.

While we restrict our analysis to two motives – Self-interested Wanting and Caring – it is important to emphasize that Care is obviously not the only prosocial motive. (Another important one for example is affiliation, i.e. the motive to belong to a social group and adhere to its norms[20].)

Insight 2: Motives are influenced by social settings.

Broadly conceived, an individual's social setting is meant to represent the entire structure of interpersonal relationships and interactions into which the individual may enter. The content of these relationships and interactions may be well-defined and structured in advance of individuals' entry into them.

Different social settings give rise to different motives. This pattern is reinforced by the fact that different motives are differentially well-suited to different social settings in terms of the outcome they generate for the decision maker. In our model, as noted, we consider Competitive settings (in which people's actions are strategic substitutes) and Cooperative settings (in which people's actions are strategic complements). We will show that the Self-interested Wanting motive is relatively well suited for Competitive settings, whereas the Caring motive is relatively well suited for Cooperative settings[21].

Insight 3: The responsiveness of motives to social settings depends on people's patterns of dispositional traits.

Inter-individual personality differences may be characterized as propensities to respond more or less to particular types of incentives (Schultheiss et al., 2010). These dispositional traits can be genetically predetermined or develop through very early life experiences (McClelland, 1965; 1985).[22] Strong propensity to activate particular decisions is frequently considered a personality trait (regarding prosocial behavior see McClintock and Allison, 1989; Van Lange et al., 1997; Balliet et al., 2009).

In our model, motives can be activated by a person's social settings and the degree to which this happens is assumed to depend on the person's dispo-

[20]This motive, which implicitly plays a major role in identity economics (Akerlof and Kranton, 2000; 2010), could be included straightforwardly in our analysis but, for brevity, we do not do so here. See also Holländer (1990), Bernheim (1994), Sugden (2000), Brekke et al. (2003), and Bénabou and Tirole (2006) for additional models of affiliation to groups in economics.

[21]The idea that social preferences flourish in environments with strategic complements and are discouraged in environments with strategic substitutes has been discussed extensively in the literature studying the evolution of preferences (see e.g. Rotemberg, 1994; Bester and Güth, 1998; Alger and Weibull, 2012). Experimental evidence that strategic complements encourage more cooperation than strategic substitutes may be found in Suetens and Potters (2007) and Potters and Suetens (2009).

[22]According to the American Psychological Association, "personality refers to individual differences in characteristic patterns of thinking, feeling and behaving (Kazdin, 2000).

sitional type.[23] In practice, other aspects of the person's internal and external environment can also activate motives. There is also much evidence that motives depend on interactions between situations and personality characteristics (see Mischel and Shoda, 1995; Roberts and Pomerantz, 2004).

As noted, our model distinguishes among three individual types: a Selfish type (inducing an individual to behave selfishly in all settings), a Caring type (inducing the individual to show concern for the interests of others in all settings, and a Responsive type (inducing the individual to show concern for the interests of others when engaged in settings entailing Cooperative interactions and to behave selfishly when engaged in settings involving Competitive interactions.) These distinctions are merely an analytical simplification of the common observation that people differ in terms of their dispositional traits and thus have different motives under any given social setting.

Insight 4: The frequencies of social contexts and their associated payoffs affect the evolution of individual types.

In our model, different individual types are differentially suited for different social settings, depending on the payoffs to these types from these settings. It is in this sense that the fitness of individual types depends on the frequency of the social settings to which people are exposed.

We will show that the Selfish type is best-suited for people who encounter predominantly Competitive settings, but poorly suited for the those who encounter more Cooperative settings. The Caring type will be shown to be best-suited for people who encounter primarily Cooperative settings, but poorly suited for those who encounter more Competitive settings. Finally, the Responsive type is shown to be best for those who encounter a more balanced mix of both Cooperative and Competitive settings.

When the frequencies of social settings change, the relative fitness of the various individual types changes as well. The fitness of these types is determined by the payoffs from the resulting motives in a given mix of social settings. Therefore changes in the frequencies of social settings affect the relative payoffs to the various types. The higher the expected payoff from a particular type, the greater are the chances that this type will be reinforced. We assume that individuals develop dispositional types that yield the highest expected payoff. This stands in for a process of social evolution or personality development that may take place over substantial periods of time, such as a lifespan or across generations.[24]

[23] For an overview of how a person's behavior depends on both their personality and the situation (for a review see Heckhausen and Heckhausen, 2010, chapter 4).

[24] Developmental psychologists have characterized personality traits as stable across settings, but plastic over the longer time scales of people's lives (Baltes, 1987). Children are particularly influenced by the environment in which they develop (Bronfenbrenner, 1979).

Insight 5: Preferences are the outcome of a reflexive interplay between individual actions and social forces.

Social settings influence people's motives, which determine the objectives of their decisions (i.e. their preferences, in terms of decision utilities). These decisions determine the payoffs from the social settings. People's types influence the motivational responsiveness to social settings; the resulting payoffs affect the evolution of individual types.

In short, preferences are not located exclusively in the individual, but rather emerge through the interaction between social settings, dispositional types and motives. This phenomenon constitutes a fundamental form of reflexivity: social settings provide a macro-foundation for microeconomic decisions, and these decisions provide a micro-foundation for the payoffs from the social settings.

Specifically, the social topography (the mapping of agents with different dispositional types into different social settings) depends on agents' relative payoffs from the Competitive versus Cooperative settings and their likelihoods of encountering each of these settings. However, since these payoffs depend on peoples' motives, whose activation arises out of an interplay between the social setting and their dispositional types, the social topography feeds back into the micro-level through reinforcement of different dispositional types. The population-level distribution of dispositional types gravitates towards an equilibrium in which people's motives and the broader social topography are co-determined.

Insofar as an individual's motives are influenced by both her social settings, preferences are not located exclusively in the individual, as assumed in traditional economic theory (where the preference mapping is, as it were, hard-wired in the person's brain)[25]. In practice, the endogenous relationships between preferences and environment can arise not only from social settings, but also cultural norms and socialization, change-oriented social movements, or any number of other forces.

Insight 6: Policies may influence the evolution of cooperation in society by affecting the reflexive interplay between individual decisions and social forces.

Our analysis stands in sharp contrast to mainstream economic theory, according to which humans are assumed to be self-interested individuals, with rational preferences that are internally consistent, temporally stable and context-independent. Under these standard assumptions, humans cooperate only in the presence of economic incentives. Problems such as the provision of public goods and common resources are associated with deficient incentives to take into account how one's decisions affect others. The desired interventions – such as taxes and subsidies, regulations, and redefinitions of property rights – are

[25]For a survey of endogenous preferences in economics, see Bowles (1998). For a defense of exogenous preferences in economics, see Stigler and Becker (1977).

understood as compensating people for their effects on others or forcing people to act as if they were compensated.

We claim that in restricting its policy purview to incentives for self-interested, rational agents with exogenous preferences, traditional economic theory overlooks significant sources of human cooperation. Beyond that, its policy prescriptions may even be counter-productive in some circumstances, since the proposed policies may crowd out pro-social motives. Our analysis addresses these deficiencies by investigating the influence of different social settings on the activation of different motives. It thereby uncovers a much wider domain of policies than those identified in mainstream economic theory to promote human cooperation in social dilemma situations.

In particular, policies can improve welfare not just by modifying agents' incentives under a given set of preferences, but also by affecting the preferences themselves by influencing their motives. Policies can shape people's motives either directly (by influencing the relative payoffs from different motives) or indirectly through the social settings in which people participate. Such policies could include institutional, political, economic and social incentives, such as corporate culture, regulations, payment schemes, and social norms. Over the long run, our analysis indicates another channel whereby policy can affect welfare, namely, through influence on dispositional traits. Changes in the frequencies of and payoffs from social settings affect the relative expected payoffs from different dispositional types and thereby influence the prevalence of these dispositional types. In turn, changes in the distribution of dispositional types over the population affect the responsiveness of motives to social settings. The resulting changes in motives affect the contributions to and payoffs from the social settings, thereby affecting agents' wellbeing.

In general, people's willingness to cooperate in the presence of social dilemmas may be influenced both by the standard pecuniary incentives of mainstream economics and by the motivation-shaping policies above. Whether pecuniary incentives complement or crowd out pro-social motivation – and hence enhance or reduce the effectiveness of motivation-shaping policies – depends on the social setting and the composition of dispositional types in the population. In particular, when pecuniary incentives based on individual performance are used to elicit pro-social behavior in the settings where strategic complements apply, these incentives can work against people's intrinsic motivation to cooperate. Because these policies help the selfishly motivated more than they help those with pro-social motives, they reinforce the prevalence of dispositional types that predispose people to selfishness. The desirability of such policies is therefore ambiguous. Our results provide structural foundations for what Bowles (1998, 2008) documents as the tendency for markets and complete contracts to crowd out voluntarily cooperative behavior and social exchange.

Where incentives are used to support pro-social behavior in settings where strategic substitutes apply however, they bolster people's motivation to consider others' wellbeing. In these settings incentives deter exploitation of cooperative people by the selfish, and therefore relatively advantage those with pro-social motives. This reinforces the prevalence of dispositional types predisposing peo-

ple to pro-sociality and amplifies the effect of the pecuniary incentives.

These considerations are relevant for the policy approaches to a wide variety of economic problems, ranging from climate change, pollution and biodiversity loss to insufficient rates of vaccination, organ and blood donation, to efforts to eradicate poverty. In all these cases, individual contributions to public goods or poverty reduction may depend significantly on their underlying motives and these may be shaped by the social settings in which they are embedded.

This framework of thought is summarized in Figure 1. Social settings create interpersonal relations that generate complementarities or substitutabilities in the participants' actions. These settings affect each participant's motives. The motivational response to the social settings is modulated through each participant's dispositional type. The motives generate behavior patterns that generate payoffs to the participants. These payoffs, in turn, affect the development of the participants' dispositional types.[26].

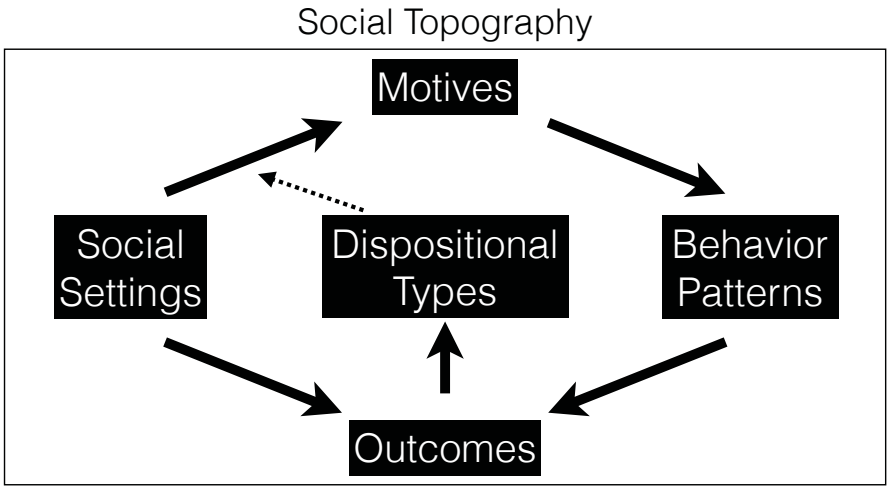

Figure 1: Conceptual framework

3 Motives and Social Settings

Our model has the following building blocks:

[26]It is reasonable to expect that the distribution of social settings may endogenously depend on the distribution of dispositional types in the population. For example a society comprised entirely of those with caring dispositional types may be rather disinterested competitive settings. Our model abstracts from this channel to focus on the causal effect that the distribution of various settings has on motives and dispositional types.

- Two motives: Self-interested Wanting (s) and Caring (c): Under Self-interested Wanting, an agent is concerned only with her own direct payoff; whereas under the Caring motive, the agent is also concerned with the direct payoff of another agent (at least to some degree).

- Two kinds of social setting: Cooperative (C) and Competitive (K). In the Cooperative setting, an agent's direct payoff depends positively on another agent's contribution to the social interaction. In the Competitive setting, an agent's direct payoff depends negatively on another agent's contribution to the social interaction. For simplicity, we assume that these social interactions are dyadic, i.e. an agent i interacts with another agent j.

- Each agent i encounters a Cooperative setting with probability $\eta_i \in [0, 1]$ and encounters a Competitive setting with probability $1 - \eta_i$. This probability is idiosyncratic and distributed in the population according to the commonly known cumulative distribution function $H(\cdot)$.

- Three kinds of agent: Selfish (s), Caring (c) and Responsive (r) agents. Selfish agents always pursue the Self-interested Wanting motive; Caring agents always pursue the Caring motive; and Responsive agents pursue the motive that is most appropriate to the setting (in terms of payoff).[27]

- We normalize the number of all agents in the population to be 1. Define the frequencies of Caring, Selfish and Responsive agents as n_c, n_s, and n_r, respectively:[28]

To begin with, we will focus exclusively on the Selfish and Caring agents. Responsive agents will be considered later.

3.1 Distinguishing Social Settings Through Strategic Complements and Substitutes

Whether or not one's cooperation makes someone else want to cooperate is essential to understanding the strategic considerations and resulting patterns of cooperation in social dilemmas. Any social dilemma may be characterized as having strategic complements, substitutes, or neither – and the well-known social dilemmas are easily classified within this framework. Social dilemmas exhibiting strategic complementarities include collaboration in teams and common goods with increasing marginal utility. In contrast, contests and common goods with diminishing marginal utility exhibit strategic substitutes.

Collaboration in teams often involves workers whose inputs are complementary. For example suppose two academics, one with theoretical and another with

[27] Since there is a one-to-one mapping between the motives of Self-interested Wanting and Care (on the one hand) and the Selfish and Caring dispositional types (on the other), these motives and dispositional types can both be associated with the same descriptors: s and c, respectively.

[28] All agents in the population must be one of these three types.

empirical expertise, decide to collaborate on an article. If the theoretician puts in low effort, then the empiricist's time may be better spent on other projects, but if the theoretician puts in high effort then the empiricist's effort becomes much more productive since a high-impact publication may require both a high-quality theoretical and empirical contribution. Cooperation in teams motivates Rotemberg (1994)'s analysis of workplace relations. He shows that altruism may be a rational way to "commit" oneself to high effort in these circumstances. Another example is open-source software. Individual programmers work on pieces of code that are then contributed to larger projects. Any one person's code may be useless without the code that others have contributed.

The other prominent instance of cooperation problems with strategic complementarities involves common (non-excludable) goods with increasing marginal utility. In these cases contributing to the common good is most efficient when a high level of provision is expected[29]. When the anticipated level of the common good is low, contributing may be very unproductive. Examples include common pool resources near their level of extinction or public goods such as transportation (one road may be rather useless without a larger network) or public broadcasting (only well-produced programs may be worth supporting).

Contests are a paradigmatic example of a social dilemma with strategic substitutes. These can include auctions and lotteries (which may be used to finance public goods)[30] and tournaments (which may be used in labor contracts, see Lazear and Rosen, 1981). The defining feature of a contest is that many people will engage in costly effort or payment and the one(s) who take the most costly action will obtain the best (expected) distribution of resources or status. Because reducing one's effort in a contest increases the likelihood that others win (holding their effort constant), this means that the expected marginal utility of others' effort is diminishing in own effort (see Dubey et al., 2006 for a formal proof of this result).

The other major type of social dilemma with strategic substitutes are common goods with diminishing marginal utility. Since contributions are less efficient at high levels of provision, people will want to reduce their contributions when they think that others are providing the common good. This is most starkly illustrated in a volunteer's dilemma. When only one person is required to stop and help a stranded motorist, people will tend to drive on if they think that someone else will help, and stop if they think that no one will. Traditional air pollutants also display this property for example, since low concentrations of pollutants may be relatively harmless.

3.2 Direct Payoffs and Utilities

Define x_i and x_j as the contributions of agents i and j, respectively, to the social interaction in the Cooperative setting. Let agent i's direct payoff from

[29] Harstad and Liski (2013) illustrate formally how optimal contributions to common goods depend on the slope of their marginal utility.

[30] See Morgan (2000) for lotteries and Goeree et al. (2005) for auctions.

the interaction with agent j in the Cooperative setting be

$$U_{ij} \equiv x_i x_j + a_C x_i + b_C x_j - \frac{d_C}{2} x_i^2 \tag{1}$$

where a_C, $b_C > 0$ and $d_C > 2$ are constants.

Define y_i and y_j as the contributions of agents i and j, respectively, to the social interaction in the Competitive setting. Furthermore, let agent i's direct payoff to the interaction with agent j in the Competitive setting be

$$V_{ij} \equiv -y_i y_j + a_K y_i - b_K y_j - \frac{d_K}{2} y_i^2 \tag{2}$$

where a_K, $b_K > 0$ and $d_K > 2$ are constants.

Under the motive of Self-Interested Wanting (s), agent i's utility from each of the two social settings is equal to the direct payoffs from these settings (U_{ij}^s and V_{ij}^s, respectively):

$$U_{ij}^s = U_{ij}$$
$$V_{ij}^s = V_{ij}.$$

But under the motive of Caring (c), agent i's utility from each of the social settings is equal to a convex combination of the direct payoffs to agents i and j:

$$U_{ij}^c = (1-\kappa)\, U_{ij} + \kappa U_{ji}$$
$$V_{ij}^c = (1-\kappa)\, V_{ij} + \kappa V_{ji}$$

where κ is a positive constant, $0 < \kappa \leq 1/2$. At one extreme, $\kappa = 0$ represents pure Self-interest, whereas $\kappa = 1/2$ represents "Perfect Care" (i.e. one's own payoff is weighted equally with the payoff of one's partner).

3.3 Equilibrium Contributions and Equilibrium Payoffs

To find agent i's utility-maximizing contribution to the social interaction in the Cooperative setting under the Self-interested Wanting motive (s), we maximize the utility $U_{ij}^s = U_{ij}$ with respect x_{ij}:

$$x_i^s = \frac{a_C + x_j}{d_C}. \tag{3}$$

Similarly, agent i's optimal contribution to the interaction in the Competitive setting under the Self-interested Wanting motive is the contribution x_{ij} which maximizes the utility $V_{ij}^s = V_{ij}$:

$$y_i^s = \frac{a_K - y_j}{d_K}. \tag{4}$$

With regard to the Caring motive (c), agent i's optimal contribution x_{ij} to the interaction in the Cooperative setting maximizes the utility $U_{ij}^c = (1-\kappa)\, U_{ij} + \kappa U_{ji}$:

$$x_i^c = \frac{(1-\kappa)\, a_C + \kappa b_C + x_j}{(1-\kappa)\, d_C} \tag{5}$$

and agent i's optimal contribution to the interaction in the Competitive setting maximizes the utility $V_{ij}^c = (1-\kappa)\, V_{ij} + \kappa V_{ji}$:

$$y^c = \frac{(1-\kappa)\, a_K - \kappa b_K - y_j}{(1-\kappa)\, d_K}. \tag{6}$$

Note that each contribution of agent i depends on the contribution made by agent j. Of course, agent j is in the same position as agent i, and thus the equilibrium contributions may be derived once we know what the motives of these two agents are. In the social settings above, the Caring agent is assumed to pursue only the Caring motive, while the Selfish agent is assumed to pursue only the Self-interested Wanting motive. Thus, for example, when two Selfish agents are paired in the Cooperative setting, each agent contributes the amount in equation (3), where $x_i^s = x_j^s$. Substituting $x_i^s = x_j^s$ into equation (3), we obtain[31]

$$x_i^s = \frac{a_C}{d_C - 1}.$$

Performing such calculations for all possible pairings in both types of social setting, we obtain the contributions in Table 1. Table 2 shows the resulting payoffs from each pairing of motives.

		Other's motive	
		Self-interest (s) *motive*	*Caring* (c) *motive*
		Cooperative (C) *setting*	
	s	$x^{ss} = \frac{a_C}{d_C-1}$	$x^{sc} = \frac{(1-\kappa)(d_C+1)a_C+\kappa b_C}{(1-\kappa)d_C^2-1}$
	c	$x^{cs} = \frac{((1-\kappa)d_C+1)a_C+\kappa b_C d_C}{(1-\kappa)d_C^2-1}$	$x^{cc} = \frac{(1-\kappa)a_C+\kappa b_C}{(1-\kappa)d_C-1}$
Own motive			
		Competitive (K) *setting*	
	s	$y^{ss} = \frac{a_K}{d_K+1}$	$y^{sc} = \frac{(1-\kappa)(d_K-1)a_K+\kappa b_K}{(1-\kappa)d_K^2-1}$
	c	$y^{cs} = \frac{((1-\kappa)d_K-1)a_K-\kappa b_K d_K}{(1-\kappa)d_K^2-1}$	$y^{cc} = \frac{(1-\kappa)a_K-\kappa b_K}{(1-\kappa)d_K+1}$

Table 1: Agents' contributions to the interactions in each social setting

Under the assumptions that

$$0 < \kappa < \frac{2\left(d_C^2 - 1\right)}{2\left(d_C^2 - d_C\right) + d_C^3}, \tag{7}$$

[31]Note that throughout we assume that agents can observe each other's motives. While this assumption is admittedly unrealistic, Frank (1988) and Guttmann (2013) among others have argued that people frequently make superficial but informative judgments about the motives of other people. Assuming that types are only partially observable does not change the qualitative result that social preferences can be beneficial in settings where strategic complements apply (Bester and Güth, 1998; Guttmann, 2013). The ongoing nature of social interactions and people's ability to build reputations makes these assumptions closer to being satisfied.

		Other's motive	
		Self-interested Wanting (s) *motive*	*Caring* (c) *motive*
		Cooperative (C) *setting*	
Own motive	s	$U^{ss} = \frac{a_C(a_C d_C + 2b_C(d_C-1))}{2(d_C-1)^2}$	$U^{sc} = U^{ss} + \frac{\kappa d_C\left(2(1-\kappa)d_C^2+\kappa-2\right)(a_C+b_C(d_C-1))^2}{2(d_C-1)^2\left((1-\kappa 1)d_C^2-1\right)^2}$
	c	$U^{cs} = U^{cc} - \frac{\kappa\left(d_C\left(2(1-\kappa)^3 d_C^2-2\kappa(1-\kappa)d_C+3\kappa-2\right)+2\kappa\right)(a_C+b_C(d_C-1))^2}{2((1-\kappa)d_C-1)^2\left((1-\kappa)d_C^2-1\right)^2}$	$U^{cc} = \frac{((1-\kappa)a_C+\kappa b_C)(a_C((1-\kappa)d_C-2\kappa)-b_C((3\kappa-2)d_C-2\kappa+2))}{2((1-\kappa)d_C-1)^2}$
		Competitive (K) *setting*	
	s	$V^{ss} = \frac{a_K(a_K d_K - 2b_K(d_K+1))}{2(d_K+1)^2}$	$V^{sc} = V^{ss} + \frac{\kappa d_K\left(2(1-\kappa)d_K^2+\kappa-2\right)(a_K+b_K(d_K+1))^2}{2(d_K+1)^2\left((1-\kappa)d_K^2-1\right)^2}$
	c	$V^{cs} = V^{cc} - \frac{\kappa\left(d_K\left(2(1-\kappa)d_K\left((1-\kappa)^2 d_K+\kappa\right)+3\kappa-2\right)-2\kappa\right)(a_K+b_K(d_K+1))^2}{2((1-\kappa)d_K+1)^2\left((1-\kappa)d_K^2-1\right)^2}$	$V^{cc} = \frac{((1-\kappa)a_K-\kappa b_K)(a_K d_K-\kappa a_K(d_K-2)-b_K((2-3\kappa)d_K-2\kappa+2))}{2((1-\kappa)d_K+1)^2}$

Table 2: Agents' payoffs in the social settings

we find that[32]

$$U^{ss} < U^{cs} < U^{sc} < U^{cc}. \tag{8}$$

In words, for the Cooperative settings, the highest utility is achieved when two Caring agents are paired, since both agents are concerned with each other's welfare and consequently internalize, at least partially, the externality arising from the complementarity of their social interaction. The second highest utility goes to a Selfish agent who is paired with a Caring agent, because the Selfish agent can take advantage of the Caring agent's concern. The third highest utility is achieved by a Caring agent who is paired with a Selfish agent, since the Caring agent exerts modest effort on behalf of the Selfish agent but gets only low effort in return. Finally, the lowest payoff is achieved when two Selfish agents are paired. Neither of them internalizes the externality arising from the complementarity of their interaction.

Furthermore, under the assumptions (7), we also find that

$$V^{cs} < V^{ss} < V^{cc} < V^{sc}. \tag{9}$$

For the Competitive setting, the highest utility is achieved by a Selfish agent when paired with a Caring agent, since the Selfish agent benefits both from her own selfishness and the altruism of the partner. The second highest utility is achieved when two Caring agents are paired, since each internalizes the Competitive externality to the other (at least partially). The third highest utility is achieved when two Selfish agents are paired, since neither of them internalizes this externality. Finally, the lowest utility is achieved by a Caring agent when paired with a Selfish agent, since the Caring agent suffers both from her own (partial) selflessness and the selfishness of the partner.

These results enable us to specify the contributions and utilities of the Responsive agents in both social settings. Observe that Cooperative settings tend to favor the survival of Caring agents, while Competitive settings rather tend to favor the survival of Selfish agents.[33] In line with our conception of multi-directedness above, we suppose that Responsive agents employ the Caring motive in Cooperative settings and the Self-interested Wanting motive in Competitive settings.[34] For this purpose, Responsive agents need to distinguish Cooperative from Competitive settings and adjust appropriately. We assume that such assessments and adjustments are not costless. In particular, Responsive agents are assumed to be subject to random mistakes in assessing their social context. For simplicity, let the expected cost of these mistakes in each

[32]Following the logic of Bester and Güth (1998), the highest κ that can be supported in a population consisting of *only* altruists (and only cooperative settings) is $1/c < 2\left(d_C^2 - 1\right) / \left(2\left(d_C^2 - d_C\right) + d_C^3\right)$.

[33]For a thorough treatment of this argument see Bester and Güth (1998).

[34]We do not interpret this as deliberately opportunistic behavior. Rather, each setting may be thought of as being associated environmental stimuli and cues. These stimuli may activate different decision-making processes. Convergent evidence from psychology and neuroscience supports the notion that humans' affect, thought patterns, perceptual sensitivity, and autonomic measures can change across contexts in ways that affect their decisions (Przyrembel et al., mimeo).

period of analysis be ξ, a positive constant. Then the utilities of the Responsive agent, alongside those of the Selfish and Caring agents, in the two social settings are given by Table 3.

		Other's dispositional type					
		s	r	c	s	r	c
		Coop. (C) *setting*			*Comp.* (K) *setting*		
Own	s	U^{ss}	U^{sc}	U^{sc}	V^{ss}	V^{ss}	V^{sc}
dispositional	r	$U^{cs}-\xi$	$U^{cc}-\xi$	$U^{cc}-\xi$	$V^{ss}-\xi$	$V^{ss}-\xi$	$V^{sc}-\xi$
type	c	U^{cs}	U^{cc}	U^{cc}	V^{cs}	V^{cs}	V^{cc}

Table 3: Payoffs of all agents in the social settings

4 The Reinforcement of Dispositional Types

We now consider the reinforcement of different dispositional types. We assume, as noted, that agents tend to develop those types for which they experience the highest expected payoffs. These payoffs depend on the probabilities with which they will encounter the Cooperative and Competitive settings, and the probabilities with which they encounter the dispositional types of others in those settings. As we have seen, the Caring type has a comparative advantage in Cooperative settings, whereas the Selfish type has a comparative advantage in Competitive settings. Responsive types are not best-suited for either setting – on account of the flexibility cost ξ – but they are better suited for Cooperative settings than Selfish agents, and better suited for Competitive settings than Caring agents.

Furthermore, within each type of setting, an individual does better if she meets another agent with the Caring motive than if she meets another agent with the Self-interested Wanting motive. That is, in Cooperative settings expected payoffs will depend on the relative proportion of Selfish agents (who have the Self-interested Wanting motive) compared to Responsive or Caring agents (who have the Caring motive); and in Competitive settings expected payoffs will depend on the relative proportion of Caring agents (who have the Caring motive) compared to Responsive or Selfish agents (who have the Self-interested Wanting motive). In this way, we may focus on two probabilities: the likelihood of encountering a selfish agent in a Cooperative setting, denoted by p_s^C, and the likelihood of encountering a Caring agent in a Competitive setting, denoted p_c^K.

Therefore the expected payoff from developing the Caring type (c), conditional on the agent's probability η_i of encountering a Cooperative setting, is

$$\Pi_i^c \equiv \eta_i \cdot \left(\left(1-p_s^C\right)U^{cc} + p_s^C U^{cs}\right) + (1-\eta_i)\cdot\left(p_c^K V^{cc} + \left(1-p_c^K\right)V^{cs}\right). \quad (10)$$

In words, the Caring agent's expected payoff in a Cooperative setting is the payoff from meeting another Care-motivated agent (U^{cc}) times the likelihood of meeting a Care-motivated agent ($1-p_s^C$), plus the payoff from meeting a

Self-interested Wanting-motivated agent (U^{cs}) times the likelihood of meeting the latter agent (p_s^C). The Caring agent's expected payoff in a Competitive setting is the payoff from meeting another Care-motivated agent (V^{cc}) times the likelihood of meeting a Care-motivated agent (p_c^K), plus the payoff from meeting a Self-interested Wanting-motivated agent (V^{cs}) times the likelihood of meeting the latter agent ($1 - p_c^K$). Since this agent encounters Cooperative settings with probability η_i and Competitive settings with probability $1-\eta_i$, each of the expected interaction payoffs is multiplied by the appropriate probability of that setting type.

The expected payoff from developing the Responsive type (r) is

$$\Pi_i^r \equiv \eta_i \cdot \left(\left(1 - p_s^C\right) U^{cc} + p_s^C U^{cs}\right) + (1 - \eta_i) \cdot \left(p_c^K V^{sc} + \left(1 - p_c^K\right) V^{ss}\right) - \xi. \quad (11)$$

In words, the Responsive agent's expected payoff in a Cooperative setting is the payoff from meeting another Care-motivated agent (U^{cc}) times the likelihood of meeting a Care-motivated agent ($1 - p_s^C$), plus the payoff from meeting a Self-interested Wanting-motivated agent (U^{cs}) times the likelihood of meeting the latter agent (p_s^C). The Responsive agent's expected payoff in a Competitive setting is the payoff from meeting a Care-motivated agent (V^{sc}) times the likelihood of meeting a Care-motivated agent (p_c^K), plus the payoff from meeting another Self-interested Wanting agent (V^{ss}) times the likelihood of meeting the latter agent ($1 - p_c^K$). Since this agent encounters Cooperative settings with probability η_i and Competitive settings with probability $1 - \eta_i$, each of the expected interaction payoffs is multiplied by the appropriate probability of that setting type. Note that the r dispositional type achieves the highest interaction payoffs in both the C and K settings, but also faces flexibility cost ξ.

The expected payoff from developing the Selfish dispositional type (s) is

$$\Pi_i^s \equiv \eta_i \cdot \left(\left(1 - p_s^C\right) U^{sc} + p_s^C U^{ss}\right) + (1 - \eta_i) \cdot \left(p_c^K V^{sc} + \left(1 - p_c^K\right) V^{ss}\right). \quad (12)$$

In words, the Selfish agent's expected payoff in a Cooperative setting is the payoff from meeting a Care-motivated agent (U^{sc}) times the likelihood of meeting a Care-motivated agent ($1 - p_s^C$), plus the payoff from meeting another Self-interested Wanting-motivated agent (U^{ss}) times the likelihood of meeting the latter agent (p_s^C). The Selfish agent's expected payoff in a Competitive setting is the payoff from meeting a Care-motivated agent (V^{sc}) times the likelihood of meeting a Care-motivated agent (p_c^K), plus the payoff from meeting another Self-interested Wanting-motivated agent (V^{ss}) times the likelihood of meeting the latter agent ($1 - p_c^K$). Since this agent encounters Cooperative settings with probability η_i and Competitive settings with probability $1 - \eta_i$, each of the expected interaction payoffs is multiplied by the appropriate probability of that setting type.

An agent will develop the c dispositional type rather than the r dispositional type when $\Pi_i^c > \Pi_i^r$. This implicitly defines a threshold $\eta = \eta^c$ above which individuals develop the c dispositional type and below which individuals develop the r dispositional type. Intuitively, this means that if agents are sufficiently

specialized in Cooperative settings, they do not develop the flexible r dispositional type because the opportunity cost of employing the Caring motive in Competitive settings, $p_c^K\left(V^{sc}-V^{cc}\right)+\left(1-p_c^K\right)\left(V^{ss}-V^{cs}\right)$, does not justify the flexibility cost ξ. Agents who will not be overly specialized in Cooperative settings will suffer the flexibility cost and develop the r dispositional type.

Likewise, an agent develops the s dispositional type over the r dispositional type when $\Pi_i^s > \Pi_i^r$. This implicitly defines a threshold $\eta = \eta^s$ below which individuals develop the s dispositional type and above which individuals develop the r dispositional type. Intuitively, this means that if agents are sufficiently specialized in Competitive settings, they do not develop the flexible r dispositional type because the opportunity cost of employing the Self-interested Wanting motive in Cooperative settings, $\left(1-p_s^C\right)\left(U^{cc}-U^{sc}\right)+p_s^C\left(U^{cs}-U^{ss}\right)$, does not justify the flexibility cost ξ. Agents who will not be overly specialized in Competitive settings will suffer this cost and develop the r dispositional type.

Solving $\Pi_i^c = \Pi_i^r$ for $1-\eta^c$, we find that

$$1-\eta^c = \frac{\xi}{p_c^K\left(V^{sc}-V^{cc}\right)+\left(1-p_c^K\right)\left(V^{ss}-V^{cs}\right)}, \tag{13}$$

meaning the share of Caring agents will depend on the ratio of the flexibility cost to the opportunity cost of forgoing high Competitive-setting payoffs.

Likewise, solving $\Pi_i^s = \Pi_i^r$ for η^s, we find that

$$\eta^s = \frac{\xi}{\left(1-p_s^C\right)\left(U^{cc}-U^{sc}\right)+p_s^C\left(U^{cs}-U^{ss}\right)}, \tag{14}$$

meaning the share of Selfish agents will depend on the ratio of the flexibility cost to the opportunity cost of forgoing high Cooperative-setting payoffs.

In order to ensure that there are positive fractions of both Caring and Selfish dispositional types, we assume that

$$0 < \xi < \min\left\{U^{cs}-U^{ss}, V^{sc}-V^{cc}, V^{ss}-V^{cs}\right\}. \tag{15}$$

Intuitively, we must assume a positive flexibility cost $\xi > 0$ because at $\xi = 0$ any individual would be able to employ the best-suited motive in any setting, regardless of how infrequently she encountered each type of setting. To assume that people are able to shift their motives to suit very unfamiliar situations seems psychologically implausible however. For even arbitrarily low flexibility costs $\xi > 0$ however, there will exist people with sufficiently high likelihoods of encountering Cooperative (Competitive) settings that they develop Caring (Selfish) rather than Responsive dispositional type. Conversely, to ensure that there are some Responsive types in the population, we must assume that the flexibility cost does not exceed the opportunity cost of employing the less-suited motivation to a particular setting. That is, the cost of being able to switch motives should be less than the gain from being able to switch motives.

Having identified the cutoffs η^c and η^s, the number of agents who develop each dispositional type is determined by the distribution H of η. Specifically,

$$n_c = 1 - H(\eta^c).$$
$$n_r = H(\eta^c) - H(\eta^s), \text{ and}$$
$$n_s = H(\eta^s),$$

5 The Social Topography

5.1 Social Opportunities

We now consider the likelihood that agents of each dispositional type will encounter Cooperative vs. Competitive settings. Depending on the relative probabilities of the Cooperative and Competitive setting – a phenomenon that depends on an individual's parameter η_i – agents of all dispositional types will encounter both social settings under different circumstances[35]. Recall that η is distributed according to $H(\cdot)$ with mean $\overline{\eta} \in (0,1)$ and density $h(\cdot)$.

The probability that an agent encountering a Cooperative setting has the s dispositional type, p_s^C, may be expressed using Bayes' rule:

$$p_s^C = P(s \mid C) = \frac{P(C \mid s) \cdot n_s}{P(C)} = \frac{E(\eta_i \mid s) \cdot H(\eta^s)}{\overline{\eta}}.$$

We may express the conditional expectation $E(\eta_i \mid s)$ by

$$E(\eta_i \mid s) = \frac{\int_0^{\eta^s} t \cdot h(t)\, dt}{H(\eta^s)},$$

as it is the mean of a truncated distribution. Therefore, we have

$$p_s^C = \frac{1}{\overline{\eta}} \int_0^{\eta^s} t \cdot h(t)\, dt, \tag{16}$$

which we term the *social opportunities function for Cooperative settings*. This represents the chance that an individual in a Cooperative setting has to find a Care-motivated partner, which has implications for how fruitful their interactions will be. Intuitively, since we need to know the likelihood of encountering a Self-interested agent in a Cooperative setting, we must take into account the total number of agents with the Selfish dispositional type ($n_s = H(\eta^s)$); but then we must also account for the fact that those who have developed the Selfish dispositional type are those least likely to encounter Cooperative settings (i.e., we must know $E(\eta_i \mid s)$). To derive the conditional (expected) likelihood of encountering a Cooperative setting for these types, we must integrate over the relevant support of the density of η, specifically over those values of η less than η^s since only those agents will have developed the Selfish dispositional type.

[35] We assume here that the distribution $H(\cdot)$ of η has no mass at either 0 or 1.

Finally, we must divide by the total likelihood of encountering a Cooperative setting $\bar{\eta}$, the probability of the conditioning event.

Likewise, the probability that an agent encountering a Competitive setting has the c dispositional type, p_c^K, is

$$p_c^K = P(c \mid K) = \frac{P(K \mid c) \cdot n_c}{P(K)} = \frac{(1 - E(\eta_i \mid c)) \cdot (1 - H(\eta^c))}{1 - \overline{\eta}}.$$

Substituting in the expression for the conditional expectation, we get

$$p_c^K = \frac{n_c - \int_{\eta^c}^1 t \cdot h(t)\, dt}{1 - \overline{\eta}}, \tag{17}$$

which we term the *social opportunities function for Competitive settings.* This represents the chance that an individual in a Cooperative setting has to find a Care-motivated partner, which has implications for how fruitful their interactions will be. Intuitively, since we need to know the likelihood of encountering a Care-motivated agent in a Competitive setting, we must take into account the total number of agents with the Caring dispositional type ($n_c = 1 - H(\eta^c)$); but then we must also account for the fact that those who have developed the Caring dispositional type are those least likely to encounter Competitive settings (i.e., we must know $1 - E(\eta_i \mid c)$). To derive the conditional (expected) likelihood of encountering a Competitive setting for these types, we must integrate over the relevant support of the density of η, specifically over those values of η greater than η^c since only those agents will have developed the Caring dispositional type. Finally, we must divide by the total likelihood of encountering a Competitive setting $1 - \bar{\eta}$, the probability of the conditioning event.

5.2 Equilibrium Social Topography

We are now able to characterize how the social topography is determined in equilibrium. Recall that the cutoffs determining the number of agents developing each dispositional type (equations 13 and 14) depend on the likelihood of encountering Care-motivated and Self-interested Wanting motivated agents in the Competitive and Cooperative setting, respectively. Since these equations captured the social forces acting on dispositional type development, they may be represented by "disposition development curves. The likelihoods of encountering each type of agent may be represented by "social opportunities" curves (equations 16 and 17), which in turn depended on the total shares of each type of agent in the population. The equilibrium is characterized by the intersection of these two curves. Because the probabilities p_c^K and p_s^C are continuous onto[36] the unit interval, we may invoke the Brouwer fixed point theorem to claim that such an equilibrium exists.

[36]Clearly, if the share of the population developing the selfish dispositional type is zero, then the likelihood of encountering a selfishly-motivated person in the cooperative setting will be zero as well. Likewise, if all agents have the selfish dispositional type, the probability of encountering this trait in the cooperative setting will be 1.

Graphically, this equilibrium can be visualized in Figures 2 and 3. The horizontal axis measures the probability of encountering an s agent in a Cooperative setting, while the vertical axis corresponds with the share of the population adopting the s dispositional type. Recall that in section 4 we derived the shares of the population developing the s dispositional type as a function of the likelihood of encountering others with the s dispositional type in a Cooperative setting. This "disposition development" curve was based on a comparison of the relative payoffs between s and r agents in the Cooperative social setting because these two dispositional types behaved equivalently in Competitive settings. On the graph, we see that the share of the population developing the s dispositional type (driving the evolution of dispositional types in the population, measured on the vertical axis) is increasing in the likelihood of meeting another s-dispositional type agent in a Cooperative setting (measured on the horizontal axis).

Recall also that in section 5.1, we derived the likelihood of encountering an s-dispositional type agent in the Cooperative setting based on the total share of s agents in the population. This was based on a straightforward application of Bayes' rule. This "social opportunities" curve indicates that the likelihood of meeting another s-dispositional type agent (describing the social opportunities facing the agents, measured on horizontal axis) is increasing in the share of the population developing the s dispositional type (i.e. if there are more of them, you are more likely to run into one).

At the equilibrium, lying at the intersection between the disposition development curve and the social opportunities curve, the share of the population adopting the s dispositional type (which is increasing in the likelihood of meeting another s-dispositional type) is consistent with the actual probability of encountering an s agent as determined by the social opportunities function for Cooperative settings. Note that the social opportunities curve intersects the origin, and the point $\left(p_s^C = 1, \eta^s = 1\right)$, while the disposition development curve lies strictly between 0 and 1 for appropriate restrictions on ξ (see previous section). Based on these restrictions we know that the two curves must cross at a point in the interior of the interval[37].

Since the social opportunities curve depends on the shape of $h(\cdot)$ however, it is possible that for sufficiently "non-uniform" distributions, this curve may intersect the disposition development curve more than once, meaning that multiple social topography equilibria could exist[38]. In the case where multiple equilibrium social topographies are possible, there will be at least one unstable equilibrium, plus another stable one. To understand the stability of the equilibrium, it is sufficient to consider whether the disposition development curve

[37]In principle, corner point equilibria in which the entire population develops one dispositional type are possible, provided that we consider extreme assumptions on the flexibility cost ξ. For $\xi = 0$, all agents would develop the Responsive dispositional type while for ξ very large no agents would develop this dispositional type.

[38]It is easy to verify that for $h(\cdot)$ uniform, the social opportunities curve is convex on the entire unit interval. Note that $\partial p_s^C / \partial \eta^s = \eta^s \cdot h(\eta^s) / \bar{\eta}$ and $\partial^2 p_s^C / \partial (\eta^s)^2 = \left(h(\eta^s) + \eta^s h'(\eta^s)\right) / \bar{\eta}$.

intersects the social opportunities curve from above or below. When the disposition development curve lies above the social opportunities curve, this means that a greater share of the population could profitably develop the Selfish dispositional type than currently have this dispositional type, which puts upward pressure on the share of Selfish agents; whereas if the disposition development curve lies below the social opportunities curve there are too many Selfish agents than could profit by developing this dispositional type, putting downward pressure on the share of Selfish agents. Therefore when the disposition development curve intersects the social opportunities curve from above the equilibrium is stable, while when the disposition development curve intersects the social opportunities curve from below, the resulting equilibrium is unstable (i.e. if there were slightly more Selfish agents, more agents would want to be Selfish and vice-versa). We consider the possibility of multiple equilibria intriguing, as "big pushes" to promote more cooperation may be possible. However, we focus on cases of unique equilibria for policy evaluation purposes[39].

The equilibrium number of Caring dispositional types is likewise determined by the intersection of the c-disposition development curve and the likelihood of encountering a c-dispositional type in a Competitive setting. Similar considerations for the equilibrium share of s agents apply here as well, with the additional possibility that the number of agents developing the Caring dispositional type may be either increasing or decreasing in the share of other Caring agents, depending on how easy they are to exploit. When $V^{cc}+V^{ss} > V^{cs}+V^{sc}$, the prevalence of other Caring agents increases the advantage of this dispositional type (as they can interact with each other more). However, if $V^{cc}+V^{ss} < V^{cs}+V^{sc}$, any increase in Caring dispositional types benefits the selfishly-motivated by more, since these agents become very profitable to exploit[40] (see right panel of Figure 3).

[39]None of the results of our policy evaluations will depend on which (stable) equilibrium is selected.

[40]Under such parameterizations a marginal increase in the share of the population with the Caring dispositional type benefits the self-interested more than the Care-motivated in competitive settings. Or rather the gain from a Self-interested agent finding a caring rather than Self-interested partner to exploit is greater than the gain from a care-motivated agent finding another care motivated agent to cooperate with.

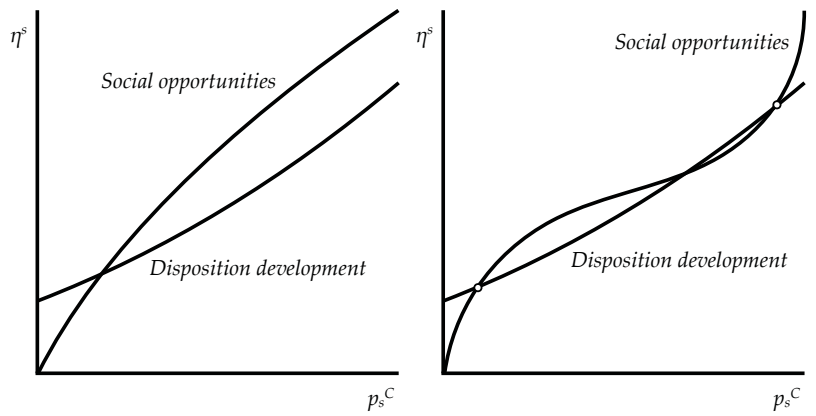

Figure 2: Determination of the equilibrium level for the Selfish dispositional type under uniform (left) and more exotic distributions of cooperative settings (right, stable equilibria indicated)

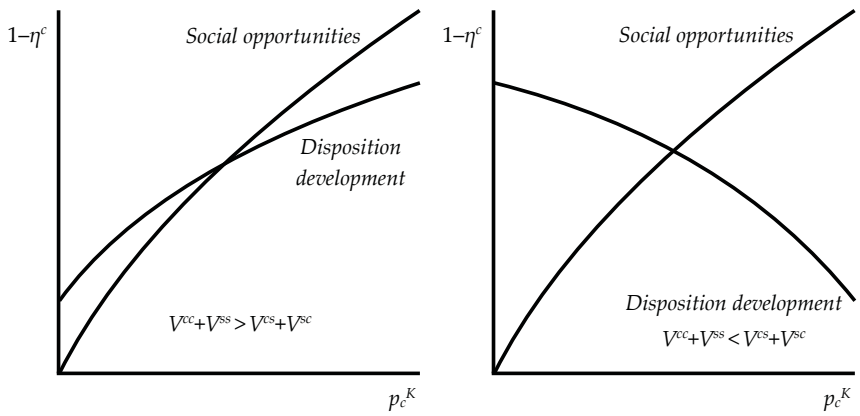

Figure 3: Determination of the equilibrium level for the Caring dispositional type – note that the fitness of the Caring dispositional type may be decreasing in the number of other caring agents if exploiting these agents is sufficiently profitable

6 Policies to Promote Cooperation

Could policies that promote more cooperative social settings lead to the evolution of a more caring society? In our model, a more caring society may be represented by a rise in the number of Caring and Responsive agents engaging in caring interactions with others.

In this section we consider three policy exercises. First, we analyze the effects of an increase in the prevalence of cooperative social settings, encouraging the development of more cooperative dispositional types. Various policies, such as incentives and institutions that discourage anti-social behavior, can serve this purpose.[41]

Second, we consider pecuniary incentives aimed at promoting cooperation. In this context, we shed light on the conditions under which such incentives "crowd-in" or "crowd-out" social preferences[42].

And finally, we examine policies promoting internal change favoring prosociality in the individual. In particular, we consider the effects of mental training that enables the recipients to derive an additional utility gain from being cooperative. This policy involves training the individual to experience more gratification from care, as distinct from policies that change the prevalence of cooperative settings or change the pecuniary incentives to promote cooperation.

6.1 Welfare Framework

We consider welfare W in terms of the sum of the expected payoffs from the Cooperative and Competitive settings, respectively, each weighted by the share of the population that encounters each setting[43]. Any potential policy changes within our framework can then be decomposed into how they affect the payoffs in the social setting holding the distribution of dispositional types fixed (static efficiency), and how these policy changes influence the distribution of dispositional types in the population:

$$dW = \partial W + \frac{\partial W}{\partial p_s^C} \cdot dp_s^C + \frac{\partial W}{\partial p_c^K} \cdot dp_c^K. \tag{18}$$

[41] As plasticity is largest in childhood, it may be desirable to structure education environments to encourage teamwork from an early age. Furthermore, since research has shown dispositional traits to exhibit some flexibility even into adulthood, social environments could be adapted through choice architecture and institutional design to foster cooperation.

[42] See Bowles and Polanía-Reyes (2012) for a comprehensive review. Numerous contributions explain why pecuniary incentives may undermine people's intrinsic motivations to act in the common interest (Titmuss, 1970; Deci & Ryan, 1985; Frey and Jegen, 2001; Sandel, 2012; Bowles, 2008). Arguments for this proposition range from people's need for self-determination (Deci and Ryan) to image concerns (Bénabou and Tirole, 2006) to displacing social norms (Sliwka, 2007) to framing decisions differently (Tversky and Kahneman, 1981), and by working to change people's preferences (Bowles, 1998).

[43] Refer to the appendix for an explicit formulation.

6.2 Promotion of Cooperative Settings

As our first policy exercise, we consider shifting the distribution of Cooperative relative to Competitive settings. In particular, we examine the effects of an increase in the average share of the population that encounters Cooperative settings, $\bar{\eta}$.

Changes in static efficiency depend on the difference in expected payoffs between Cooperative and Competitive settings. Such static analysis suggests that increasing the prevalence of Cooperative settings is only worthwhile when they are more productive than Competitive settings. As shown below, however, this may be mistaken in a dynamic context.

An increase in the share of cooperative settings would have a direct impact on the social opportunities curves, both in Cooperative and Competitive settings. Recall that in equations 16 and 17, the likelihood p_s^C of encountering an s-dispositional type in the Cooperative setting was decreasing in $\bar{\eta}$, while the likelihood of encountering a c-dispositional type in the Competitive setting was increasing in $\bar{\eta}$. Using our equilibrium graphs, we can see that when the social opportunities curve shifts up in the Cooperative setting, it intersects the disposition development curve at a lower share of Selfish dispositional types (Figure 4, left panel). Similarly, when the social opportunities curve shifts down in the Competitive setting, it intersects the disposition development curve at a higher share of Caring dispositional types (Figure 4, right panel)

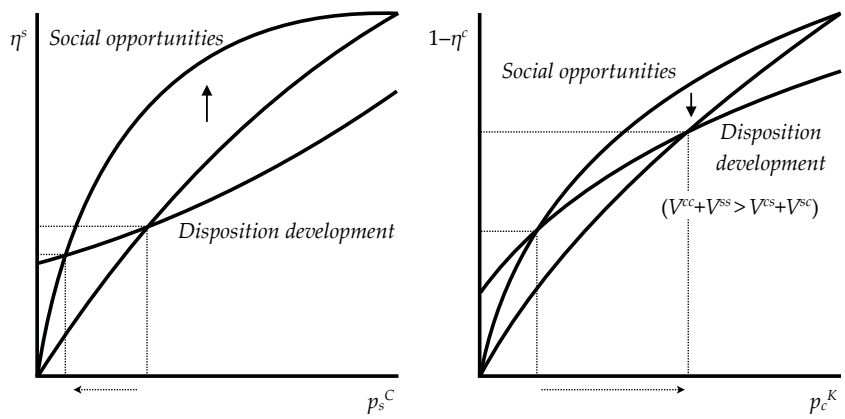

Figure 4: Effect of increase in Cooperative settings on dispositional type development

Intuitively, here the increase in Cooperative settings changes where the marginal types lie in the overall distribution. Some people who would have been selfish may now develop the responsive dispositional type since their likelihood of encountering a Cooperative setting goes up, and some people who would have been responsive types may now develop the caring dispositional type since their likelihood of encountering a Competitive setting goes down. Since over-

all welfare is decreasing in p_s^C and increasing in p_c^K, this policy has beneficial second-order effects that go beyond considerations of whether Cooperative settings are more productive than Competitive settings.

6.3 Pecuniary Incentives

Let us now consider how pecuniary incentives – taxes on Competitive activities and subsidies for Cooperative activities – affect the evolution of cooperation. We will show that when pecuniary incentives increase the payoffs to those with Caring dispositional type relative to those with Selfish dispositional type, that incentives support the development of that dispositional type and crowd in voluntary cooperation. Such policies work by discouraging exploitation of the cooperative by the selfish. If pecuniary incentives work to increase the payoffs to the selfishly motivated relative to those motivated by care however, then they discourage the development of caring dispositional type and crowd out voluntary cooperation. In this way the incentives substitute for social preferences, allowing the selfish to achieve for themselves what was formerly achievable by the other-regarding alone.

6.3.1 Taxes on Competition

Consider a policy that takes the form of a tax the Competitive activity. Though policy makers are generally not accustomed to thinking of incentive schemes in these terms, many policies do in fact have the character of discriminating among cooperative and competitive activities. For example, many jurisdictions impose taxes on luxury (positional) goods. Additionally, many policies provide stronger incentives at lower levels of pro-social behavior. For example, charitable contributions deducted from income subject to a progressive income tax are effectively subsidized at the marginal tax rate. In this way, smaller charitable contributions are subsidized at a higher rate because the resulting taxable income is subject to a higher tax bracket.

Within our framework, we consider a constant marginal tax of τ_i on each unit of effort y_i exerted in the Competitive setting. We assume that since differently motivated individuals engage in different levels of competition, the tax can effectively discriminate by motive, for example by taxing only high levels of competition[44]. Assume also that it can allocate the revenue from this task in lump sum form. The payoff to the Competitive activity now becomes

$$V_{ij} \equiv -y_i y_j + (a_K - \tau_i)\, y_i - b_K y_j - \frac{d_K}{2} y_i^2 + R_i \tag{19}$$

where R_i is a lump sum transfer.

[44]While such an incentive scheme may seem strange, many policies do in fact have the character of providing stronger incentives at lower levels of pro-social behavior. For example, charitable contributions deducted from income subject to a progressive income tax are effectively subsidized at the marginal tax rate. In this way, smaller charitable contributions are subsidized at a higher rate because the resulting taxable income is subject to a higher tax bracket.

The institution must balance its budget, meaning that it can disburse as a lump sum an amount equal to the total amount of tax revenue T_K.

Since policymakers often focus on the "worst offenders", we assume that taxes are only imposed on those engaging the most intensely in competition (meaning the Care-motivated are not taxed, (i.e. $\tau_c = 0$, see above) and that all the tax revenue is remitted to those taxed[45] (i.e. $R_c = 0$). We first derive the interaction payoffs under this tax scheme, and show that there are static welfare gains from imposing the tax.[46] We will then show that by narrowing the difference in Competitive setting interaction payoffs between c and r/s agents, that this policy also reduces the opportunity cost of developing the c dispositional type, and therefore has additional positive externalities.

Static Efficiency

Imposing a tax on effort in the Competitive setting increases static efficiency because taxes reduce the intensity of competition. Since there are negative externalities associated with increased competition in this setting, agents exert too much effort relative to what is socially optimal. Therefore appropriately set taxes can curb competition and thereby raise efficiency.

Effects on Reinforcement of Dispositional Types

It can be shown that raising taxes narrows the relative payoff disadvantage that Caring dispositional types face in the Competitive setting (i.e. $d\left(V^{cc} - V^{sc}\right)/d\tau_s$, $d\left(V^{cs} - V^{ss}\right)/d\tau_s > 0$).[47] Using our equilibrium graphs, we can see that when the disposition development curve shifts up, it intersects the social opportunities curve at a higher share of Caring dispositional types (Figure 5).

Note also that the tax does not affect the payoffs in Cooperative settings (i.e. $dp_s^C/d\tau_s = 0$). Intuitively, here the tax acts as a relative penalty to developing the Selfish or Responsive dispositional type, so some individuals whose probability of participation in Cooperative settings was sufficiently low that they developed the r dispositional type before, will now find it less costly to develop the c dispositional type. Since overall welfare is increasing in p_c^K, this policy has beneficial second-order effects as well.

6.3.2 Subsidies for Cooperation

We now consider a policy that promotes prosocial activity through the use of subsidies to interactions in the Cooperative setting. Specifically, suppose that a policymaker provides a constant marginal subsidy of σ_i to each unit of effort x_i

[45] Even though all tax revenue may be refunded to the s- and r-dispositional-type agents, their utility may go down. One can in principle however imagine more complex tax and transfer schemes that yield Pareto improvements. Because we will show that total welfare must go up, it will be possible to disburse the resulting surplus in a way that leaves the more competitive agents no worse off and still promotes development of the c dispositional type.

[46] Refer to the appendix for a formal analysis.

[47] See appendix.

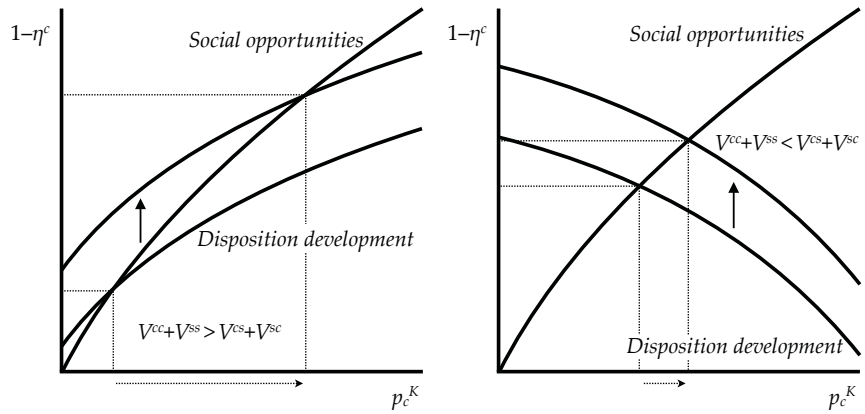

Figure 5: Effect of competition tax on Caring dispositional type development

exerted in cooperation. We assume that since effort contributions are monotonic in type, that the subsidy can discriminate by motive, for example by subsidizing only low levels of cooperation. Assume also that it can fund this subsidy through lump sum taxes. The payoff to the Cooperative activity now becomes

$$U_{ij} \equiv x_i x_j + (a_C + \sigma_i)\, x_i + b_C x_j - \frac{d_C}{2} x_i^2 - R_i. \tag{20}$$

where R_i is a lump-sum tax.

The institution must balance its budget, meaning that it must raise a lump sum equal to the total amount of subsidy that it disburses. Call this amount T_C.

Since policymakers often focus on the "worst offenders", we assume that subsidies are only provided to low cooperators (i.e. $\sigma_c = 0$) and that all the tax revenue is raised from those subsidized (i.e. $R_c = 0$). We first derive the interaction payoffs under this subsidy scheme, and show that there are static welfare gains from providing the subsidy.[48] We will then show that by narrowing the difference in Cooperative setting interaction payoffs between s and c/r agents, that this policy also reduces the opportunity cost of developing the s dispositional type, and therefore has countervailing negative externalities.

Static Efficiency

Giving a subsidy for effort in the Cooperative setting increases static efficiency because subsidies increase the intensity of cooperation. Since there are positive externalities associated with increased cooperation in this setting, agents exert too little effort relative to what is socially optimal. Therefore appropriately set subsidies can bolster cooperation and thereby raise efficiency.

[48] Refer to the appendix for a formal analysis.

Effects on Reinforcement of Dispositional Types

It can be shown that increasing subsidies narrows the relative payoff disadvantage that Selfish dispositional types face in the Cooperative setting (i.e. $d\left(U^{cc}-U^{sc}\right)/d\sigma_s,\ d\left(U^{cs}-U^{ss}\right)/d\sigma_s<0$).[49] Using our equilibrium graphs, we can see that when the disposition development curve shifts up, it intersects the social opportunities curve at a higher share of Caring dispositional types (Figure 6).

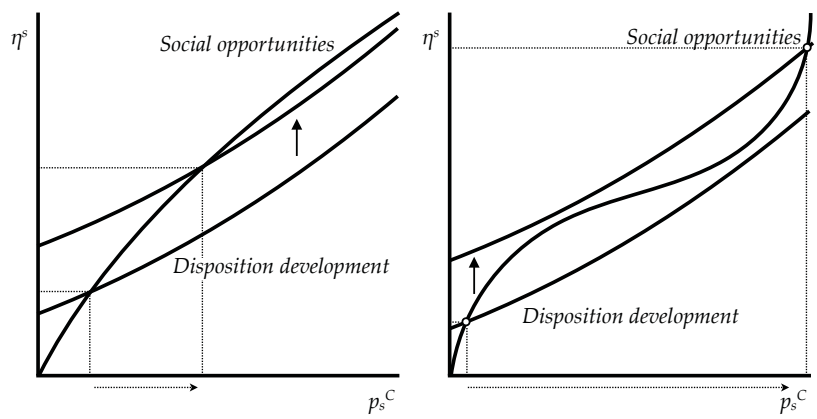

Figure 6: Effect of cooperation subsidy on Selfish dispositional type development

Note also that the subsidy does not affect the payoffs in Competitive settings (i.e. $dp_c^K/d\sigma_s=0$). Intuitively, here the subsidy acts as a relative penalty to developing the Caring or Responsive dispositional types, so some individuals whose probability of participation in Competitive settings was sufficiently low that they developed the r dispositional type before, will now find it less costly to develop the s dispositional type. Since overall welfare is decreasing in p_s^C, this policy has countervailing second-order effects.

6.4 Internal Change

We now consider the effects of a policy that promotes the individual's welfare from caring interactions. This may take the form of support for mental training that enables the individual to achieve a greater payoff from providing Care. In particular, we suppose that an individual who has received such training experiences an extra payoff when engaging in the caring activity (but no extra payoff when engaging in self-interested activities).

More specifically, we suppose that individuals are given the opportunity to pursue "compassion meditation" apart from their interactions within Cooperative or Competitive social settings. While we conceive of this activity as a

[49] See appendix.

setting, it is one which people can choose to participate or not. Furthermore, we assume that agents' utilities from this meditation activity do not depend on the actions of others. In particular, suppose that compassion meditation yields a utility of

$$U_{ij} \equiv \left(1 - p_s^C\right) U^{ic} + p_s^C U^{is} + \begin{cases} M - d_M & \text{motive} = c \\ -d_M & \text{motive} \neq c \end{cases} \tag{21}$$

with M, $d_M > 0$ and $M > d_M$. The parameter d_M represents the cost of participation. Those individuals who can employ the Care motive (those with the r and c dispositional types) experience an additional utility of $M - d_M$ from engaging in compassion meditation, whereas those who cannot (the s dispositional types) would experience only the cost d_M and would refrain from choosing to participate in this setting.

These self-chosen settings will have an effect on dispositional type reinforcement. The cost parameter d_M affects the cutoff at which people switch between r and s dispositional types. In this context, the cutoff equation 14 may be modified as follows:

$$\eta^s = \frac{\xi - M + d_M}{\left(1 - p_s^C\right)\left(U^{cc} - U^{sc}\right) + p_s^C\left(U^{cs} - U^{ss}\right)}. \tag{22}$$

An internal change policy can affect the equilibrium distribution of dispositional types by making compassion meditation more widely accessible, for example. This would take the form of a reduction in d_M. The resulting shift in the disposition development curve is illustrated in Figure 7.

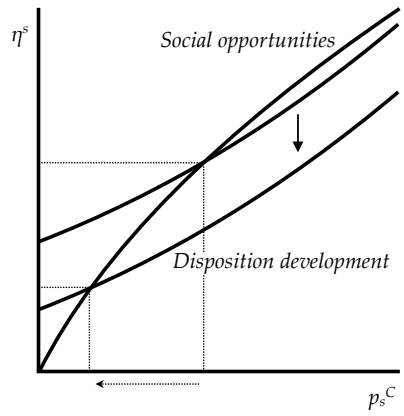

Figure 7: Effect of internal change on Selfish dispositional type development

The figure shows that this internal-change policy leads to a fall in the likelihood of encountering a selfish agent in a Cooperative setting (p_s^C) and a fall in the share of selfish agents (η^s).

7 Concluding Remarks

One major reason for the evolutionary success of humans lies in their ability to cooperate. Mainstream economics views humans as self-interested, rational individuals, and thus all cooperation among them must take the form of economic synergies. These synergies occur when individuals exploit available gains from trade, as described by Adam Smith's "invisible hand" mechanism. This paper, by contrast, explores social avenues of cooperation that have been largely ignored in economics thus far. We began from the premise that economic cooperation presupposes social cooperation, since people who lack trust in and goodwill toward one another are unlikely to be either willing or able to exploit existing gains from trade. The paper addresses both proximate and ultimate sources of cooperation. The proximate sources explain people's Cooperative decisions in terms of their objectives, which may be both individualistic (as in mainstream economics) and pro-social. Unlike the standard models of pro-social behavior in behavioral economics (where pro-social preferences are assumed to be located in the minds of individuals and thus wholly ascribable to them), in our model preferences are the outcome of the interplay between the individual and her environment, influenced by the individual's dispositional type and her social interactions within her social settings. The ultimate sources of cooperation explains these preferences in terms of the evolutionary forces acting on them. In this sense, our model is a preliminary attempt to account for the existence and prevalence of particular dispositional types and their relation to social settings and the broader social topography.

Our analysis addresses these phenomena by recognizing that all economic decisions are motivated and that people's motives depend on their dispositional types and social settings. For simplicity, we considered three dispositional types: Caring, Selfish and Responsive. The motives of Responsive individuals may be elicited by their social setting, of which we considered two: one entailing Cooperative interactions (where people's contributions are complementary) and one entailing Competitive interactions (where their contributions are substitutable). The sum of these social settings, aggregated over all agents, determines the social topography. The relative fitness of the various dispositional types determines their evolution.

In short, the degree of cooperation in an economy is portrayed as the outcome of reflexive interplay between people's individual economic decisions and the social forces to which they are subject. Their individual decisions determine the outcomes of their social interactions, which influence the relative fitness of dispositional types and their social topography, which in turn influences their individual decisions. In the course of these reflexive interactions, people's social preferences and their social topography co-evolve. In our model, Selfish dispositional types do not necessarily drive Caring and Responsive dispositional types out of existence. Rather, all dispositional types coexist in an environment of idiosyncratically varying social contexts. This coexistence of disposition types represents a "social balance", in which different agents survive in process of interacting with one another.

In this way, our paper is a first modest step towards explaining why people are prone to cooperate beyond what individualistic responses to economic synergies would imply; how this cooperation depends on their dispositional types, settings and social topography; why people vary in their willingness to cooperate even under identical environmental conditions; and how their dispositional type, preferences and social topography evolve. We contribute to the literature on the evolution of preferences (e.g. Robson and Samuelson, 2011) by analyzing how a variety of contexts can support a stable population of heterogeneous types. This literature typically features considerations of only how one game favors the survival of a particular set of preferences, but usually does not consider multiple environments with differing incentives, and to our knowledge has never considered preferences that change across environmental contexts.

Finally, our analysis is a preliminary attempt to explore how policies that promote pro-social behavior (by affecting people's internal and external environments) may affect the evolution of dispositional types and social contexts. In this respect, our analysis seeks to identify new conceptual avenues for addressing insufficient cooperation in society. Whereas mainstream economics addresses them by proposing pecuniary incentives, regulations, and institutional changes that induce self-interested, rational agents to internalize existing externalities, our approach both cautions that these policies may have unintended consequences and points to social, educational and institutional changes, which might affect people's degree of pro-sociality through the settings they encounter, their payoffs from these settings, and the fitness of different dispositional types.

References

[1] Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. Quarterly Journal of Economics, 115(3), 715-753.

[2] Akerlof, G. A., & Kranton, R. E. (2010). Identity Economics: How Our Identities Shape Our Work, Wages, and Well-being, Princeton: Princeton University Press.

[3] Alger, I. & Weibull, J. W. (2012). A generalization of Hamilton's Rule - Love others how much? Journal of Theoretical Biology, 299, 42-54.

[4] Allport, G.W. (1937), Personality: A Psychological Interpretation, New York: Holt, Rinehart and Winston.

[5] Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. Economic Journal, 100(401), 464-477.

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

[7] Atkinson, J. W., & Feather, N. T. (eds.). (1966). A theory of achievement motivation. New York: Wiley.

[8] Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. American Economic Review, 96(5), 1652-1678.

[9] Balliet, D., Parks, C. D., & Joireman, J. (2009). Social value orientation and cooperation: A meta-analysis. Group Processes and Intergroup Relations, 12, 533-547.

[10] Baltes, P. B. (1987). Theoretical Propositions of Life-Span Developmental Psychology: On the Dynamics Between Growth and Decline. Developmental Psychology, 23(5), 611-626.

[11] Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. American Economic Review, 97(2), 170-176.

[12] Bernheim, B. D. (1994). A theory of conformity. Journal of Political Economy, 102(5), 841-877.

[13] Bernheim, B. D., & Rangel, A. (2004). Addiction and cue-triggered decision processes. American Economic Review, 94(5): 1558-1590.

[14] Bester, H., & Güth, W. (1998). Is altruism evolutionarily stable? Journal of Economic Behavior and Organization, 34(2), 193-209.

[15] Bolton, G. E. & Ockenfels, A. (2000), ERC: A theory of equity, reciprocity, and competition. American Economic Review, 90(1), 166-193.

[16] Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. Journal of Economic Literature, 36(1), 75-111.

[17] Bowles, S. (2008). Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments. Science, 320, 1605-1609.

[18] Bowles, S., & Polanía-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? Journal of Economic Literature, 50(2), 368-425.

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

[20] Bronfenbrenner, U. (1979). The Ecology of Human Development: Experiments by Nature and Design. Cambridge, MA: Harvard University Press.

[21] Bulow, J. I., Geanakoplos, J. D., & Klemperer, P. D. (1985). Multimarket oligopoly: Strategic substitutes and strategic complements. Journal of Political Economy, 93(3), 488-511.

[22] Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. Quarterly Journal of Economics, 117, 817-869.

[23] Cox, J. C., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. Games and Economic Behavior, 59(1), 17-45.

[24] Deci, E. L., & Ryan, R. M. (1985). Intrinsic Motivation and Self-Determination in Human Behavior. New York: Plenum Press.

[25] Dubey, P., Haimanko, O., & Zapechelnyuk, A. (2006). Strategic complements and substitutes, and potential games. Games and Economic Behavior, 54(1), 77-94.

[26] Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. Games and Economic Behavior, 73(2), 459-478.

[27] Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. Games and Economic Behavior, 47(2), 268-298.

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

[29] Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. Games and Economic Behavior, 54(2), 293-315.

[30] Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. Quarterly Journal of Economics, 114(3), 817-868.

[31] Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally Cooperative? Evidence from a public goods experiment. Economics Letters, 71(3), 397-404.

[32] Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. Journal of Personality and Social Psychology, 80(6), 1011-1027.

[33] Fosgaard, T. R., Hansen, L. G., & Wengström, E. (2014). Understanding the nature of cooperation variability. Journal of Public Economics, 120, 134-143.

[34] Frank, R. (1988). Passions Within Reason: The Strategic Role of the Emotions. New York: Norton.

[35] Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. Journal of Economic Surveys, 15(5), 589-611.

[36] Gilbert, P. (2013), Mindful Compassion, London: Constable and Robinson.

[37] Harstad B., & Liski M. (2013). Games and Resources. In: J.F. Shogren (ed.), Encyclopedia of Energy, Natural Resource, and Environmental Economics, 2, 299-308. Amsterdam: Elsevier.

[38] Gelcich, S., Guzman, R., Rodriguez-Sickert, C., Castilla, J. C., &, Cárdenas, J. C. (2013). Exploring external validity of common pool resource experiments: Insights from artisanal benthic fisheries in Chile. Ecology and Society, 18(3): 2.

[39] Gneezy, U. & Rustichini, A. (2000). Pay enough or don't pay at all. Quarterly Journal of Economics. 115(3), 791-810.

[40] Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. Journal of Economic Perspectives, 25(4), 191-210.

[41] Goeree, J. K., Maasland, E., Onderstal, S., & Turner, J. L. (2005). How (not) to raise money. Journal of Political Economy, 113(4), 897-918.

[42] Guttmann, J. M. (2013). On the evolution of conditional cooperation. European Journal of Political Economy, 30, 15-34.

[43] Heckhausen, H. (1989). Motivation und Handlung. Berlin: Springer.

[44] Heckhausen, J. (2000). Evolutionary perspectives on human motivation. American Behavioral Scientist 43(6): 1015-1029.

[45] Heckhausen, J., & Heckhausen, H. (2010). Motivation und Handeln, Berlin: Springer.

[46] Holländer, H. (1990). A social exchange approach to voluntary cooperation. American Economic Review, 80(5), 1157-1167.

[47] Jackson, D. N., & Paunonen, S. V. (1980). Personality structure and assessment. In: M.R. Rosenzweig and L.W. Porter (eds.), Annual Review of Psychology, 31, 503-552. Palo Alto: Annual Reviews, Inc.

[48] Kazdin, A. E. (Ed.) (2000). Encyclopedia of Psychology. New York: Oxford University Press.

[49] Laibson, D. (2001). A cue-theory of consumption. Quarterly Journal of Economics, 116(1), 81-119.

[50] Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. Journal of Political Economy, 89(5), 841-864.

[51] Levine, D.K. (1998). Modeling altruism and spitefulness in experiments. Review of Economic Dynamics, 1, 593-622.

[52] Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology, 57(3), 426-441.

[53] McAdams, D. P. (1980). A thematic coding system for the intimacy motive. Journal of Research in Personality, 14(4), 413-432.

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

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

[56] McClelland, D. C. (1965). Toward a theory of motive acquisition. American Psychologist, 20(5), 321-333.

[57] McClelland, D. C. (1967). Achieving society. New York: Free Press.

[58] McClelland, D. C. (1985). How motives, skills, and values determine what people do. American Psychologist, 40(7), 812-825.

[59] McClintock, C. G., & Allison, S. T. (1989). Social Value Orientation and helping behavior. Journal of Applied Social Psychology, 19(4), 353-362.

[60] McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of disposition theories: Theoretical contexts for the Five-Factor Model. In: J. Wiggins (ed.), The Five-Factor Model of Personality: Theoretical Perspectives, 51-87. New York: Guildford Press.

[61] McDougall, W. (1932). The Energies of Men. London: Methuen.

[62] McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. Games and Economic Behavior, 10(1), 6-38.

[63] McKelvey, R. D., & Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. Experimental Economics, 1(1), 9-41.

[64] Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of disposition: Reconceptualizing situations, dispositions, dynamics, and invariance in disposition structure. Psychological Review, 102(2), 246.

[65] Morgan, J. (2000). Financing public goods by means of lotteries. Review of Economic Studies, 67(4), 761-784.

[66] Murray, H.A. (1938). Explorations in Personality. New York: Oxford University Press.

[67] 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, 30-71. Oxford: Oxford University Press.

[68] Potters, J., & Suetens, S. (2009). Cooperation in experimental games of strategic complements and substitutes. Review of Economic Studies, 76(3), 1125-1147.

[69] Przyrembel, M., Chierchia, G., Bosworth, S. J., Snower, D. J., & Singer, T. (2015). Beyond approach and avoidance: Towards a motivation-baed decision making model. Mimeo.

[70] Rabin, M. (1993), Incorporating fairness into Game Theory and Economics. American Economic Review, 83(5), 1281-1302.

[71] Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. Review of General Psychology, 8(3), 179-193.

[72] Roberts, B. W., & Pomerantz, E. M. (2004). On traits, situations, and their integration: A developmental perspective. Personality and Social Psychology Review, 8(4), 402-416.

[73] Robson, A. J., & Samuelson, L. (2011). The evolutionary foundations of preferences. In: J. Benhabib, A. Bisin, and M. O. Jackson (Eds.), Handbook of Social Economics, 221-310, Amsterdam: North-Holland.

[74] Rotemberg, J. J. (1994). Human relations in the workplace. Journal of Political Economy, 102(4), 684-717.

[75] Sandel, M. J. (2012). What Money Can't Buy: The Moral Limits of Markets. New York: Farrar, Straus and Giroux.

[76] Schultheiss, O. C., & Brunstein, J. C. (Eds.) (2010). Implicit Motives. Oxford: Oxford University Press.

[77] Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. American Economic Review, 97(3), 999-1012.

[78] Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. American Economic Review, 67(2), 76-90.

[79] Suetens, S., & Potters, J. (2007). Bertrand colludes more than Cournot. Experimental Economics, 10(1), 71-77.

[80] Sugden, R. (2000). The motivating power of expectations. Rationality, Rules, and Structure, 28, 103-129.

[81] Titmuss, R. M. (1970). The Gift Relationship: From Human Blood to Social Policy. London: George Alien & Unwin Ltd.

[82] Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

[83] Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and Competitive orientations: Theory and preliminary evidence. Journal of Personality and Social Psychology, 73, 733-746.

[84] Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. Journal of Economic Behavior & Organization, 81(2), 664-676.

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

[86] Wiggins, J. S. (1973), Personality and Prediction: Principles of Personality Assessment. Reading, MA: Addison-Wesley.

Appendix A: Details of policy exercises

Welfare framework

The total welfare function W is expressed

$$W = \bar{\eta}\left(\left(1-p_s^C\right)\left(\left(1-p_s^C\right)U^{cc}+p_s^C U^{cs}\right)+p_s^C\left(\left(1-p_s^C\right)U^{sc}+p_s^C U^{ss}\right)\right)+$$
$$(1-\bar{\eta})\left(p_c^K\left(p_c^K V^{cc}+\left(1-p_c^K\right)V^{cs}\right)+\left(1-p_c^K\right)\left(p_c^K V^{sc}+\left(1-p_c^K\right)V^{ss}\right)\right)+T.$$

where the term T represents any relevant lump-sum transfers.

We can collect terms and rearrange:

$$W = \bar{\eta}\left(\left(1-p_s^C\right)^2 U^{cc}+p_s^C\left(1-p_s^C\right)\left(U^{cs}+U^{sc}\right)+\left(p_s^C\right)^2 U^{ss}\right)$$
$$+(1-\bar{\eta})\left(\left(p_c^K\right)^2 V^{cc}+p_c^K\left(1-p_c^K\right)\left(V^{cs}+V^{sc}\right)+\left(1-p_c^K\right)^2 V^{ss}\right)+T.$$

Promotion of Cooperative settings

Static efficiency

Let us partially differentiate welfare W with respect to $\bar{\eta}$ (ignoring any second-order effects on dispositional types for the moment):

$$\frac{\partial W}{\partial \bar{\eta}} = \left(1-p_s^C\right)^2 U^{cc}+p_s^C\left(1-p_s^C\right)\left(U^{cs}+U^{sc}\right)+\left(p_s^C\right)^2 U^{ss}$$
$$-\left(\left(p_c^K\right)^2 V^{cc}+p_c^K\left(1-p_c^K\right)\left(V^{cs}+V^{sc}\right)+\left(1-p_c^K\right)^2 V^{ss}\right).$$

Whether this is positive or negative depends on whether the Cooperative or Competitive setting produces higher payoffs.

Taxes on competition

Interaction payoffs and static efficiency

We will use the small change in parameterizations to modify Table 1 in a straightforward way. Table 4 displays the contributions under the tax scheme.

		Other's motive	
		Self-interest (s) *motive*	*Caring* (c) *motive*
Own	s	$y^{ss} = \frac{a_K-\tau}{d_K+1}$	$y^{sc} = \frac{(1-\kappa)((d_K-1)a_K-d_K\tau_s)+\kappa b_K}{(1-\kappa)d_K^2-1}$
	c	$y^{cs} = \frac{((1-\kappa)d_K-1)a_K-\kappa b_K d_K+\tau_s}{(1-\kappa)d_K^2-1}$	$y^{cc} = \frac{(1-\kappa)a_K-\kappa b_K}{(1-\kappa)d_K+1}$

Table 4: Agents' contributions to competitive setting interactions under competition taxes

The tax revenue T_K can be expressed

$$T_K = \tau\left(1-\bar{\eta}\right)\left(1-p_c^K\right)\left(p_c^K y^{sc} + \left(1-p_c^K\right) y^{ss}\right),$$

meaning that the tax is levied only in the Competitive setting, on the most competitive agents (those with r and s dispositional types), to a degree that depends on their level of competitiveness, which in turn depends on their likelihood of meeting a c dispositional type.

Let us partially differentiate the welfare W with respect to τ_s (ignoring any second-order effects on dispositional types for the moment). It will clarify to further decompose this into the partial effect of the tax τ_s on welfare holding the y_{ij} fixed, and the effect of the tax on the y_{ij}:

$$\frac{dW}{d\tau_s} = \frac{\partial W}{\partial \tau_s} + \sum_{y_{ij}} \frac{\partial W}{\partial y_{ij}} \cdot \frac{dy_{ij}}{d\tau_s}.$$

Note that $\partial W/\partial \tau_s = 0$ since there are revenue-neutral lump-sum transfers involved. We know that the $\partial W/\partial y_{ij}$ are negative by equation 19, and we can see that $dy^{ss}/d\tau_s$, $dy^{sc}/d\tau_s$, and $dy^{cc}/d\tau_s$ are non-positive. While $dy^{cs}/d\tau_s$ is positive, we can note that y^{cs} goes up by less than the amount y^{sc} goes up by under our parametric assumption $d_K > 2$. Furthermore, due to the nature of the multiplicative term in equation 19, $\partial W/\partial y^{sc}$ is larger than $\partial W/\partial y^{cs}$ since $y^{sc} > y^{cs}$ for small taxes.

Effects on reinforcement of dispositional types

This result relies on the fact that $d\left(V^{cc} - V^{sc}\right)/d\tau_s =$

$$\frac{d_K^2\left(1-\kappa\right)^2\left(a_K\left(d_K-1\right) - d_K\tau_s\right) + b_K\left(d_K^2\left(1-\kappa^2\right)-1\right)}{\left(1-d_k^2(1-\kappa)\right)^2} > 0$$

and $d\left(V^{cs} - V^{ss}\right)/d\tau_s =$

$$\begin{aligned}
&\left(a_K\left(d_K^2\left(\left(d_K\left(d_K^2+d_K+2\right)+1\right)\kappa^2 - 2\left(d_K+1\right)d_K^2\kappa\right.\right.\right.\\
&\left.\left.\left.+\left(d_K^2+d_K-1\right)d_K+\kappa-1\right)-\kappa\right)\right.\\
&+b_K\left(d_K+1\right)\left(d_K\left(d_K\left(d_K\left(1-2\kappa\left(1-\kappa\right)\right)-\kappa\left(1-\kappa\right)+1\right)-1\right)-\kappa-1\right)\\
&-d_K\tau_s\left(d_K\left(d_K\left(d_K^2\left(1-\kappa\right)^2+4\kappa-3\right)+4\kappa-2\right)+2\kappa\right)\\
&/\left(\left(d_K+1\right)^2\left(1-d_K^2\left(1-\kappa\right)\right)^2\right)\\
&> 0.
\end{aligned}$$

Subsidies for cooperation

Interaction payoffs and static efficiency

We modify the calculations of Table 1 in a straightforward way to show the contributions of the agents under cooperation subsidies in Table 5.

		Other's motive	
		Self-interest (s) *motive*	*Caring* (c) *motive*
Own	s	$x^{ss} = \frac{a_C+\sigma_s}{d_C-1}$	$x^{sc} = \frac{(1-\kappa)((d_C+1)a_C+d_C\sigma_s)+\kappa b_C}{(1-\kappa)d_C^2-1}$
	c	$x^{cs} = \frac{((1-\kappa)d_C+1)a_C+\kappa b_C d_C+\sigma_s}{(1-\kappa)d_C^2-1}$	$x^{cc} = \frac{(1-\kappa)a_C+\kappa b_C}{(1-\kappa)d_C-1}$

Table 5: Agents' contributions to the cooperative setting interaction under cooperation subsidies

Note that we are now considering only the cooperation subsidy in isolation. The cost of the subsidy will equal

$$T_C = \sigma\bar{\eta}p_s^C\left(\left(1-p_s^C\right)x^{sc}+p_s^C x^{ss}\right),$$

meaning that the subsidy is given only in the Cooperative setting, to the least cooperative agents (those with s dispositional types) to a degree that depends on their level of cooperativeness, which in turn depends on their likelihood of meeting a c or r dispositional type.

Let us partially differentiate the welfare W with respect to σ_s (ignoring any second-order effects on dispositional types for the moment). It will clarify to further decompose this into the partial effect of the subsidy σ_s on welfare holding the x_{ij} fixed, and the effect of the subsidy on the x_{ij}:

$$\frac{dW}{d\sigma_s} = \frac{\partial W}{\partial \sigma_s} + \sum_{x_{ij}} \frac{\partial W}{\partial x_{ij}} \cdot \frac{dx_{ij}}{d\sigma_s}.$$

Note that $\partial W/\partial\sigma_s = 0$ since there are revenue-neutral lump-sum transfers involved. We know that the $\partial W/\partial x_{ij}$ are positive by equation 20, and we can see that each of $dx^{ss}/d\sigma_s$, $dx^{sc}/d\sigma_s$, $dx^{cs}/d\sigma_s$, and $dx^{cc}/d\tau_s$ are non-negative.

Effects on reinforcement of dispositional types

This result relies on the fact that $d\left(U^{cc}-U^{sc}\right)/d\sigma_s =$

$$\frac{b_C\left(1-d_C^2\left(1-\kappa^2\right)\right)-d_C^2\left(1-\kappa\right)^2\left(a_C+d_C\left(a_C+\sigma_s\right)\right)}{\left(1-d_C^2\left(1-\kappa\right)\right)^2} < 0$$

and $d\left(U^{cs}-U^{ss}\right)/d\sigma_s =$

$$\begin{aligned}
&\left(b_C\left(d_C-1\right)\left(d_C^3\left(2\kappa\left(1-\kappa\right)-1\right)+d_C^2\left(1-\kappa\left(1-\kappa\right)\right)+d_C-\kappa-1\right)\right.\\
&+a_C\left(d_C^2\left(d_C-1+d_C^2\left(1-\kappa\right)^2-d_C^3\left(1-\kappa\right)^2+\kappa-\left(2d_C-1\right)\kappa^2\right)-\kappa\right)\\
&-d_C\left(2\kappa+d_C\left(2+d_C^3\left(1-\kappa\right)^2-4\kappa+d_C\left(4\kappa-3\right)\right)\right)\sigma_s\Big)\\
&/\left(\left(d_C-1\right)^2\left(1-d_C^2\left(1-\kappa\right)\right)^2\right)\\
&< 0.
\end{aligned}$$