

A Wedding Toast Is Not a State of the Union Address: Other-Regarding and Others' Opinions Regarding Preferences As Relevant Factors in Shaping Human Preferences

Jerome Nikolai Warren (Universität-Gesamthochschule Siegen)

Introduction

There is increasing certainty around the idea that rationality, state dependence and other regarding preferences are not mutually exclusive.¹ The belief that they are itself derives from a mis-equivocation of the *Homo economicus* model with rationality. This is a mistaken belief. The *Homo economicus* model, which rests on a Benthamite understanding of human behavior as *self-interested*, *utility-maximizing* and chained to the "two sovereign masters" of pain and pleasure, is severely limited in its relevance to human behavior at best, mistaken at worst, and cannot be redeemed by accounting for limited information and uncertainty. We will discuss the most glaring defects of the model in the following two chapters, but we can satiate ourselves at present with the assertion that there exist certain *other regarding preferences* in human beings not reflected in the model, including a desire for *esteem* and certain elements of *altruism* that cannot be explained by self-interested utility maximization, and that these are often times influenced by external contexts (*states*).

The paper is organized into five chapters of varying length, with chapters two and three making up the largest share. Chapter one introduces the idea of rationality, and attempts to briefly analyze human behavior as it relates to rationality. Chapters two and three study extensively occurrences of other-regarding preferences in human agents. Chapter four looks at how these preferences are influenced by states (contexts), and chapter five synthesizes the findings of the preceding chapters and looks at the case for picking up more fully the arguments presented in the paper in mainstream economics research.

In the following, rationality as it relates to organic life, and particularly human beings, will be discussed in short. There is a lack of consensus in the social sciences today with respect to rationality. Certain branches, such as economics, make abundant use of it, whereas others, like psychology, often

¹ These terms will be defined below.

reject (what they feel are) its fundamental premise. As we described above, the interpretation of rationality that is rejected is often times a mistaken one, and there are obvious drawbacks to this high level of diversity in basic micro-foundations in the social sciences: contradiction, overlap of research, inability or constraint on trans- and multidisciplinary research, to name a few. A more thorough-going understanding of the rational actor model might provide a useful basis for a future synthesis of the social sciences, as Gintis (2009) argues.

Gintis (2009, 1), drawing on Morgenstern and Von Neumann (1953), describes the rational actor model simply as representing “an individual with consistent preferences.” The criterion of consistency is measured on the basis of three major assumptions, namely:

1. completeness of preferences: this assumption merely states that any event B is at least weakly preferred to itself. Consequences of completeness are reflexivity and exclusion.
2. transitivity of preferences: if $A \succcurlyeq B$; and $B \succcurlyeq C$; then $A \succcurlyeq C$.
3. independence of irrelevant alternatives: an individual's preferences for A & B do not depend on his/her preference for C.

These three conditions form the core of modern rational agent theory. They essentially allow the modeling of preferences as a continuous field and lie at the root of contemporary microeconomic theory.

1. Bayesian Rationality

In addition to the above conditions, we can state the conditions for Bayesian rationality, a relatively more demanding form of rationality that consists of preference updating via expected utility maximization, a fundamental building block of *homo economicus*. Given a lottery with n outcomes of the sort $i \in n$, whose individual probabilities, p_i sum such that $\sum (i \times p_i) = 1$, then

1. The preference for $a \succcurlyeq b$ depends solely on the states of nature" these two events or outcomes entail. In other words, if $a' = a$ and $b' = b$, where $a \neq b$ and $a' \neq b'$, then $a \succcurlyeq b \leftrightarrow a' \succcurlyeq b'$.
2. if P is a non-zero set from the underlying population of events and $a \succcurlyeq b$, then all outcomes where a occurs are preferred to outcomes where b occurs: $(f = a|P) \succcurlyeq P (f = b|P) \leftrightarrow a \succcurlyeq b$.
3. Non-wishful thinking: beliefs about the probability of an outcome are independent of the payoffs of that outcome. Analytically, if we prefer a to b and a' to b' , then we prefer an outcome O given event p (e.g., a coin toss) only if we prefer the outcome Q given given event r (e.g., a separate coin toss). Axiomatically:

If $O = a, b|p$, $Q = a', b'|r$, $O' = a, b|(1-p)$, $Q' = a', b'|(1-r)$, then $O \succcurlyeq O' \leftrightarrow Q \succcurlyeq Q'$.

4. Higher payoff outcomes are preferred: For any event p , if an outcome a has a higher payoff than a set of outcomes $b(\pi)$, with π being a subset of p , then a should be strictly preferred to b : $a \succ b(\pi)$, then $O = a|p \succ Q = b$.
5. Lastly, there is a condition that allows event subsets to be ranked ordinarily. For a set of outcomes a, b, a' and $b' \in \Omega$ where $a \succ b$, there exist disjoint sets x_i such that for any x_i if $a' = O$ for $\omega \in x_i$ and $a'(\omega) = a(\omega)$ for $\omega \notin x_i$, then $a' \succ b$.

Granted these five complementary assumptions, Bayesian rationality obtains.

2. Are People Rational?

The next question is, of course, whether the conditions outlined above obtain in human beings. In order to shed light on whether or not this is true, it might be wise to consider when they do not obtain and thereby deduce when they do. This we do in the present section. Indeed, an appreciation of the rational agents model or of *homo economicus* requires knowledge about whether people are or aren't rational. This is an uncontroversial statement. The answer to this query is, however, not so clear. Traditional economic models (*homo economicus* being the most widely known) have been based on the premise that individuals are 1) utility-maximizing, 2) self-interested 3) individual agents. Thus, the primary medium that models have depended upon is sometimes referred to as a "selfish get". While there is a great deal of evidence that people are by and large rational, the empirical basis for this rationality, and particularly assumptions 1-3 above, has not been seriously explored in the literature until relatively recently.

In the following, a number of results will be summarized. We will attempt to show that rationality and utility maximization are not necessarily coterminous.

2.1. Biological basis for rational behavior

Many would read a biological basis into rationality. There is also wide-ranging empirical data confirming rationality in human beings. This is the case even when apparent contradictions with rationality are present. For instance, there is a pronounced time in-consistency in much of human discounting behavior (Gintis 2009, 8). Time inconsistency refers to preferences by which the reward accrued by two different outcomes is discounted by a non-constant factor, i.e., the preference for a particular reward depends on the time from the present in which the reward is received. Human agents are generally time-inconsistent. This type of inconsistency is especially apparent when individuals are making choices with both long- and short term costs and benefits. Smoking is an example of this: smoking has immediate benefits in the calming effects of nicotine, but significantly increases the risk of future illness and disease. Many

nevertheless “irrationally” smoke, discounting future costs in favor of immediate benefit.

This type of inconsistency appears to be systematic in nature (Gintis 2009, 9). At closer inspection, smoking does not appear to be irrational at all, although it evidently contradicts the principle of expected utility maximization. If we include a preference for cigarettes and reflect a certain degree of “present bias”, it is easy to show that a rational agent may develop a preference for cigarettes (or unprotected sex, or any number of other “present biased” activities).

It appears that long-term regarding preferences are a relatively late development in the evolution of homo sapiens. Long and short term-decision making are in fact localized in different parts of the brain, long term-decision making being localized in the (much more recently evolved) pre-frontal cortex. It is notable that when this region is damaged, an excessive level of “present bias” is observed. Indeed, in general, human beings seem to reflect a certain degree of “present bias”. However, “humans are much closer to time consistency and have much longer time horizons than any other species, probably by several orders of magnitude” (Gintis 2009, 11).

There are a number of other systematic biases or contradictions *per se* to rationality of human behavior. *Status quo* bias, endowment and framing bias, and the use of heuristics in decision making. Most of these biases and contradictions can be dealt with by building them into the model of human preferences. This will be shown at present.

2.1.1. *Prospect Theory*

People tend to view value outcomes differently according to whether they involve losses or gains. This might be seen to be complementary to status quo bias. According to (Gintis 2009), this suggests an emphasis on change in position, rather than on magnitude or “level” of the present position. At face value, this seems to contradict rationality, but on closer inspection, it does not violate rationality conditions (Kahneman & Tversky 1979) suggest a solution to this difficulty. Given that we design our utility function to reflect the above emphasis (a bias towards change in position), we can design a function that in fact has a higher slope to the side of the graph representing negative gains (losses) than the side representing positive gains. In other words, we would wind up with a “kinked” function. There is a wide-ranging literature confirming the existence of “kinked” functions in a number of animal species, especially with regards to time (preferring the present to the past, etc.), so this is not so difficult to imagine.

In fact, numerous experiments confirm the existence of “kinked” utility functions in humans. Ultimately, accounting for this, and accounting for differences in the magnitude of loss aversion and in “time discounting” should go a long way to resolving apparent contradictions between human behavior

and conditions of rationality. Similar contradictions with regards to framing, endowment and the use of heuristics also turn up similar situations. Including parameters in the utility function that reflect such biases seems to eliminate most of them in so far as they contradict with rationality. The case of heuristics is a bit more complex, but we don't intend to deal with it at length here. For more on the role of heuristics in decision making, see (Gintis 2009, 26).

The same goes for other regarding preferences. Including these in individuals' utility functions should resolve many of the contradictions the rationality model faces in these situations. According to the model *Homo economicus*, people should in fact just consider others in so far as doing so maximizes their expected material payoffs. In the following chapters, we show that this is often times not the case, confirming the limitations of the model and the need for alternative models of human behavior.

2.2. Expected utility maximization in practice

As pointed out above, failures of the expected utility principle do not equate irrationality. We will outline at present why this is the case. Subjects who systematically violate the principal of expected utility often display preferences that can easily be modeled and incorporated into the rational actor model. Time discounting, "present bias" and "folk probability" are all examples of this type of violation. Other types result from ignorance of the choice set or the payoffs thereof. These two types of violations are, in turn, referred to as "systematic" and "non-systematic performance errors". Like an unpracticed opera singer who intends to sing a high-C, but falls off flat, performance errors can be dealt with via practice and education. Examples include programs to spur prudence and saving in the general populace.

Irrationality, on the other hand, would entail much deeper contradictions of behavior from the precepts outlined above and in Chapter 1. Such behavior is rarely found in the animal world, from the simplest multicellular organisms to the most complex vertebrae, including human beings.

The two most striking failures of expected utility maximization are the Allais and Ellsberg paradoxes. These both describe situations where individual behavior systematically deviates from rationality.² These will be discussed in some detail at present.

2.2.1. Allais paradox

The Allais paradox, first presented in (Allais 1953), involves a situation in which an individual presented with two lotteries makes a choice in both that, while individually uncontroversial, reveal a contradiction in utility maximization when taken together. This will be shown briefly: if an individual

² It should be noted that non-systematic deviations from rationality do not conflict with the underlying assumptions of the model.

is offered the following two lotteries, empirical evidence (see: Machina 1987, 136 for an overview) suggests that many individuals tend towards choosing 1a and 2b.

	Option a	Option b
Lottery 1	50 cents	.1*\$1+.89*50 cents +.01*\$0
Lottery 2	.11*50 cents + .89*\$0	.1*\$1+.9*\$0

Table 1. Allais Paradox.

Simply calculating the expected utility functions from these two choices shows a paradox. This means individuals systematically violate the higher payoff loving and transitivity conditions of the rational agent model. Gintis (2009) offers two possible explanations for this inconsistency. Firstly, an incorrect choice may encourage feelings of regret, thus leading to bending of the conditions of utility maximization. Secondly, it is possible that this is an example of status quo based loss aversion: an individual may see the “anchor point” of the two lotteries as being different, in the one it is 50 cents, in the other \$0. The differing anchor points might therefore lead to differing beliefs about likely outcomes.

Whatever the correct explanation, this may be more evidence for state-dependent preferences (discussed in the concluding chapter) than of inconsistency.

2.2.2. Ellsberg paradox

The Ellsberg paradox describes a different aspect of irrational behavior. Individuals are presented the choice of two urns, both filled with red marbles and white marbles. The distribution of the first urn is known, whereas the second urn's distribution is unknown. Individuals are given the choice to pick a marble twice from one of the two runs. Players are given some amount of money in the first round for one red marble and none for a white marble, and the same amount for one white marble and none for a red marble in the second round. Given the knowledge of the color distribution of the first urn, utility maximizing individuals would only select one marble from the first urn. This can be explained by the idea that, given the knowledge of the first urn's color distribution, we assume it has a higher level of red or white marbles, respectively. To increase my chances in the second round, I would reflexively choose a marble from the urn with less chance of having more of the wrong color. If I believed the second urn had more of the wrong color in the second

round, then I would have picked a marble from it in the first round, when the goals were switched.

Empirically, however, people tend to choose both marbles out of the first urn when the number of each color approaches, but is not equal to, fifty percent of the total. Rationality would dictate different behavior. To see this, let us assume urn 1 has a distribution of 49 white and 51 red marbles. The expected utility of choosing from urn 1 in the first round is $.51(\$1) + .49(0)$ (if \$1 were the amount in question). The expected utility of choosing from urn 2 would be $p(\$1) + (1-p)(0)$. The choice of a marble from urn 1 would imply that $p > .49$. In the second round, many again choose from urn 1, whose expected utility is $.51(0) + .49(\$1)$, and urn 2's expected utility is $p(0) + (1-p)(\$1)$. Choosing from urn 1 would imply that $p < .49$. The above is, of course, logically infeasible. Gintis (2009, 18) describes the Ellsberg and the Allais paradoxes as a performance errors. Subjects assessing the situation described by the Ellsberg paradox may simply assess "known" risk differently than uncertain risk. "Of course, with some relatively sophisticated probability theory, we are assured that there is in fact no such additional risk" (Gintis 2009, 18).

2.3. Charitable giving

Sugden (1982) describes some of the basic contradictions between the principle of expected utility maximization and social practice with reference to the issue of charity. He finds that theoretical levels of marginal utility derived from charitable giving are not congruent with many features of charitable giving found in reality. Namely, deriving a model with a few simple specifications congruent with the "public-minded" theory of charitable giving reveals that, in order to be plausible, individuals would either have to react strongly to changes in giving by others to charity (i.e., have a strong discount rate for charitable giving with respect to others' giving), or to approach 100% of allocation of additional income to charity (i.e., have a very low income elasticity for giving). Both of these outcomes seem highly improbable, given, as Sugden (1982) points out, that they imply that individuals would either increase their charitable contributions in reaction to lower contributions by others, or not spend any amount of additional income on personal consumption.

All of this points to flaws in the theory of expected utility, but one shouldn't completely discount it as a viable theory for explaining any aspect of human behavior. There is a difference between a theory's failure to provide universal insight and its complete irrelevance. A useful undertaking of future scholarship will be to determine where the principle of expected utility maximization does and does not obtain.

3. Nash Equilibrium & the Prisoners' Dilemma

One of the fundamental concepts of modern economic theory is that of the Nash equilibrium. Assume two individuals exhibiting the above conditions of Bayesian rationality, who play a game in which the payoffs are as follows: A long-standing debate has ensued in the social sciences as to the “dominant” strategy for each player in the game. It is true that, were both players to be Benthamite hedonists, there would be little possibility for any other strategy besides defection. This can be seen by viewing the dominant strategy for player 1 if player 2 chooses cooperation. In this case, player 1 would receive 50 cents by cooperating, yet could expect \$1 if s/he were to defect. Therefore, a Benthamite hedonist would choose defection. The second player's response to this best response would also be to defect, such that the inevitable outcome would be that each receives nothing. This is the case because no matter how we run our thinking process, a hedonistic, payoff maximizing individual will always choose the defect strategy. It forms what has been come to be referred to as a Nash equilibrium, which is a stable equilibrium that is arrived at in a non-cooperative game (a game without collusion) where both individuals can gain nothing by individually switching strategies.

2 \ 1	Cooperate	Defect
Cooperate	50 cents	(1,-1)
Defect	(-1,1)	0

Table 2. Nash equilibrium.

Of course, this interpretation of events is exceedingly misleading, as it is unlikely that people behave like Benthamite hedonists in most cases in reality, save for a limited number of settings (such as in criminal interrogations). If it were to generally obtain, it is arguable that such behavior amounts to sociopathy where found, as Gintis (2009) argues. Ethical and moral principles ultimately prescribe we behave in certain ways, and we tend to apply these pretty strictly, often disregarding material payoffs to uphold moral principles. As such, significant deviations were found between actual behavior and that prescribed by the Benthamite view, both in field experiments and in the laboratory (Henrich *et al.* 2004). We get into some of these in the following chapters.

4. Conclusion on rationality

It would appear that our fundamental discoveries in this chapter have been that rationality is a concept often misunderstood in social science research. Moreover, the argument has been attempted to be established that including a particular model of rationality drawn from microeconomics literature into the foundations of modern social theory would greatly benefit the coherence and consistency of social science scholarship. Evidence has been offered that human beings, like most other animal species, generally follow rational behavioral precepts, and where they do not, adequate measures can be taken in many instances to reflect systematic deviations from underlying principles (such as utility maximization). One conclusion we have hoped to have evidenced in the above is that the first underlying premise of the *homo economicus* model, namely utility maximization, is often invalidated in practice. This does not, of course, mean that the model is useless or incorrect. As has been pointed out above, human beings are in general an order of magnitude more consistent in their preferences than other animal species. Yet, an understanding of the limitations of a model is important in order to know where it does apply.

Altruistic Preferences

By “social preferences”, we mean certain implicit or explicit desires or preferences that do not conform to the expectations of a model based on self-interest. These behaviors consist of feelings (and the actions they motivate) of altruism, fairness, reciprocity and inequality aversion (Fehr & Camerer 2007), as well as a general concern for the thoughts of others (Ellingsen & Johannesson 2007). We refer to the first type in this paper as “altruistic preferences”, and the second as “esteem preferences”. Kropotkin (2012), a seminal author in the modern scientific study of human reciprocity, speaks of this type of behavior as existing in numerous animal species, including bees:

These small insects, which so easily might become the prey of so many birds, and whose honey has so many admirers in all classes of animals from the beetle to the bear, also have none of the protective features derived from mimicry or otherwise, without which an isolated living insect hardly could escape wholesale destruction; and yet, owing to the mutual aid they practice, they obtain the wide extension which we know and the intelligence we admire, By working in common they multiply their individual forces; by re-sorting to a temporary division of labor combined with the capacity of each bee to perform every kind of work when required, they attain such a degree of well-being and safety as no isolated animal can ever expect to achieve however strong or well-armed it may be (Kropotkin 2012).

Indeed, we will argue presently that one can view this sort of behavior in humanity as well, and that the existence of such behavior as a general rule runs contrary to the precepts of Benthamite utilitarianism (i.e., that human agency consists primarily of self-regarding preferences). If we are in fact successful in demonstrating that individuals do not always (or even generally) maximize own payoffs instead of holding to some other principle of agency (such as rule-based utilitarianism, Kantian de-ontology, etc.), then the case for retaining self-regarding preferences as the *sin qua non* of human social preferences is severely weakened.

Many popular and iconic historical examples of self-sacrifice and other deviations from self-regarding preferences exist. One example is the figure of Pheidippedes, a courier in ancient Greece who supposedly died after running a little more than 42km in order to relay the news of the Battle of Marathon, which the Athenians had won. Likewise, one can view the actions of German soldiers during the last stages of WWII, who maintained a fortified “corridor” for East Prussian civilians to escape as depicting these preferences. Self-regarding bees, couriers and soldiers in a losing army would not concern themselves with the well-being of others insofar as their own personal material welfare was unaffected. A self-interested bee would not sacrifice its life by discharging its stinger's venom. It would wait for another bee to do this, thus “free riding” on that other bee's generosity. A self-interested courier would take his or her time informing the Athenians of their military victory. Likewise, self-interested soldiers in a losing army would abandon their posts, and not worry about any “corridors” for civilians.

Social preferences are not restricted to bees and couriers and soldiers. They are clearly visible in everyday human behavior. They repeatedly occur in experimental settings established to measure general strategic interaction³. In the following, we will address empirical evidence for the existence of social preferences, both in experimental and non-experimental settings. We generally follow Ellingsen and Johannesson's (2007) division into “other regarding” and “other's thought regarding” social preferences, and restrict our-selves primarily to the first in this chapter, moving to the latter in the following one. First, we will attempt to address what, if any, origins social preferences have.

5. Where Social Preferences Come From

The real origins for social preferences (in human beings) likely lie back in the linguistic, biological and cultural origins of *Homo sapiens*. Certainly, the existence of such values would puzzle anyone with a basic understanding of natural selection. Darwin himself wrestled with the idea:

³ Though the applicability of such results might be limited (Bowles & Polania-Reyes 2012, 11).

But it may be asked, how within the limits of the same tribe did a large number of members first become endowed with these social and moral qualities, and how was the standard of excellence raised? It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents, or of those who were the most faithful to their comrades, would be reared in greater numbers than the children of selfish and treacherous parents belonging to the same tribe. He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature. The bravest men, who were always willing to come to the front in war, and who freely risked their lives for others, would on an average perish in larger numbers than other men. Therefore, it hardly seems probable that the number of men gifted with such virtues, or that the standard of their excellence, could be increased through natural selection, that is, by the survival of the fittest; for we are not here speaking of one tribe being victorious over another (Darwin 2004).

Inadvertently, Darwin may have stumbled upon the answer. The problem, it seems, may lay in the fact of attempting to distinguish inter and intragroup advantages, whereas the influence may be quite significant, from one to the other. And, thus, Darwin continued later:

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase (Darwin 2004).

This circumstance – low individual fitness but high social or “deme” fitness – as a prerequisite to the evolution of altruism is famously captured by taking Hamilton's formula for inclusive fitness, which states the precondition for the evolution of altruism as consisting of higher benefits than costs to the behavior in question, and taking a sum over a population. Axiomatically, the formula reads

$$rb > c$$

where r is the relatedness of the individual upon whom one is conferring a benefit, b is the level of benefit of the action to the individual and c is the cost

to the individual carrying out the action. This means that summing the formula for i groups of j individuals would entail

$$rb_{k6=j} > c_{i;j}$$

It is easy to see that, according to the formula, the ability for altruistic behavior to develop and evolve would necessarily be limited to societies with a high level of relatedness, and would even there be quite limited to actions that confer extraordinary benefit to the recipient when measured with respect to the costs accrued. We will return to this issue momentarily. To illustrate, let us take a relatively high level of median relatedness, .2, and a benefit to cost ratio of .2, and a benefit to cost ratio of 5 to 1 has to occur for altruism to be able to spread.

It is important, it would appear, to devise a thorough and exhaustive view of the origins of altruism, and merely stating that it depended on close-knit communities does not seem to capture the sensitivity of altruism to outside conditions. Although, as Axelrod and Hamilton (1981) discuss, there may be certain stability conditions that allow a certain gene-cultural mix to persevere in the face of "invading" types, it is likely ulterior factors are at play in maintaining altruism as a viable behavioral type. The mechanism by which these factors are relevant for economic discourse, in that the existence of markets and modern states requires a large degree of trust and reciprocation among citizenry. This makes a nuanced view of the evolution and maintenance of reciprocal altruism (altruistic behavior that results from the expectation that it may be reciprocated in the future) quite central to modern economic theory, writ large.

5.1. Cosmopolitan ancestors

Indeed, the premise that our ancestors lived in close-knit groups appears quite suspect. Moreno-Gómez *et al.* (2009) discuss the potential for the existence of cosmopolitan ancestors. There is considerable evidence for this theory. The archaeological record Ambrose (1998), for instance, speaks for significantly less genetic diversity than the parochial, close-knit communities individuals certain biologists, like Dawkins, assume to have preceded modern man would in fact possess. Additionally, ethnographic study of contemporary "pre-contact" societies⁴ reveals a high degree of continual and consistent contact with outside groups (trade and warfare⁵). Indeed, the high degree of conflict (Bowles & Gintis 2011, 102)⁶ and the extreme level of climatological volatility

⁴ See Salzano *et al.* (1977) for a seminal study in the field; Lins *et al.* (2010) and Henn *et al.* (2011) are more a more recent studies making similar claims.

⁵ In fact, migration and trading of individuals between groups seems to be quite frequent among such societies.

⁶ Indeed, there is a good argument that warfare likely strongly contributed to the evolutionary stability of altruism, via parochial altruism, which we don't mean to

that accompanied the late Pleistocene (Andersen *et al.* 2004) would have necessitated frequent and sometimes extreme levels of migration and interaction.

On a more abstract level, when a number of repeated generation computer simulations were run, it was shown that even mild levels of perceptual error, that is, error in interpreting behavior and relatively innocuous cost-benefit ratios (Hamilton) reduced quite extraordinarily the upper limit of group members allowing for altruistic behavior to develop. Even for a relatively low perceptual error rate of 2%, maximum group size where altruistic behavior forms a dominant equilibrium ranged from 2-8, much smaller than the size we understand groups to have consisted of in the late Pleistocene/Early Holocene, which was likely closer to 80 (Moreno-Gómez *et al.* 2009). Another explanation for the development of altruistic behavior is therefore necessary.

5.2. Multi-level selection

The theory Wilson and Sober (1994) advance to further explain for the proliferation of altruism in society in general introduces the notion of payoff differences among groups (“demes”) and the evolutionary (dis)advantages that these differences may involve. According to this process, groups or demes that display certain traits are more likely to survive than those that do not. In effect, members of these groups have an evolutionary advantage over members of other groups. This is the case even though individual members of the deme who have the trait are at an evolutionary disadvantage. This process is referred to as “multi-level selection”, meaning that the evolutionary process occurs on both the inter and intragroup levels.

According to the theory of Wilson and Sober (1994), three conditions are required in order for altruism to be considered an evolutionarily stable trait despite the theory of inclusive fitness: 1) that group life is essential to human survival, 2) that successful groups tend to destroy and/or absorb less successful groups, and 3) that groups containing higher levels of altruists tend to cooperate more and that cooperative groups tend to persevere in the case of frequent inter-group conflict (and, by extension, to survive the severe environmental crises that characterized early human civilization). Certainly, to validate this theory one would have to find evidence of such multi-level selection in practice. This is what Bowles and Gintis (2011) in fact attempt to do. Their theory contradicts what has been a dominant position in the social sciences for decades at least.

Bowles’ and Gintis’ (2011) theory consists of a number of co-evolutionary processes that may have contributed to the evolution of altruistic behavior. These will be outlined in the following.

address at present, but which provides a very compelling assessment of the facilitator of the evolution of cooperation and altruism.

5.3. Selective extinction

The process of selective extinction consists in the advantage over and against other groups that a group possessing a certain trait might have. In terms of social preferences, while the social individual may incur a selective loss in comparison with the less socially inclined, there may in fact be a between-group advantage for groups featuring socially inclined individuals. This might be the case in case of frequent or severe inclement conditions or frequent inter-group conflict.

$$\frac{(P)_j}{P_{ij}} \neq \frac{i}{G}$$

We can represent this effect by the inequality where $(P)_j$ is the variance of population within one group or deme with respect to the trait in question, P_{ij} is the inter-group population variance with respect to the trait, i is the individual cost of switching from not having to having the trait and G is the change in the level of fitness of the j trait in any group with respect to a shift in the number of individuals displaying that trait. Being that we are concerned here with social preferences, of which altruism is the most iconic, we will assume that the parameter i is negative, and G positive.

In order for selective extinction to bestow a particular trait with an evolutionary advantage, this inequality has to hold. The inequality would hold when there is a large intergroup diversity in types (e.g., when there are many different levels of cooperation between groups) and/or when within-group variance is very low (e.g., when there are groups of homogeneous altruists and others of homogeneous sociopaths). As such, a trait like cooperation can spread by increasing the fitness of groups possessing it, so that, in practice, inter-group conflict leaves these groups with a higher chance of success, thus passing on the trait to further generations (and, in effect, spreading by colonizing or subjugating groups with a greater likelihood that do not feature the trait⁷. It in effect works on G .

5.4. Reproductive leveling

Reproductive leveling consists of measures which in fact reduce the payoff difference between traits. They, in turn, are comprised of conventions that are in every group member's interest to uphold, as long as others uphold them. As such, they are an example of reciprocal altruism. As an example, a trait like the sharing of food or land redistribution has obvious evolutionary disadvantages to the individual, as property belonging to individuals is withdrawn without direct material compensation. However, if benefits of this type of behavior for the group are randomly distributed within the group, and groups endowed with a high number of individuals who share food have a selective advantage

⁷ Bowles and Gintis (2011) cite this as a potential reason for the spread of European culture in the last five centuries.

over groups that do not, then these two facts themselves spell out two distinct effects that reduce the former disadvantage. The first includes the term n^b , where n is the group size, representing the possibility of any altruistic individual benefiting from his or her own altruistic behavior. The second is a bit more complex. It involves estimating the total effect of the switch from not having the trait to having the trait, and not simply viewing this switch in isolation, where the effect is only represented by a cost parameter, c .

Ultimately, the effect of switching from not having to having the trait, i described above, can be generally represented by the following equation, taken from Bowles and Gintis (2011, 118):

$$i = \frac{dw_{ij}}{dp_{ij}}$$

where dw_{ij} is the reproductive fitness of members with trait j in group i , and the term in the denominator on the left-hand side is the fraction of deme i who have the altruism trait. The dw_{ij} is, in turn, equal to:

$$c + b + \frac{\kappa}{\lambda_A}$$

where the first term represents the first effect discussed, and the last term the second effect. κ is a sort of “violence” or “volatility parameter”; λ_A is the change in group survival chances with respect to a change in the percentage of members displaying altruistic behavior. If a particular group engages in reproductive leveling conventions, these clearly reduce the cost to the individual of engaging in altruistic behavior. The above equation becomes:

$$\tau_r c + b + \frac{\kappa}{\lambda_A}$$

where τ_r is the extent of reproductive leveling. The equation represents the fact that, in the presence of reproductive leveling, the cost of altruistic behavior is a declining function in the percentage of group members engaging in reproductive leveling.

5.5. Selective assortment

If we assume that the distribution of altruism is not random, then it is imaginable that some level of self-selection occurs with respect to this parameter. Self-selection and migrational patterns might then theoretically lead to a distribution where a large number of homogeneous altruist and sociopath communities exist. This behavior is referred in the literature to as “selective assortment” (Bowles & Gintis 2011, 122), and could certainly contribute to a stable equilibrium where altruistic traits spread, via the above detailed effects contributing to selectional advantage.

6. Genetic & Cultural Evolution of Altruism

Bowles and Polania-Reyes (2012) suggest that social preferences likely originate to a great extent from social contexts such as parental influence, religious and community affiliation, socialization and on a number of other exogenous factors (what Bowles and Polania-Reyes (2012) refer to as “baseline values”). Additionally, they argue, certain preferences are endogenous, a sort of preference which Bowles and Polania-Reyes (2012) explain as a type of long-term “preference updating”, featuring elements of persistence beyond the short term⁸.

There are likely also neurological bases. Fehr and Camerer (2007) report activation of the ventral striatum (in communication with a human cooperator, as well as in correlation with offers perceived as “fair” in experimental settings), dorsal striatum (in punishing defectors in Prisoner Dilemma games), nucleus accumbens and the orbitofrontal cortex (in men in accompanying punishment of unfair partners), ventral tegmental (in response to either receipt of monetary reward or the act of donating to charity⁹). Most of these areas of the brain are specifically involved with processing reward, and are interestingly often activated both when giving and receiving (in experimental settings) endowments.

Likewise, a number of areas of the brain implicated in both reward, motivation and executive functions have been found to activate in situations where social preferences are likely active. Activation in the ventromedial prefrontal cortex (PFC), (in situations of punishment for behavior perceived as “unfair”) in the dorsolateral PFC, the anterior insula and the anterior cingulate cortex (in the rejection of low offers). All of these areas of the brain are involved variously in risk assessment, reward and motivation, pleasure, fear, stress and conflict resolution and are likely implicated in parsing conflicts between social preferences and self-interest, as Fehr and Camerer (2007) suggest.

7. The Issue of Free Riding

Institutions that rely on a great deal of social preferences, such as shrimp fishing, jury deliberations, taxation and countless others, rely to a greater or lesser extent on compliance. In some cases, such as taxation, it may be more or less possible to make credible threats to the effect that noncompliance is tied

⁸ Bowles and Polania-Reyes (2012) describe the endogeneity by referring to the authors’ residence in Italy as potentially affecting long-term food preferences: “It might be that we have newly come to enjoy the taste of pasta, perhaps through extensive exposure to it while in Italy.” This would be an example of “preference updating” tantamount to preference endogeneity.

⁹ Apparently more generous giving was correlated with higher activity in this region, suggesting that “giving has its own reward” (Fehr & Camerer 2007, 422).

to severe repercussions, but in other cases few such mechanisms exist (or where they do they are ineffective). In such cases, we are said to face the problem of free riding. Free riding consists of all incidents where public (non-excludable, non-rivalrous) goods are used without proper (pecuniary or non-pecuniary) compensation.

The problem of preventing, eliminating and punishing free-riding is a serious one, and one which the present work is not extensive or involved enough to deal with in its entirety. We will suffice ourselves with mentioning at present a few examples where free riding might occur in order to provide a general picture, and return to the issue in a future section. Fehr and Leibbrandt (2011) discuss a clear case where free-riding presents a real problem. They studied two groups of Brazilian shrimp fishermen. There is a convention among shrimp fishermen to cut holes in the traps. If holes are too small, not enough young shrimp escape in order to repopulate the depleted areas.

It is easy to see that each individual fisherman has an incentive to cut small holes as long as everyone else is abiding by the convention. This is an obvious example of a free riding problem. The same case holds for paying taxes and a large number of other behaviors where private and public utility are not congruent.

8. Other-Regarding Preferences in Practice

We see many examples of other regarding preferences in practice. A few we will address in more detail in the next section, but it may be worth noting certain practical examples of the first type of social preference at present. Definitive evidence of social preferences was found in a so-called gift game, wherein principals played the role of employers, and agents those of employees. In the version of the game conducted by Fehr *et al.* (1998), significant deviations from self-regarding utility maximizing behavior were recorded, both on the side of “employers” and “employees”. Additionally, numerous trials with ultimatum games reveal distinct deviations from the behavior prescribed by the *Homo economicus* model. Camerer and Thaler (1995) found significant evidence of “manners” in decisions in both determining and accepting (or not accepting) distributions of the initial endowment in the game. Offers were typically significantly more than the minimum allowable (the average was around 50%), and positive offers were frequently rejected when considered “low” or “unjust”. Other ultimatum experiments confirm these findings (see, for instance, Oosterbeek *et al.* (2004) for a meta-study).

Differently from the ultimatum game, there is no ability to reject offers in the dictator game. This means the esteem effects discussed below play no role in such cases. Nevertheless, significantly positive distributions are the rule

here, in general. In experiments conducted by Andreoni and Miller (2002), the average distribution of the endowment was more than 25%, far above the 0% hedonistic utility maximization would prescribe. We can expect that the difference in distributions between ultimatum (50%) and dictator (25%) games to be attributed to some form of “reputation” or esteem valuation, or to fear of rejection. We will discuss this in more detail below.

9. Conclusion

In this chapter, it has been our goal to show that another element of the *Homo economicus* model, self-regarding preferences, is deeply flawed. Human beings do, in fact, possess certain social preferences that include a regard for others that has both an economic, as well as a moral character. The economic character includes the efficiency which other regarding beliefs and behavior facilitate, as well as the implications of certain “social resources” such as esteem on behavior. The moral dimension stems from a heuristic or rule that certain communities abide by that in fact is unable to be adequately represented in the standard utility function. “We are nice because we are nice people.”

Moreover, we have summarized some of the main findings of biology, anthropology and archeology as they relate to the development and evolution of social preferences, including outlining some mechanisms by which altruism came about, such as multi-level selection and co-evolution. We have outlined some of the main pitfalls to social preferences, including free riding and the seeming paradox of charitable giving. In the following chapter, we will widen the scope of our analysis to include esteem and the general consideration of “the thoughts of others” as they influence our own preferences.

Esteem Preferences

The following chapter will outline the relevance of esteem preferences for economics research. Esteem preferences can loosely be understood to refer to preferences regarding “what others think”, following Ellingsen and Johannesson (2007). We first assess what the origins of esteem are, and address what it is (and therefore is not), and follow this with a discussion of the hedonistic paradox, a circumstance that obfuscates esteem from working like a traditional good. We then outline a possible economy of esteem, following Brennan and Pettit (2004), and then describe several problems facing it.

10. Origins & Characteristics of Esteem

Esteem is a word rich with many meanings. As such, it holds to Nietzsche's claim that "words are like pockets: they hold now one thing, now another". Esteem should be seen as a sentiment of approbation felt towards a particular individual. In common usage, one speaks of both "self-esteem" and the esteem of others, what is referred to as "social esteem". The latter is the core concern of the economic study of esteem, though as Brennan and Pettit (2004, 16) argue, the two are closely correlated. Though this connection is of more than precursory significance, we do not intend to expound on it here.

The origins of esteem are of recent interest in the psychological and sociological literature. Indeed, the entire field of moral epistemology has arisen in recent years, wherein a chief aim is to discover why it is collectives develop a sense of good and bad behavior, and how these are engendered in individual agents. Certainly, esteem alone would be a poor answer as to the 'what' of human morality, or even its ends, but it is likely a good starting point – and perhaps a good model for capturing – the 'how' of how seemingly unlikely developments in behavior and morality occur and spread throughout populations.

We have already discussed in the last chapter how activities which themselves have no or even negative utility payoffs for individuals can potentially benefit the group as a whole, thereby potentially having desirable social outcomes that outweigh the costs to the individual of performing the action. The question of the motor or facilitator of the spread of these types of behavior could potentially be addressed by the existence of esteem. For example, the effectiveness of gossip at eliciting socially desirable behavior is itself likely the result of the evolutionary advantage of having a "gossip receptor", broadly seen. For theoretical and empirical data on the origins of morality (and, by extension, of esteem), see, *inter alia*, Kurzban and Leary (2001), Kurzban and Houser (2005) and De Waal (1996).

Brennan and Pettit (2004, 2-15) describe esteem as being an attitudinal good, which are "goods (...) that come into being by virtue of what people think and feel about the person esteemed: that is, by virtue of their attitudes rather than their actions." However, esteem is not merely an attitudinal good. It consists as well of evaluative components, i.e., judging or "ranking" individuals according to either carrying out – or carrying out to a certain extent – an action. In addition to this, esteem also comprises comparative components, i.e., the placing into a specific context – for instance, that of average and ideal performances – of estimable actions. For instance, I cannot disesteem a plumber for his poor pedagogical skills when comparing him with a high school teacher, but I can esteem one teacher more than another, for the reason that their career aspirations as teachers should be fairly similar: i.e., they are comparable. Giving one teacher a higher level of esteem naturally eliminates the ability to esteem at this level of esteem all

teachers who perform at less than this level. Finally, esteem consists of directive components, i.e., being directed at actions in particular, and generally those with regard to which the individual being judged has some degree of self-control. I don't esteem someone in total, but in so far as they engage in activities deemed estimable.

11. The Hedonistic Paradox

The attitudinal quality of esteem is further colored by the fact that, given that esteem is a good, it is not a “normal” good: it cannot be given as a gift or traded voluntarily, there-fore one cannot let off or eliminate “excess esteem” (pardon the pun) by supplying it via an external market, and so on [Brennan & Pettit 2004, 51]. A voluntarist quality of esteem is, in fact, virtually non-existent, according to Brennan and Pettit (2004), although this supposition has been called into question by at least one author¹⁰. If the sup-position holds, this means that I cannot freely (cf. arbitrarily) confer esteem upon others. Ultimately, these attitudinal dispositions are given, to a large extent, exogenously: “we cannot help but be sources of light or gloom in [the] lives [of those whom we esteem or disesteem]” (Brennan & Pettit 2004, 54). This, combined with the fact that the active seeking of esteem usually results in the opposite reaction (namely, disesteem), creates a problem which in philosophy is referred to as the “hedonistic paradox” or what Brennan & Pettit refer to as the “teleological paradox”: like pleasure or fear or any number of attitudinal goods (bads), one cannot offer or receive esteem directly¹¹.

To illustrate this, take the example of pleasure. In most cases, it is true that one receives pleasure from doing something pleasurable. Pleasure is then an indirect consequence (a payoff) of engaging in the activity. This is easy to imagine. Picture the person smiling and laughing while eating an ice cream cone with a friend on the ledge of a canal on a sunny day in summer. The pleasure in this case is derivative of eating the ice cream, being with the friend, and enjoying a sunny summer's day. Seeking pleasure per se (*plaisir pour la plaisir*) is more difficult to imagine: “It is incoherent to think of making spontaneity a targeted goal: spontaneity consists precisely in not having such a self-focused aim” (Brennan & Pettit 2004, 36).

The same thing that obtains for pleasure or spontaneity is arguably also the case for esteem. Demanding esteem directly is a contradiction in terms as demanding spontaneity or pleasure is one. Not only is this a contradiction in terms, it is also likely self-defeating. The reasoning behind this is clear: someone who plays a virtuoso piece merely for the purpose of impressing an audience reduces or even eliminates the esteem he may have

¹⁰ See Cowen's review of Brennan and Pettit's book (Cowen 2005).

¹¹ For the problem of “open demand” of esteem, see (Brennan & Pettit 2004, 36), for the case of supply, see (Brennan & Pettit 2004, 58).

accrued, as soon as his true intent is revealed. However, on the other hand, little or no publicity removes the ability for the individual to have his/her behavior esteemed in the first place.

Ultimately, these conditions mean that esteem works at first sight like a scarce good in an economy, though not a normal one, by any measure. The paradox we've just outlined ensures that the economic exchange of esteem must find more indirect means in order for flows to occur in any direction. We outline these means at present.

12. An Economy of Esteem?

If esteem is to be studied in the domain of economics, it first must be established if and how it fits within the strictures of the discipline¹². As Brennan and Pettit (2004, 2) state, "Economies are systems whereby scarce resources are allocated among competing parties". They continue: "In particular, they are systems of allocation that have an interactive or aggregative dimension. What individual agents do, gives rise to aggregate patterns that feed back in turn into the things that determine what individual agents do". It seems, then, that the requirements of analyzing esteem within the economics discipline are that one or more of the following holds: Esteem is a scarce resource; it is allocated among competing parties; and it has an interactive and an aggregative dimension. We have already shown that the first is the case. When I give esteem to one individual, I automatically forgo giving an equal level of esteem to all other individuals belonging (by my judgment) to the same category performing at inferior levels to this person. The second condition obtains from the fact that the first condition obtains and from the fact that esteem is desirable. Indeed, it is difficult to imagine the first two conditions obtaining without the third also holding true. We will talk more about the aggregative components of esteem when we discuss esteem associations.

For there to be an economy of esteem, furthermore, there must be both a demand and a supply of esteem. We turn to these issues next.

12.1. The demand for esteem

The demand for esteem works around the impasse created by the paradox outlined above by retreating from the grand boulevards and wide avenues of active demand" (like the demand for expensive watches from renowned Paris boutiques), and settling in the subterranean depths of virtual demand (Brennan & Pettit 2004, 40). Although esteem acts as an underlying motivating factor at these depths, it does so only indirectly, via intermediary

¹² It is likely that even if there is no direct or explicit way to model esteem in economic terms, the ability for discourse to ensue may still be possible, a la Posner's "law & economics".

wants and desires. We can view this situation analytically by considering the following: imagine an individual with a demand for x_i different goods of the sort $\sum x_i$. We imagine further that the demand for x_1 cannot be fulfilled directly, but only via a secondary good, of the sort x_{1+i} . If the particular good, say x_2 , does not accomplish the desired effect of increasing esteem, the individual attempting to maximize his or her own esteem selects another activity of the sort x_{2+i} . As a practical example: if the (indirect) desire for esteem causes one to take up playing basketball, the inability to succeed at this latter task may cause one to develop a demand for learning chess. The latter demand, while to a certain extent autonomous, is nevertheless called into existence by the fact of the failure of basketball as contributing to an acquisition of esteem. The demand for esteem is generally one that can be included in utility functions in this indirect way.

12.2. The supply of esteem: Esteem services

Brennan and Pettit (2004, 56) state that the primary means of regulating or “meeting” the virtual demand for esteem are so-called esteem services. These are actions that largely serve as proxies for the allocation of esteem. Brennan and Pettit (2004) describe three main categories of esteem services: attention, testimony and association. Attention is a prerequisite for the supply of esteem. Attention is also a scarce resource, so (in this con-text) it also falls under the rubric of an economic good. In many instances, attention may itself lend a form of esteem, as the famed “Hawthorne effect” suggests.

Testimony consists of publicly making certain statements regarding the beliefs of one individual for another, thereby increasing the likelihood of esteem effects for the affected individual. This works also in reverse: the refusal to state negative views of what is viewed as a poor performance can also be seen as an esteem service, by mitigating disesteem. Association consists in the proximity with which one is seen with the individual in question. Esteem associations, which we discuss shortly, can serve the role of facilitating this (but also the fulfillment of the other services mentioned above). Each of these services can work in either direction: esteem as well as disesteem can be facilitated via esteem services.

12.3. The intangible hand

No discussion of the economy of esteem is complete without mentioning the intangible hand. Like Adam Smith's invisible hand of the market for goods, and Weber's iron hand of state power, the intangible hand deals in ways that are not immediately tangible to achieve some end that is neither intended nor planned by parties involved. Different from Smith's invisible hand, however, the intangible hand deals primarily with attitudes. Indeed, it has in many ways the potential to shape and influence how individual agents act and react in certain situations, even influencing the functioning of markets and state

power. In a nut-shell, the intangible hand can be understood as follows: if the above outlined economy of esteem exists, then people's natural desire for esteem will lead them to behave in ways that are conducive to impressing the suppliers of esteem in ways that cause desirable out-comes for all individuals involved.

It is important to note that all of this occurs before any regulation or impositions by government or other third parties. As we will point out in the following chapter, it is possible that incentives and mechanisms are put into place that alter what is perceived to be “good” or “normal behavior” in certain circumstances. This can work in several directions, as will be pointed out. A few examples include so-called “broken-windows” policies and policies to fine littering. Used correctly, such institutions can achieve socially desirable outcomes. But the potential exists to misuse them and to undermine the desired ends.

13. Esteem Associations

One way to organize the allocation of esteem services are so-called “esteem associations”, which are institutions that organize the allocation of esteem in a dimension of shared interest or competence. A professional association or a trade union are examples of esteem associations. Esteem associations work using a number of means in order to accommodate the desire for esteem of all their members. Of course, granted that esteem is a scarce good, the association must coordinate several ethical problems, such as whether certain members are more deserving of esteem than others, or whether esteem should be distributed relatively evenly throughout the organization, and others.

One of the main tools that most esteem associations have to their disposal is the fact that they are voluntary in nature. This means that members have the right of exit, and the association, moreover, has the right of expelling members under certain circumstances. Indeed, part of the repute that many such organizations have stems directly from the selective nature membership entails. Ivy League colleges are an example of this, where no doubt the highly selective nature of the school plays at least some role in propping up the reputation of the school, regardless of the nature of the education offered. We will address in the following section what implications removing this right has on the workings of esteem associations. We will see that some of the outcomes of the situation are quite interesting.

13.1. Involuntary associations

The right of exit is an important one. Without the ability for members to voluntarily leave a group (exit right), or for group members to collectively expel members (expulsion right), the group loses many of the advantages which esteem associations are endowed with. What results is referred to in the

literature as an involuntary association. Negro baseball leagues were an example of one type of involuntary-association. This is the case because all good black baseball players were necessarily relegated to them, as they had no re-course to the (white) major leagues.

Thus, entry into the group was only loosely of a voluntary nature. As players were unable to leave the leagues while still retaining their professions as baseball players (no outside option), exit was restricted, if not rendered impossible. These are relatively typical features of involuntary associations. This certainly has an implication on the association and its members to glean a certain level of esteem. Among these are greater volatility in welfare and lower mean levels of esteem (a good player in the black leagues is not a good player, per se; he is a good black player, etc.).

Probably the best example of an involuntary association is the state of unemployment. Members of the "Reserve Army of Labor", if we term the club such, have absolutely no control of their membership into a class of individuals that is both despised and relegated to a subordinate role *viz.* the rest of society, a situation mostly of no doing of their own (hence the involuntary character of entry). As exit is determined by the successful match of employee and employer – a process of some arbitrariness and chance – exit is also, at the limit, involuntary. One would therefore expect to see many of the negative aspects of involuntary associations among the unemployed. Indeed, one does see much more arbitrary shifts in the welfare of the unemployed than in employed circles.

For a quick example, Munsch (2003) discusses a set of programs within the state of Saxony in Germany wherein benefits to the aged and unemployed are cut, and a percent-age (less than 100%) could be redeemed by working in – to all intents and purposes – less than desirable forms of occupation. Collective bargaining and the strictures of the labor market would likely prevent anything similar from occurring on a systematic basis to the employed.

13.2. Overlapping esteem associations

Another pertinent issue with regards to the institution of esteem associations is the case of esteem associations with overlapping purviews or boundaries. Where this is the case, certain defining qualities are shared among several independent groups. One example of this might be a national Psychiatric and Psychological trade group in any country. Both of these groups consist of members who claim to be knowledgeable about the workings of human mental faculties, and yet their expertise are disciplinarily distinct and therefore entail scruples, knowledge bases, routines and perhaps even rituals that vary between the two groupings. It will be our argument going forward that the existence of such overlaps entails both desirable and undesirable effects.

13.2.1. Desirable effects of overlapping boundaries of esteem associations

In many cases where the boundaries of esteem associations overlap, there may be actual benefits to be gleaned. For instance, a recent literature has sprung up around the Cuban baseball leagues, which in fact arose out of necessity as a result of the U.S. imposed embargo against Cuba since 1961. The embargo may disappear as a phenomenon someday, when and if Cuba-American bilateral relations normalize. It wouldn't be hard to argue that Cuban baseball has developed a distinctive flair and culture, despite the hardships the embargo has imposed on the island.

Remaining in the Caribbean, we can interpret the localized development of a distinctive musical tradition – whether it be Afro-Cuban jazz or reggae – in a similar manner. The overlaps in these areas certainly point to the fallacy of pointing to “economies of scale” or similar benefits in the case of dealing with overlapping competences, as these would be achieved at the cost of eliminating the particular flair, sound and feel of Caribbean music as seen *vier* and against standardized popular music or even Caribbean music's more popular neighbor from New Orleans.

13.2.2. Undesirable consequences of overlapping boundaries of esteem associations

There are also, of course, undesirable consequences to esteem associations whose area of competence overlap. One of these is a tendency for such overlapping competences to lead to segmentation, overspecialization and eventually even polarization among members of the respective associations. Brennan and Pettit (2004) discuss this with reference to the fact that at the limit of such a movement, disesteem from one competing association may actually equate to esteem in another group, due to the level of polarization of the constituents. An example of this can be seen in the cable news industry, where a hypostatization of advertisers, producers, commentators and ultimately viewers leads to an outcome that is obviously undesirable.

A reason such situations should be regarded as undesirable is that such internecine competition leads to an abundance of wasted effort. If we imagine the national Psychological and Psychiatric associations – going back to our earlier example – to be engaging in a similar struggle as the cable news industry, then more energy would potentially be expended on discrediting the other science than, say, in curing and treating chronic mental illness.¹³ Certainly, where possible, Afro-Cuban jazz-style overlaps should be encouraged, and cable-news-like overlaps discouraged, precisely for this reason.

¹³ It is even possible that such controversies do occur within the medical professional, for instance between the mainstream and “alternative medicine” advocates, or between the former and homeopathy, *inter alia*.

14. Publicity

Publicity facilitates the supply and demand of esteem, if we continue to view it in an economic sense. It presents the public with the behavior that needs to be evaluated, either by feelings of esteem or disesteem. As such it comprises one or more of the esteem services described above. The absence of publicity can likewise inhibit the working of esteem, leading to a situation referred broadly as pluralistic ignorance. A great example of this is a set of experiments conducted by Prentice and Miller (1993). It was determined that the tolerance for alcohol consumption was negatively influenced by the knowledge of the preferences of others. It appears that individuals change their preferences, adapting to (what they falsely believe to be) outside conditions without explicit knowledge that other students are more dismissive of alcohol consumption than (falsely) assumed. Additionally, an experiment conducted in a female restroom in a college library determined that the number of individuals maintaining the hand-washing norm more than doubled with a second person present (Munger & Harris 1989). This goes to show that the publicity, which is inextricably tied to esteem, has a significant effect on shaping behavior.

Empirical evidence points to changes in behavior (e.g., increases in cooperation in “one-shot” public goods games) in the case of non-material “punishment” in terms of “negative reputation points” and gossip in experimental settings (Barr 2001). It could be argued that publicity in such cases communicates disdain for a particular behavior, facilitates communication of those norms and thereby promoting self-corrective behavior in future interactions. Interestingly enough, all of this occurs without constraining material pay-off, giving some credence to the idea that non-material punishment is an effective tool at eliciting pro-social behavioral responses.

One can argue, furthermore, that non-material punishment or publicity punishment solves the informational problem inherent in anonymous interaction, where the determination of defectors as well as appropriate behavior sometimes remains muddled in uncertainty and risk. A good example of this might be the “positive” and “negative” feed-back system on auction sites like *eBay* (Resnick *et al.* 2006). Indeed, negative esteem is a central part of what contributes to the tenability of cooperation in many societies. In general, empirical investigations made by Fehr and Gächter (2002) reveal that cooperation is more difficult to sustain in the absence of punishment. Certainly, negative esteem is one central component of punishment. Going back to our earlier discussion of the origins of esteem in human society, one can imagine that in situations where gossip and other forms of non-material reproach are not possible (impossible to interact with the “esteem receptor”) the ability for esteem effects to elicit desirable outcomes will be somewhat

subdued. Ultimately, where such cases systematically arise, changes to institutions that contribute to these poor outcomes might be desirable.

15. Problems and Pitfalls of Esteem

15.1. Pitfalls of publicity

A clear signal that esteem has aggregative or interactive aspects is that when individual behavior deviates significantly from the norm, or when the standard deviation from the mean is very high (that is, many individuals are deviating from the prescribed behavior or compliance is inconsistent throughout the population), non-material publicity might actually be counterproductive. Brennan and Pettit (2004) discuss this predicament with reference to “the light of publicity”. In fact, the “light” may not always reveal a reality congruent with expectations, as there is no guarantee that perceived compliance is always lower than actual compliance. In fact, if the opposite is true, a tax authority, say, would have an incentive to keep true compliance a secret. On the other hand, restraint and silence might encourage the type of behavior one is attempting to thwart: “A widespread practice of averting one's gaze in the face of shameful behavior blunts that incentive” (Brennan & Pettit 2004, 171). These are all issues that must be considered in a policy context and on more practical levels as well.

Brennan and Pettit (2004) describe a situation in which the publication of false beliefs about the normal behavior of individuals may actually have desirable consequences. For instance, if the perceived norm is lower than observed behavior (perhaps because of misinformation or in-transparency), and an agency misreports observed behavior, say, in this case, exaggerating compliance with the norm in question, the action can potentially “push up” the level of the perceived norm, and therefore have a positive effect on observed behavior. Of course, this deceitful behavior raises ethical questions, and places the reputation of the publishing authority in question. These observations certainly point to a dilemma of public policy, or at least to a definite trade-off.

15.2. Status as a good with externalities

Esteem complements certain goods, like, for instance, happiness, and substitutes for certain other goods, like wage level. Indeed, a survey conducted by the author in the city of Siegen, Germany discovered a very strong correlation between various indicators of lack of social esteem – surveillance, lack of positive commendation, low level of paid vacation time – and a certain “money premium”. In other words, people who have less respect on the job need to compensate this lack of influence and respect with a certain “wage premium”. Receipt of esteem, respect, autonomy and a general sense of

contributing to social welfare (altruistic preferences outlined above) therefore appear to substitute for this “wage premium”. Frank (1985) also discusses this as a likely cause for the – relatively – flattened quality of the wage distribution in industries like car salesmen and university professors.

The price of status may be drawn from Frank (1985), who speaks of the dilemma in two dimensions. Firstly, when considering the fact that workers are typically not paid their marginal product, as mainstream economic theory would have it, Frank suggests that a possible explanation of this situation is that an individual's local position is either associated with certain costs, or must be compensated, depending on whether the individual has a high or low esteem within the organization. Individuals with low reputations within the organizations must be compensated for the negative status that they receive. They therefore earn more than their marginal product. Individuals with higher reputations are in the reverse position: their relative high status demands a premium from them: their salaries are less than their marginal product. According to this formulation, the wage distribution is “flattened out”, which is confirmed in *praxis*.

The second, related, dimension is that of an individual deciding whether to join an organization in which he or she is a “top dog, but may earn less than if he or she shifts allegiances to a competitor, or to switch and enjoy higher earnings, but lose the “top dog status. In essence, this is the choice Frank (1985) refers to as “choosing the right pond”: one can be a “big fish in a small pond”, or a “small fish in a big pond”. Each choice has benefits and each has costs. In theory, the dilemma and the respective choices can explain for the multiplicity of salaries of relatively similar workers and for the relatively “compressed” quality of salaries in the job market.

15.3. The “Adam Smith problem”

There has been a long ranging discussion within both the economics profession, and within ethics writ large, that there exists a contradiction between the ideas of human nature espoused by the moral philosopher and economist Adam Smith in the works *The Wealth of Nations* and *Theory of Moral Sentiments*. The one book, it is argued, espouses a theory of individuals as thoroughly moral and other-regarding (“sympathetic”), whereas the other seemingly advances a contradictory position of individualistic, self-regarding behavior, giving rise to the famous quotation that the butcher and the baker do not do business with each other out of mutual sympathy, but self-interest. In the next chapter, we will address why this is not, in fact, an issue. In short, there are many instances that suggest a strong contextual influence on human preferences, such that certain situations may be gauged more “sympathetically”, and others in a more utilitarian fashion, simply on the basis of contextual clues. If this is the case, as we will show it is in a wide array of

contexts, then it is likely the case that the famed “Adam Smith problem” is, in fact, not a problem at all.

16. Conclusion

In this chapter, we have attempted to outline the economy of esteem, in both its supply and demand aspects. Additionally, we discussed the role that esteem preferences and esteem goods have on human behavior, the latter exerting influence on the fulfilling of the former, and the fact that esteem is not a good in the normal sense, and can only be demanded virtually. This is evidenced by the reproach which publicity-seeking behavior produces in general, among others, via what is referred to as the hedonistic dilemma. Interaction between esteem and behavior are visible, for instance, in research on alcohol consumption of college students, among others, and certain theoretic models exist to capture the interaction between esteem goods and esteem preferences, such as the “big fish little pond, little fish big pond” conception of Frank (1985). Furthermore, we described the intangible hand, and how its employment can help achieve desirable (and undesirable) social consequences). We additionally surveyed a number of features of the esteem economy, including esteem services, publicity and esteem associations. In addition to this, we outlined the effect of punishment and status on the esteem economy. Lastly, we summarized a few key problems and pitfalls of the esteem economy, including false esteem signals and “Adam Smith problem”, which will be taken up more extensively in the following chapter.

State-Dependence

The model of state-dependent preferences offers some tools for including extraneous environmental factors and personal assessments of these into behavioral models. As such, it offers some useful tools for including certain of the features outlined above (namely, social preferences) into a standard microeconomic agent-based model. The model augments and is compatible with Bayesian rationality. In a nutshell, “state dependence” means that an individual's preferences (a utility scalar) consist of a utility function and an assessment of the likelihood of the particular outcome or payoff. The latter in turn depends on a set of prior states and the acts that elicit the outcomes, and it is this last element which represents the “state dependence” in the utility scalar: preferences which are state dependent do not display independence with regard to the underlying state of nature.

We therefore have a model with three basic parameters, states, acts and outcomes, where a state is a particular state of nature of the type “it is raining”, acts are actions taken by the individual agent of the type “watching

the soccer match in an open stadium” and out-comes are payoffs of the type “getting wet”. These we can represent respectively by the symbols s , f and x , where $s \in S$; $f \in L$ and $x \in X$. Both acts and outcomes are dependent on states, and acts are additionally contingent upon outcomes. The model proceeds with expected utility maximization, where $u(s)$ is a utility function.

The first step in showing that state dependence in preferences is a possibility consists of representing preferences as consisting of a combination of the utility function described above and the set of acts $f \in L$. A preference using this design would resemble the following:

meaning f is weakly preferred to. An individual’s preferences for f weakly dominate those for g . This result is featured famously in Friedman and Savage (1948). The important feature of this equation is that the preference relation $f \geq g$ does not depend on the utility of the respective state.

Ultimately, the preference relation for two states of nature may, in fact, differ, even if the outcomes for both are the same. To see this, we need to be able to separate beliefs from tastes, which is not directly possible in the above. We can achieve the desired effect by introducing belief updating. We introduce a parameter $p(s) = \sum f(s)$, which represents a prior probability distribution in the Bayesian sense. $p'(s)$, in turn, represents a posterior probability distribution, which is, in effect, an updated belief about the state of nature. Preference relations on $p'(s)$ are marked by \geq' . An alternate interpretation of this parameter is that $p'(s)$ represents hypothetical probabilities, such as there is a 10% chance that it may rain today”.

Preferences of the $w(s)$ sort display an affine quality. To see this, we represent the following relation: $w'(s) = p'(s)u(s)$. If both $p(s)$ and $p'(s)$ are positive for some constant s , then $w(s) = c(s)w'(s)$. This ultimately means preference relations on the state s of the type \geq and \geq' coincide (i.e., are affine).

Delimiting the choice subset to a finite set of states and outcomes, we can easily come up with a subjective probability distribution with the following conditions: a strictly positive probability distribution on S , knowledge of the decision makers choices between each and every pair of acts contingent on the hypotheticals and knowledge of the decision makers actual choices between pairs. An example of an application of this might be seen in [?], which studies state-dependence in unemployment. In this example, there are only two states: employment and unemployment, and hypotheticals are simply exponential functions on the present state. Actual statistical data on unemployment, can of course be obtained either with reference to statistical agencies (mostly stock-flow data) or by performing natural experiments.

17. Applications of State-Dependence

Applications to the state dependent model are wide-ranging. Heckman and Borjas (1980), for instance, devise a model for determining state dependence in unemployment. It is easy to imagine how this would occur. The state set is strictly finite, with two states: employment and unemployment, the probability of which depends on either the present state (Markovian dependence), or on past states (occurrence dependence), among others. We will not here derive the model employed by Heckman and Borjas (1980), but instead merely describe the methodology and the relevance of state dependence in the model. The main premise is that if it is the case that (present or past) states of unemployment influence the chance of employment, then the likelihood of having a job in period $t+i$ is, in fact, state dependent.

Keane (1997) applies state dependence to consumption behavior of households. If state dependence in consumption exists, then having purchased a good in period $t-1$ would affect the likelihood of purchasing the same good in period $t+i$. Hyslop (1999) similarly interpret labor market participation of married women. An influence of marriage and pregnancy on labor force participation was ultimately discovered. Melino and Yang (2003) discovered significant state dependence in risk aversive behavior of investors. A number of experiments using the “trust game” revealed strong dependence of fair-minded behavior in distribution on the level of trust established by the game's parameters. Additionally, evidence for “parochial altruism” was established in various games involving participants displaying an affinity for either Klee or Kandinsky (Bowles & Gintis 2004). A “religion effect” was established in “third party punishment” Dictator games, wherein members of the three large monotheistic religions displayed much less fair-minded behavior in the presence of a third player who served as a “judge” of their behavior (Barr *et al.* 2009). These results applied to modernized as well as “primitive” societies (see, for instance, Bernhard *et al.*, (2006) for examples among hunter-gatherers in Papa New Guinea).

Another issue of relevance, both with regards to altruism and its dependence on external states is neurological in nature: to what extent does the activation of certain neurotransmitters play a role in eliciting cooperative or altruistic behavior? Rilling *et al.* (2012) and Declerck *et al.* (2010) studied the issue, with interesting results. Oxytocin and Vasopressin both seemed to enhance the ability of trust relationships, upon which cooperation depends, to develop. This is an interesting result for understanding the circumstances in which cooperation may arise.

18. The Role of Incentives

In contrast with the assumption of much of neoclassical economic theory, which assumes preferences to be static and separable, a consensus is building around the idea that much more dynamic forces are at work in determining the content, development and manifestation of preferences. A seminal text in this regard is Lucas Jr (1976). In addition, the idea that preferences are truly separable is certainly increasingly cast in a doubtful light (Bowles & Polania-Reyes 2012, 38). Separability involves the exclusive economic implications of material incentives (here seen as things like taxes, subsidies and the like) on preferences; meaning that it is assumed that incentives only alter the relative costs of choices, but do not influence preferences in any other, more fundamental, way (for in-stance, signaling of one or another sort). As we will show in the following, this is an extraordinarily naive assumption.

Bowles and Polania-Reyes (2012) describe two types of preferences which incentives chosen by policymakers might influence: so-called state dependent and endogenous preferences. They analyze and summarize hundreds of trials of various principal-agent games in numerous cultures. They discuss three main avenues of incentives influencing, or “crowding”, state dependent preferences: “bad news”, “moral disengagement” and “control aversion”. The first of these is referred to as “bad news”, and refers to the fact that certain desires or motivations of the principal are made known to the agent, typically through the principal's own actions. In other words, the use of incentives by principals reveals information about the principal, such as his beliefs about the agent or his desired outcome of the interaction. The second effect, “moral disengagement” consists of the situational cues that tend to shape our behavior (often unwittingly).

The analytical difference between this mechanism and the former is that it only reflects the dynamics of the underlying social structure (e.g., in experiments, the possible decision tree), and does not convey information about the principal or other agents. As such, the effect is distinct from incentives set by the principal or other agents in that the underlying social structure reflects more appropriate behavior than conveying anything meaningful about other actors involved in the game or interaction. The third way incentives can influence preferences, broadly referred to as “control aversion” consists of the existence or nonexistence of certain feelings of autonomy or independence on the part of the agent, the nonexistence of which can have drastic – and independent – consequences on the behavior of the agent.

We can refer to the three effects in the following way: as the term “bad news” does not capture the opposite effect of “good news”, we will henceforth refer to this effect as the “informational effect”. The second effect, likewise, is vastly direction-dependent, and we can refer to it as the “ethicality effect” (Bowles & Polania-Reyes 2012, 7). The third parameter we can best refer to as

the “autonomy effect”. We now discuss the theoretical and empirical basis for these three effects briefly.

18.1. Informational parameter

The informational parameter pertains to the perception of revelation, as a consequence of the implementation of certain strategies or actions by the principal upon the agent, of beliefs of the principal towards the agent. Fehr & Rockenbach (2003) is an example of such an effect at work. German college students were designated as either “investors” or “trustees”. Investors were given the chance to transfer some amount to the trustees, who then had the amount tripled by the experimenter. The latter group could then transfer back as much or as little of the transferred amount to the investor group. One interesting feature of the experiment is that investors could specify a desired amount to be transferred back at the end of the experiment. The experimenters designed the game with a mechanism by which investors could choose to punish trustees who returned less than the specified amount.

The findings would surprise anyone holding fast to the Benthamite model of *homo economicus*: trustees granted generous transfers with higher back-transfers. The use of the fine reduces the amount of back transfers with respect to the amount transferred. Additionally noteworthy is that only one-third of investors refused the optional fine, and that their back transfers were fifty percent higher than those of investors who employed the fine.

Clearly, the fact of the principal (here, the “trustee”) implementing a tactic signaling distrust (or similar interpretations) of the agent affects the agent's response (negatively). That is not all. According to research conducted by Li *et al.* (2009), similar threats of sanctions as those described above elicited de-activation of the ventromedial PFC, whose functions we discuss in later chapters¹⁴, and which Li *et al.* (2009) conclude means a “perception shift” to more self-interested behavior (Bowles & Polania-Reyes 2012, 23), ultimately amounting to a “crowding out” of social preferences. Similar results were also discovered cross culturally in other experiments (Bowles & Polania-Reyes 2012, 22).

Of course, the effect could also be imagined to function in the other direction. And indeed, as Bowles and Polania-Reyes (2012, 28) analyze, such “crowding in” or “good news” effects were repeatedly demonstrated.

18.2. Ethicality effect

The second state dependent parameter by which incentives influence preferences is the way in which incentives change the signals determining appropriate behavior. We could also call this the framing effect, though the

¹⁴ In that the area is responsible for balancing conflicting “emotional” and “cognitive” preferences, signaling a cost involved in the decision.

influence we wish to describe is not merely restricted to decision-making, but includes their influence on ethical beliefs, dispositions and behaviors, as well. A number of studies have been conducted which support the argument that decisions made by experimenters in the organization (or even the naming) of games has a significant impact on behavior elicited subsequently. The argument given ultimately runs along the lines that different “framing” setups influence the nature of the decision dynamic, and thereby motivate the use of an ethical “on/off switch”. As established by Schotter *et al.* (1996), market-like competition among agents reduces agents’ assessment of the importance of fairness. The “religion effect” is of a similar nature. Shifting from a Dictator to a “Third party punishment” game reduces the difference in endowment allocated between members of the three monotheistic world religions and others by nearly 70%. This is a remarkable result. Presumably, it can be explained by the shift in the perceived importance of ethical norms when a third party is present to adjudicate the situation and react accordingly.

Again, the argument works in the opposite way, as outlined in the previous chapter. Obligatory fines may be an example of ethical “crowding in” (Cardenas 2004). If these coalesce with pre-existing social norms, the fines may reinforce certain (positive) social behaviors.

18.3. Autonomy parameter

The last effect by which state dependent preferences are influenced is the so-called “autonomy parameter”. Although this parameter intends to measure the extent to which individual responds to certain situations of control and independence, it is clear that its effects will often be correlated (or even coterminous with) the other two. As such, it is analytically challenging, but its empirical relevance is clear. Therefore, we will briefly outline its workings as discussed by Bowles and Polania-Reyes (2012).

When given a choice to employ a demand for a certain endowment on promise of punishment in an experimental setting, Falk and Kosfeld (2006) discovered that those principals who used the demand received 50% less than those who did not make use of the demand. A similar case in a non-experimental setting occurred with regards to sick day leave in the Boston Fire Dept:

On December 1, 2001 the Boston Fire Department terminated its policy of unlimited paid sick days, replacing it with a 15-day sick day limit; pay would be docked for firemen exceeding the limit. The firemen responded to the new incentives: those calling in sick on Christmas and New Year's Day increased tenfold over the previous year (Bowles & Polania-Reyes 2012, 2).

It is clear that incentives affecting autonomy, like the other state dependent preference parameters above, may exhibit a positive “crowding in” effect, as well¹⁵.

19. Conclusion and Synthesis

A significant literature exists revealing a use and relevance for state-dependence in preferences. There is both theoretical and empirical evidence for the validity of state dependence in a wide array of environments. Incentives do not work separately on preferences, meaning that one cannot “cleanly” separate preferences into economic and moral elements. Therefore, incentives have the characteristic of being able to “crowd in” or “crowd out” either or both of these elements. This occurs by one or more of a number of vehicles, including providing information about others’ beliefs and intentions, “framing” appropriate behavior and by restricting (or enabling) a sense of individual autonomy on the part of the individual agent. Examples include the change of policy in the Boston fire department, where social preferences were likely “crowded out” by a shift in the frame and a loss of autonomy, along with a number of other examples cited.

In closing the paper, we make the traditional Hegelian spring from thesis, antithesis, and at last to synthesis, and attempt to close full circle on the issue of how social preferences and state-dependent preferences can be brought into further focus in economics research. The chapter will discuss, in the order mentioned, the issue of co-evolution as a means of maintaining certain institutional status quos; an assessment of the mainstream views on social preferences; the importance of making use of the intangible hand in policymaking; and the basis of human nature as one not, indeed, captured by Benthamite self-regard, but of a social, rounded human agency that considers the wider social context for cues and has an extent of sympathy for others. The chapter will close by mentioning some computational tools for integrating

¹⁵ One could also speak of endogenous preferences when addressing the present discussion. However, we will delimit our discussion on that issue and provide references to the relevant literature at present. Bowles and Polania-Reyes (2012, 9) describe the difference between strict state dependent and endogenous preferences in the following manner:

An example unrelated to incentives may clarify the difference between endogenous and state-dependent preferences. As Italian residents, your authors now eat a lot more pasta than we did in our countries of origin. Abstracting from possible international price differences, this could be another case of “when in Rome, do as the Romans”. Or it might be that we have newly come to enjoy the taste of pasta, perhaps through extensive exposure to it while in Italy. Which case it is – state-dependent or endogenous preferences – would be revealed by what we will eat back in Bogotá or Santa Fe. If we go back to arepas or potatoes, then our taste for pasta was state-dependent. If we remain pastaphiles, then our preferences have endogenously changed.

some of the discussed issues into research. As such, we hope to weave the thread of the above essay to a close by presenting clearly the gaps in present practice as discussed above, in addition to arguing for the merits of changing present practice and ultimately providing at least one concrete methodological tool to potentially execute this change, i.e., agent based computational modeling.

20. Co-Evolution

Institutional co-evolution refers to the simultaneous evolution of certain traits in societies via the dual route of genetics and culture. Given a certain probable base level of altruism (via mutation or migration, for instance), the process of multi-level selection can predispose certain behaviors to advance evolutionarily via the mechanisms outlined in chapter 4. Certainly, culture, imitation and convention play a role, and Brennan and Pettit (2004) discuss interesting ways for capturing the process by which such institutional co-evolution occurs from an economic standpoint. We will label the two views the “vanguard model” and the “whistle-blower model” (my own terminology). These ultimately are two vehicles facilitating the spread and maintenance of certain ideas, institutions, beliefs and practices.

20.1. Vanguard model

Assume that compliance with a convention is a binary matter (for simplicity's sake). Second, assume a heuristic approach to convention compliance in the population generally, the major levers for which are ideal and standard levels of compliance. Then, it is easy to imagine the spread of a certain convention as a monotone increasing function of the number of adherents to the convention. This would, at first glance, seem to be a truism. However, the situation demands a more thorough representation than a mere assertion. As such, we take the approach employed by Brennan and Pettit (2004) to suggest a function with three distinct co-domains: a first section in which no convention exists, a second where a small “vanguard” carries out the behavior prescribed by the (new) convention, and a third wherein virtually all citizens, or a significant number thereof, abide by the convention.

This could be understood, as Brennan and Pettit (2004) take it to be, to reflect the various incentives that arise with reference to esteem, as discussed in chapter 9. Indeed, to jump ahead, one can imagine an incentive to act in a certain manner as being absent in the first co-domain, when few or none engage in the practice. Brennan and Pettit (2004) use the example of recycling. As more individuals both espouse the benefits of and engage actively in the practice, the level of positive approbation attached in the activity increases. However, as no overwhelming social norm has been established at this point, no negative attitude is associated with not adhering to the convention. In the

final co-domain of the compliance function, the practice is so widespread that, while at some points there may be a certain (declining) level of positive esteem tied to performing the action, increasingly it is negative beliefs and perceptions about those failing to comply with the (now) norm that drive compliance in the final stages of the adoption of the norm.

It is easy to see that this model captures one organic route by which new institutions develop and proliferate throughout societies. It is possible that such institutions have the ability to evolve without the intrusion of state power, but it is imaginable that, at the limit, a certain amount of “nudging” might be desirable. The heavy state involvement in the spread of renewable energy – via feed-in tariffs and the like – is one example. In this case, state power can be employed to help the future “vanguard” overcome the strictures of institutional paralysis and inertia.

20.2. Whistle-blower model

The whistle-blower dilemma consists less of a concept that directly aids in institutional development and serves more in terms of guiding institutional maintenance. Whistle-blowing clearly has benefits for the maintenance of institutions at risk of decline or destruction via opaqueness or corruption. These benefits accrue via the aid that whistle-blowing in general offers institutions like public opinion in reacting correctly to undesirable behavior that might otherwise remain secret.

The whistle-blowing dilemma assumes a certain degree of rational preference-updating behavior, different from the vanguard model, which is more concerned with aggregate behavior. In essence, an individual who becomes privy to corrupt or intolerant behavior within an institution must engage in a continual process of both assessing his level of knowledge of the behavior in relation to the standard by which it is measured, and of putting this in relation with the context of silence among any colleagues or peers. The whistle-blower dilemma serves as a check on coarsely reporting information that may more efficiently be dealt with internally. As Brennan and Pettit (2004) state:

In lots of cases, I will not be the only person in a position to have information about a suspected case. In deciding whether to blow the whistle in such a case, I rationally focus my attention on the circumstances in which my action will make a difference. And the only case in which this is so is when no one else blows the whistle. For if someone else blows the whistle I don't need to. This fact gives me pause. Because there is, after all, some uncertainty about whether whistle-blowing is actually justified – uncertainty about the facts of the case, about the consequences of going public, and so on. And I realize that each other person is making an independent assessment of these considerations and that, in the only case that matters (i.e., where my action is

decisive), none of those others has decided that the circumstances justify proceeding.

Ultimately, whistle-blowing in principle aids the maintenance of social norms by compensating for a failure of one or more of the other institutions discussed above, which can fail for a number of reasons, including a process of “reverse Darwinism”, where nepotism facilitates a process of cronyism and corruption.

Whistle-blowing is linked to the issue of social esteem, because whistle-blowers tend to suffer loss of esteem by coming forward, and whistle-blowing can destroy the esteem of the implicated organizations. The very real risks of extraordinarily adverse implications for both individuals and institutions naturally mean that institutions have a natural incentive to mitigate whistle-blowing from occurring that may go above and beyond the necessary.¹⁶ This situation can have potentially detrimental effects on the ability of whistle-blowing to function properly. This situation is basically one of moral hazard, where two parties in a contract have split incentives. It provides a strong argument for the need of neutral, nonpartisan institutions that serve to protect and facilitate whistle-blowers and the whistle-blowing process. Wide-ranging immunity and political asylum – as is present with respect to cartel law – are examples of such necessary protections. They, in fact, serve to insure the proper working of whistle-blowing as a check on opaqueness and corruption of fixed social institutions.

21. An Assessment of the Mainstream View

A number of field experiments have shown that a failure to adequately reflect on certain social preferences (see chapter 4 above) leads to poor policy, or, at least, policy which is not fully “thought out”. The Boston fire department example shows how this might be the case. Incentives and institutional mechanisms which obfuscate social preferences, for instance, by clouding agent autonomy or by revealing antipathetic beliefs or intentions on the part of principals or other agents, can lead to “crowding out” of moral as well as of economic motivations (to be seen largely as non-separable). This in turn can lead to a distortion in the effectiveness of policies, what Bowles and Polania-Reyes (2012) refer to as “naïve” policy. Naive policy consists of measures that fail to adequately reflect on their own impact on beliefs and preferences, thereby leading to a “crowding out” of some or all of the intended effects. The cases of the trust and gift games discussed above should serve to concretize

¹⁶ One can see this, for instance, in the Obama administration's extremely vigilant attitude toward discovering and prosecuting whistle-blowers internal to the administration and government.

this. Most of these shortcomings are discussed in further detail in the above work, and are only summarized here to point out the main conclusions drawn in the preceding sections.

21.1. Preference separability

Important problems for social policy with regards to social preferences in economics research is the often occurring (and, as above described, incorrect) assumption of preference separability. Under this rubric, “moral preferences can be neatly distinguished from “economic ones: we read *The Wealth of Nations* at the workplace, and the *Theory of Moral Sentiments* in bed at night. Contradictions to this belief are drawn up in Bowles and Polania-Reyes (2012). As pointed out in chapter 16, there are a number of fallacies with the assumption of preference separability, including the outlined autonomy, informational and ethicality (and other) parameters of state dependent preferences, and the lack of any substantive evidence for the phenomenon. The fact of interdependence between “ethical” and “economic” preference should therefore play a more prominent role in economic (and more general social) modeling.

21.2. Policy trade-offs

One manifestation of preference interdependence is encountered in the public goods example of dealing with free-riding. Policymakers can attempt to discourage this behavior in a number of ways, including by subsidizing pro-social behavior. However, doing so might create a situation where the perception of motives by people behaving in a pro-social manner is of having been done for pecuniary reward (so-called “he’s doing it for the money” perception). As such, esteem and altruistic incentives may be “crowded out” by policy frameworks focusing only on the free-riding problem to the exclusion of preference interdependence.

To continue with the present example, it may in fact be determined to be beneficial to have a subsidy, in the end, either because it is at such a low level, or the “crowding out” effect so small in comparison with the benefits accrued from the subsidy that the latter effects outweigh any undesirable outcomes. It is worth pointing out that “naïve” policy is not necessarily coterminous with “excessive” policy: in many cases, a libertarian point of view can create just as undesirable a situation with respect to social policy as a situation where incentives are excessively employed. Financial regulation is one clear example of this. “Naïve”, ultimately, entails simply any situation where policymakers fail to reflect adequately on the interplay between preferences and incentives. Discussion of this issue can be found in Cardenas (2004), Al-Ubaydli *et al.* (2013), Vertova and Galbiati (2010).

22. Making Use of the Intangible Hand

Having made the case for the existence of social preferences, one can ask the famous question “whither now?” or “what is to be done?”. How does one incorporate the findings and revelations discussed and described in the above into economic theory and research in general? It will be our argument in this section that a great tool for policymakers to employ when attempting this task is the intangible hand, discussed and outlined above. We will discuss a number of cases that briefly illustrate how policymakers can potentially elicit desirable responses with reference to the intangible hand.

In one case (Lewin 2015), a recent report suggests that thousands of Chinese women fly the U.S. every year in order to give birth to American citizens. These women often pay upwards of \$60,000 in order to receive this privilege. There are obvious moral reservations as to this type of behavior, not least of which is the obvious disparities this creates between Chinese with and without the means to solicit such services. States can take a wide range of actions to curb the behavior. One example might be publicizing the views of the broad public on women who pay to give birth to their children in the U.S. The principle of pluralistic ignorance suggests that the real views of citizens on this issue may be different from perceived norms. Having the real views of the citizenry published may in fact reveal the stigma attached to the activity, and hasten to curtail the motivation to do it in the first place. Additionally, the principle of parochial altruism tells us that this policy may be more effective if enacted in China, rather than in the U.S., since members of many societies value the beliefs, judgments and approbation of their fellow citizens more than those of outsiders.

With regards to recent revelations of illegal activities by a number of large international banks, it may be the case that more stringent action is ultimately necessary to curb the undesirable activity than merely publishing a list of “bad banks”. However, the intangible hand can prepare the waters for the wider berth of the iron hand of state power. Indeed, the issue of publicity discussed above may come to the fore here. As discussed in chapter 9, publicity has the potential to push perceived standards either upward or down, and revelations of clearly systematic extortion, bribery, manipulation and money laundering would, according to common sense, push them downwards. As discussed above, this situation could in some instances be undesirable. However, it may push them down far enough that a tipping point is reached, whereby a consensus forms around the idea that the present system and its rules are corrupt or in need of reform.¹⁷ In this regard, the intangible hand can serve both to extend the reach of the invisible hand, as well as to determine where and when the iron hand of state power is necessary.

¹⁷ See, for instance, Arnold (2015).

Lastly, it is important to remember, when considering policy alternatives, that there may be a fallacy in simply taking preferences as given. As Brennan and Pettit (2004, 251) state: "Mainstream economics does not have to judge on the merits of the preferences or the individuals in play; it simply works with the assumption that whatever people's preferences in the sphere of the material economy, it is better to have more preference satisfaction than less. In that sense economics operates with a thin theory of the good: it does not introduce any exogenous or thick criteria of assessment." Indeed, a relevant question for policymakers is how to use the intangible hand in order to achieve socially desirable outcomes in situations where the absence of external influence might lead to a less desirable outcome. Examples of this might be smoking or teenage pregnancy. As pointed out in our introductory chapter, more preference satisfaction is not necessarily a good thing!

23. Homo Socialis?

Although a growing segment of economics research incorporates elements of other-regarding preferences in models and theories¹⁸, the continuity of this trend with the rest of the discipline (in particular with introductory coursework) is lacking. The dominant model in economics, as taught in textbooks and recited in lecture halls around the world in various undergraduate and graduate level economics programs, is that of *homo economicus*, a self-interested, utility-maximizing individual agent. It is likely that this model fits well with the dictates of neoclassical theory, and this explanation may serve to account for the lack of alternative behavioral models of human behavior. However, as we have shown in the above, this model has deep problems – some obvious, some less so. Most principally, human agents in many cases display systematic deviations from the utility maximizing dictate of the *homo economicus*. Including this fact in models of human agency would serve not only the strength of economic models in terms of predictive and explanatory capacity, but would also lend some level of continuity between models and reality, an element that certain economists argue is lacking in the discipline today¹⁹, sacrificed for the sake of parsimony and rigor. In this regard, Herbert Gintis speaks of a "physics envy" among the economics profession.

The lack of regard for esteem and altruistic preferences in research, despite the wide array of evidence for them, has the potential to discredit

¹⁸ For a prime example, consider the wide array of research into so-called behavioral economics.

¹⁹ Additionally, it could be argued that the falsifiability of a science depends to a great deal on the ability for theories within that science to be tested and verified. It is hard to test a theory that has virtually no bearing on reality.

social science research in general by diminishing its legitimacy (that is, its legitimacy as a genuine scientific undertaking). It is likely that a main reason for the lack of esteem preferences in economic modeling is the complexity of the emotions and states involved. Nevertheless, as we will argue in the following section, the state of modern computing is such that even extremely intricate and subtle details of behavior and preferences can be relatively accurately represented in certain modeling systems. It would arguably behoove researchers to familiarize them-selves with the methods detailed below to more fully come to grasps with the full range of human behavior and action.²⁰

And there clearly is a range of human behavior. The above text has attempted to point out the failure of models that imply a thoroughly rational, self-regarding agent with exogenous preferences to adequately reflect the breadth and scope of human behavior, to their own detriment. New models are needed to augment (and potentially replace) *homo economicus* as benchmark behavioral schemes. One such model is *homo socialis*, which assumes a varying degree of social preferences in individuals, that may and can have “feed-back effects” on economic preferences. As outlined in the above chapters, this model much more accurately captures real dynamics of human behavior and has the ability to capture the more full range of human behavior – even as it relates to economic realities. As such, the author will make the rather strong assertion that, save for specific contexts where the significance is spelled out clearly, the *homo economicus* model should be updated to the standards of *homo socialis*.

24. Agent-Based Behavioral Modeling

Agent-based behavioral computational economic (ACE) modeling offers an expansive palate of tools that researchers can apply to solving a lot of the systematic problems outlined above in engaging in economic analysis. Agent-based modeling goes back at least to the middle of the last century, to the von Neumann machine. This and other early examples of agent-based systems allowed relatively simple systems with rudimentary parameters to replicate and coordinate autonomously, thereby allowing for a certain level of abstraction or simulation. The science has certainly advanced leaps and bounds since the 1940s, and much of what is possible today with agent-based modeling can include many subtleties and nuances of human (and environmental) behavior and predispositions. Tesfatsion (2003) describes agent-based behavioral modeling as “encouraging economists to address

²⁰ Certainly, it can be argued that for the purpose of “back of the envelope” calculations, an optimization here and a Lagrangian there might do the job required, but where more detailed results are necessary, one should not shy away from more sophisticated methods merely because they weren't featured in the textbook.

growth, distribution, and welfare issues in a more comprehensive manner embracing a variety of economic, social, political, and psychological factors, thus restoring the broad vision of early political economists.”

Indeed, it is arguably intellectually more challenging to derive a model that is not only predictively but also conceptually accurate. Opposing Friedman (1953), that the only measures of the value of an economic model should be simplicity and fruitfulness, this point of view would advance as a reservation that contemporary computational power is sufficient in order to entertain more complex facets of economic and social life within the scope of models, while not losing any of the predictive power or “fruitfulness” of the model. In fact, such a view would continue, holding on to various elements of neoclassical analysis derived in the 19th century to the detriment of newer and more efficient methods of analysis serves to hold the discipline of economics in the terrain of non-evolutionary sciences, as Thorsten Veblen argued 100 years ago.

What role computation and programming play in this is not certain. What is certain is that economies are social systems that are also prone to the laws of evolution delineated briefly above, and therefore change in terms of their parameters and configurations and ordering over the course of time and history. Assessing, describing and interpreting such a dynamic system arguably requires more complex tools than neoclassical analysis provides, and simplicity and elegance are in this case not necessarily desirable, as social systems are messy. Computation allows a researcher to capture some of that messiness without losing any of the explanatory power of good models. As Tesfatsion (2006) states:

For economic purposes, computer programming is in some ways more powerful [than differential calculus] in that it facilitates the modeling of complex interactive processes involving kinks, jumps, and other forms of discreteness imposed or induced by empirical constraints. Consequently, programming frees us to adapt the tool to the problem rather than the problem to the tool²¹.

25. Concluding Remarks

It appears then that a number of fundamental features of the *homo economicus* model, namely that of expected utility maximization, Benthamite self-regard and methodological individualism, are premised on assumptions that at best hold up only tentatively, either in severely constrained environments, or with the addition of highly implausible preconditions. This means that, while the model may apply in certain settings, and particularly those where a large amount of cost-benefit analysis devoid of moral constraints is encountered, its

²¹ Tesfatsion (2006) ends this passage by calling for “every graduate economics program” to include programming language requirements.

applicability is by no means universal. Alternate models that reflect other aspects of human agency were suggested in the above, and particular emphasis was given to models that include social preferences and preferences that are context dependent (so-called state-dependent preferences). Ultimately, there are critical aspects that have been glossed over in the present work. This should by no means discount the salience or significance of these aspects. The present work, however, being limited in scope (and length), was written with certain emphases that resulted from the given constraints of the project at hand.

References

- Al-Ubaydli, O., Houser, D., Nye, J. V., Paganelli, M. P., & Pan, X. 2013. "The Causal Effect of Market Participation on Trust: An Experimental Investigation Using Randomized Control." *PLoS One* 8(3): 1–8. doi: 10.1371/journal.pone.0055968.
- Allais, M. 1953. "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine." *Econometrica: Journal of the Econometric Society* 21(4): 503–546.
- Ambrose, S. H. 1998. "Late Pleistocene Human Population Bottlenecks, Volcanic Winter, and Differentiation of Modern Humans." *Journal of Human Evolution* 34(6): 623–651.
- Andersen, K. K., Azuma, N., Barnola, J.-M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H., *et al.* 2004. "High-Resolution Record of Northern Hemisphere Climate Extending Into the Last Interglacial Period." *Nature* 431(7005): 147–151.
- Andreoni, J. and Miller, J. 2002. "Giving According to Garp: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica: Journal of the Econometric Society* 70(2): 737–753.
- Arnold, M. 2015. "Public Outrage Is the Perfect Stick to Prod HSBC Into Break-Up." *The Financial Times* March 2.
- Axelrod, R. & Hamilton, W. D. 1981. "The Evolution of Co-Operation". *Science* 211(4489): 1390–1396.
- Barr, A. 2001. *Social Dilemmas and Shame-Based Sanctions: Experimental Results from Rural Zimbabwe*. Oxford: Centre for the Study of African Economies.
- Barr, A., Ensminger, J., Henrich, J., Wallace, C., Barrette, C., Bolyanatz, A., Cárdenas, J.-C., Gurven, M., Gwako, E., Lesorogol, C., *et al.* 2009. *Homo Aequalis: A Cross-Society Experimental Analysis of Three Bargaining Games*. Oxford: Department of Economics.
- Bernhard, H., Fischbacher, U., & Fehr, E. 2006. "Parochial Altruism in Humans." *Nature* 442(7105): 912–915.

- Bowles, S. & Gintis, H. 2004. "Persistent Parochialism: Trust and Exclusion in Ethnic Networks." *Journal of Economic Behavior & Organization* 55(1): 1–23.
- Bowles, S. & Gintis, H. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, NJ: Princeton University Press.
- Bowles, S. & Polania-Reyes, S. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?." *Journal of Economic Literature* 50(2): 368–425.
- Brennan, G. & Pettit, P. 2004. *The Economy of Esteem: An Essay on Civil and Political Society*. Oxford: Oxford University Press.
- Camerer, C. & Thaler, R. H. 1995. "Anomalies: Ultimatums, Dictators and Manners." *The Journal of Economic Perspectives* 9(2): 209–219.
- Cardenas, J.-C. 2004. "Norms from Outside and from Inside: An Experimental Analysis on the Governance of Local Ecosystems." *Forest Policy and Economics* 6(3): 229–241.
- Cowen, T. 2005. Review Essay: The Economy of Esteem". *Politics, Philosophy & Economics* 4(3): 374–382. doi: 10.1177/1470594X05056609.
- Darwin, C. 2004. *The Descent of Man*. London: Penguin Publishing.
- De Waal, F. B. 1996. *Good Natured*. Cambridge, MA: Harvard University Press.
- Declerck, C. H., Boone, C., & Kiyonari, T. 2010. "Oxytocin and Co-Operation Under Conditions of Uncertainty: The Modulating Role of Incentives and social information." *Hormones and Behavior* 57(3): 368–374.
- Ellingsen, T. & Johannesson, M. 2007. "Paying Respect." *The Journal of Economic Perspectives* 21(4): 135–150.
- Falk, A & Kosfeld, M. 2006. "The Hidden Costs of Control." *The American Economic Review* 96(5): 1611–1630.
- Fehr, E. & Camerer, C. F. 2007. "Social Neuroeconomics: The Neural Circuitry of Social Preferences." *Trends in Cognitive Sciences* 11(10): 419–427.
- Fehr, E. & Gächter, S. 2002. "Altruistic Punishment in Humans." *Nature* 415(6868): 137–140.
- Fehr, E., Kirchsteiger, G., & Riedl, A. 1998. "Gift Exchange and Reciprocity in Competitive Experimental Markets." *European Economic Review* 42(1): 1–34.
- Fehr, E. & Leibbrandt, A. 2011. "A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons." *Journal of Public Economics* 95(9): 1144–1155.
- Fehr, E. & Rockenbach, B. 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature* 422(6928): 137–140.
- Frank, R. H. 1985. *Choosing the Right Pond: Human Behavior and the Quest for Status*. Oxford: Oxford University Press.
- Friedman, M. 1953. *Essays in Positive Economics*. Vol. 231. Chicago, IL: University of Chicago Press.

- Friedman, M. & Savage, L. J. 1948. "The Utility Analysis of Choices Involving Risk." *The Journal of Political Economy* 56(4): 279–304.
- Gintis, H. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, NJ: Princeton University Press.
- Heckman, J. J. & Borjas, G. J. 1980. "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* 47(187): 247–283.
- Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J., Kidd, J. M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., *et al.* 2011. "Hunter-Gatherer Genomic Diversity Suggests a Southern African Origin for Modern Humans." *Proceedings of the National Academy of Sciences* 108(13): 5154–5162.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Hyslop, D. R. 1999. "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women." *Econometrica: Journal of the Econometric Society* 67(6): 1255– 1294.
- Kahneman, D. & Tversky, A. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica: Journal of the Econometric Society* 47(2): 263– 291.
- Keane, M. P. 1997. "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business & Economic Statistics* 15(3):310–327.
- Kropotkin, P. 2012. *Mutual Aid: A Factor of Evolution*. Mineola, NY: Courier Dover Publications.
- Kurzban, R. & Houser, D. 2005. "Experiments Investigating Cooperative Types in Humans: A Complement to Evolutionary Theory and Simulations." *Proceedings of the National Academy of Sciences of the United States of America* 102(5): 1803–1807.
- Kurzban, R. & Leary, M. R. 2001. "Evolutionary Origins of Stigmatization: The Functions of Social Exclusion." *Psychological bulletin* 127(2): 187.
- Lewin, T. 2015. "California Apartments Raided in Federal Investigation of Chinese 'Birth Tourism'." *New York Times* March 3.
- Li, J., Xiao, E., Houser, D., & Montague, P. R. 2009. "Neural Responses to Sanction Threats in Two-Party Economic Exchange." *Proceedings of the National Academy of Sciences* 106(39):16835–16840.
- Lins, T. C., Vieira, R. G., Abreu, B. S., Gentil, P., Moreno-Lima, R., Oliveira, R. J., & Pereira, R. W. 2010. "Genetic Heterogeneity of Self-Reported Ancestry Groups in an Admixed Brazilian Population." *Journal of Epidemiology/Japan Epidemiological Association* 21(4): 240–245.

- Lucas Jr., R. E. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46.
- Machina, M. J. 1987. "Choice Under Uncertainty: Problems Solved and Unsolved." *Journal of Economic Perspectives* 1(1): 121–154.
- Melino, A. & Yang, A. X. 2003. "State-Dependent Preferences Can Explain the Equity Premium Puzzle." *Review of Economic Dynamics* 6(4): 806–830.
- Moreno-Gámez, S., Bowles, S., & Wilkins, J. 2009. *Cosmopolitan Ancestors: Simulations Calibrated with Genetic and Ethnographic Data Show That Prehistoric Populations Were Not Small and Isolated*. Santa Fe, NM: Santa Fe Institute.
- Morgenstern, O. & Von Neumann, J. 1953. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Munger, K. & Harris, S. J. 1989. "Effects of an Observer on Hand Washing in a Public Restroom." *Perceptual and Motor Skills* 69(3): 733–734.
- Munsch, C. (Hrsg.) 2003. *Sozial Benachteiligte engagieren sich doch: über lokales Engagement und soziale Ausgrenzung und die Schwierigkeiten der Gemeinwesenarbeit*. Weinheim - München: Beltz Juventa.
- Oosterbeek, H., Sloof, R., & Van De Kuilen, G. 2004. "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis." *Experimental Economics* 7(2): 171–188.
- Prentice, D. A. & Miller, D. T. 1993. "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm." *Journal of personality and social psychology* 64(2):243.
- Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. 2006. "The Value of Reputation on Ebay: A Controlled Experiment." *Experimental Economics* 9(2): 79–101.
- Rilling, J. K., DeMarco, A. C., Hackett, P. D., Thompson, R., Ditzen, B., Patel, R., & Pagnoni, G. 2012. "Effects of Intranasal Oxytocin and Vasopressin on Cooperative Behavior and Associated Brain Activity in Men." *Psychoneuroendocrinology* 37(4): 447–461.
- Salzano, F., Neel, J., Gershowitz, H., & Migliazza, E. 1977. "Intra and Intertribal Genetic Variation Within a Linguistic Group: The Ge-Speaking Indians of Brazil." *American Journal of Physical Anthropology* 47(2): 337–347.
- Schotter, A., Weiss, A., & Zapater, I. 1996. "Fairness and Survival in Ultimatum and Dictatorship Games." *Journal of Economic Behavior & Organization* 31(1): 37– 56.
- Sugden, R. 1982. "On the Economics of Philanthropy." *The Economic Journal* 92(366): 341–350.
- Tesfatsion, L. 2003. "Agent-Based Computational Economics: Modeling Economies As Complex Adaptive Systems." *Information Sciences* 149(4): 262–268.

- Tesfatsion, L. 2006. "Agent-Based Computational Economics: A Constructive Approach to Economic Theory." *Handbook of Computational Economics* 2:831–880.
- Vertova, P. & Galbiati, R. 2010. "How Laws Affect Behavior: Obligations, Incentives and Cooperative Behavior." *Incentives and Cooperative Behavior*: 1–36.
- Wilson, D. S. & Sober, E. 1994. "Reintroducing Group Selection to the Human Behavioral Sciences." *Behavioral and Brain Sciences* 17(04): 585–608.

Jerome Nikolai Warren (Siegen, Germany)

***A Wedding Toast Is Not a State of the Union Address:
Altruism and Esteem Preferences as Relevant Factors
in Shaping Human Agency***

Abstract: This paper describes using current literature and research a problem that has plagued social scientists for centuries, that of "moral sentiments". Human beings are inherently social by nature and hold certain regard for others' opinions (esteem preferences) as well as for others generally (altruism). It is argued in the article that such preferences may in fact be consistent with a core rational human agent. It is furthermore argued that the lack of regard for such preferences in social sciences research (and particularly within the domain of economics) severely weakens models and theories in the respective disciplines. A few potential avenues for including social preferences writ large into social science (read: economic) modeling are outlined.

Keywords: esteem, altruism, social preferences, *homo economicus*

Ethics in Progress (ISSN 2084-9257). Vol. 6 (2015). No. 2. 104-152.

doi: 10.14746/eip.2015.2.7