



**University of  
Zurich** UZH

University of Zurich  
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 26

**Tastes, Castes, and Culture:  
The Influence of Society on Preferences**

Ernst Fehr and Karla Hoff

August 2011

---

# **Tastes, Castes, and Culture: The Influence of Society on Preferences**

Ernst Fehr<sup>a)</sup> and Karla Hoff<sup>b)</sup>

August 8 2011

*Abstract:* Economists have traditionally treated preferences as exogenously given. Preferences are assumed to be influenced by neither beliefs nor the constraints people face. As a consequence, changes in behaviour are explained exclusively in terms of changes in the set of feasible alternatives. Here we argue that the opposition to explaining behavioural changes in terms of preference changes is ill-founded, that the psychological properties of preferences render them susceptible to direct social influences, and that the impact of “society” on preferences is likely to have important economic and social consequences.

Running short title: **The Influence of Society on Preferences**

Key words: endogenous preferences, culture, caste, frames, anchors, elicitation devices

JEL classification: A12, A13, D01, K0

---

\*<sup>1</sup>) This paper is part of the Research Priority Program on the “Foundations of Human Social Behavior—Altruism vs. Egoism” at the University of Zurich.

<sup>a)</sup> Ernst Fehr, Department of Economics, Laboratory for Social and Neural Systems Research, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland, email: [ernst.fehr@econ.uzh.ch](mailto:ernst.fehr@econ.uzh.ch)

<sup>b)</sup> Karla Hoff, World Bank, email: [khoff@worldbank.org](mailto:khoff@worldbank.org)

... one does not argue about tastes for the same reason that one does not argue over the Rocky Mountains – both are there, will be there next year, too, and are the same to all men.“ (Stigler and Becker 1977, p. 76)

“Almost all sociologists take it as obvious that individuals’ preferences are formed by society and that society, so to speak, exists within persons.“ (Peter Hedstrom, Oxford University, personal communication)

There is a standard view in economics that to explain outcomes, one can begin by characterizing the preferences of individuals and then go on to describe the society that the individuals will create. The preferences remain the same regardless of the society that emerges. This view contrasts sharply with the conception of the individual in other social sciences, in which institutions and other socially defined variables influence preferences. Casual observation suggests that there are vast behavioural differences across societies and even across groups within societies, and cross-cultural research has documented striking differences in preferences across societies (*e.g.* Henrich *et al.* 2005, Henrich *et al.* 2010). However, despite the documentation of large cultural differences, conclusive scientific evidence that preferences are causally shaped by social institutions and cultural practices is very rare.

Take, for example, the study of Henrich *et al.* (2010), which shows that the preferences of individuals to be fair in anonymous transactions and to punish unfairness increase with the level of the society’s market integration, measured as households’ average percentage of calories that are purchased. It is a plausible hypothesis that market integration shapes preferences for fairness because in market exchanges the *quid pro quo* is a salient characteristic of interactions that may support the development of such preferences. However, it is also possible that stronger fairness preferences, *e.g.* preferences to honour informal contractual obligations, foster market integration.<sup>1</sup> Thus, one cannot unambiguously conclude from the observed correlations that the cultural and economic institutions prevailing in the different societies have causally shaped the preferences for fairness and punishment.

---

<sup>1</sup> The positive correlation between altruistic punishment and community size can also be interpreted in two ways. It is a plausible hypothesis that larger communities develop more elaborate forms of institutional punishments that, in turn, increase the willingness of individuals to punish altruistically. However, it is also plausible that a higher willingness to engage in altruistic punishment enables small-scale societies to enforce the social norms that support the development of larger, more productive communities.

There are several reasons why the study of the causal influence of institutional and cultural factors on preferences is difficult. One obvious reason is that societies and groups typically differ in a large number of dimensions but the researcher cannot control for all of them. Another reason is that due to the existence of multiple equilibria, societies may end up with different institutions, cultural practices, and preferences despite the fact that they started with similar factor endowments and preferences. In this case, the co-variation between institutions, cultural practices, and preferences need not indicate a causal impact of institutions on preferences. The co-variation may merely reflect the fact that other background variables (which may be idiosyncratic historical events such as a war or a natural disaster) led to different developmental trajectories.

Any investigation of causal influences of institutional and cultural factors on preferences requires settings and techniques in which these potential confounds are addressed. The papers in this symposium use a caste status divide in India, a language border within Switzerland, and a manipulation of the decision format in a blood drive to plausibly identify causal influences on preferences. In this introduction, we put these papers in the perspective of the emerging literature on endogenous preferences.<sup>2</sup>

This paper is divided into three parts. We first address the well-known objection that explaining behavioral changes by preference changes introduces *too many free variables*. We argue that today such explanations should be considered on an equal footing with constraint-based explanations. Recent progress in game theory has provided many free, unobservable, variables that can explain outcomes, whereas recent progress in modeling and measuring preferences has put more constraints on preference-based explanations.

In the second part, we discuss the progress in psychology in pinpointing the determinants of preferences. Psychological research has shown that preferences can be affected by the way they are elicited, by the framing of situations, by anchoring devices, and by the priming of individuals' identities. *Thus, to the extent to which social institutions prime individuals' identities and act as elicitation, framing, and anchoring devices, to this extent the institutions also shape preferences.*

---

<sup>2</sup> Earlier surveys of the literature on endogenous preferences are Bowles (1998) and Lichtenstein and Slovic (2006, chapter 1).

Finally, in the third part, we explore some potential economic consequences of endogenous preferences. The impact of social institutions on preferences may contribute to the persistence of group inequality. The impact of violent conflicts on preferences may contribute to the difficulty of resolving the conflicts. Minimum wage laws may create entitlement effects. Changes in the way of *eliciting* preferences (*e.g.*, the use of default options) may affect the extent to which individuals *construct* their preferences, in turn inducing swings in the choices they make. The development of the welfare state may change the preferences parents instill in their children. Most of the implications we discuss in this part have the status of untested hypotheses and so provide a rich menu of opportunities for empirical research.

### **1. Economists' Reluctance to Explain Behavioural Changes by Preference Changes**

As discussed above, the standard approach in economics is to begin by characterizing individuals' preferences and the constraints that individuals face when making choices. Then changes in behaviour are explained as responses of optimizing agents to changes in prices, information, and technology that change the payoffs and the available set of actions. Preferences are assumed to be exogenous and stable with regard to the changes in the constraints. There is no doubt that this approach has been very useful, and in many applications the assumption that preferences are – at least in the short run – unaffected by changes in constraints may be justified. However, as we will argue later, the assumption that social institutions generally leave preferences unaffected is implausible in view of the evidence that we discuss in Section 2. In addition, this assumption may well prevent us from a deeper understanding of interesting and important phenomena, as we discuss in Section 3.

The convention against invoking preference changes as an explanation for behavioural changes has been very strong in economics. In some circles, invoking preference changes was condemned as bad science: "... assumptions of unstable tastes have been a convenient crutch to lean on when the analysis has bogged down. They give the appearance of considered judgment, yet really have only been *ad hoc* arguments that disguise analytical failures" (Stigler and Becker, 1977, p. 89).

One reason for the opposition to invoking preference changes as explanatory factors is the fear that such explanations introduce so many free variables that the researcher can explain

everything, therefore he explains nothing. Any strange behaviour, or behavioural change, can be “explained” by positing a direct preference for the behaviour. Obviously, such explanations are empty and meaningless unless there are constraints on the set of preferences that one can reasonably put forward as an explanans.

Where should the constraints imposed on preference-based explanations come from? One way to discipline such explanations is to apply knowledge about the causal determinants of preferences. Historically, two obstacles have stood in the way of producing this knowledge. First, the ability to infer causality from non-experimental data was quite limited, and the econometric tools that helped overcome some of these limits (*e.g.* natural experiment approaches, instrumental variables) have been widely applied only in the last two decades. Second, economics and the social sciences generally had a poor understanding of the “nature” of risk, time, and social preferences because the standard approach – expected utility theory, exponential discounting theory, and the assumption of purely self-interested preferences – failed to capture important preference patterns. It is not possible to properly measure preference change if one lacks a full understanding of the fundamental patterns of preferences. What appears as a change in preferences under the wrong theory may, in fact, reflect a stable preference.<sup>3</sup>

Independent empirical measures of preferences and of preference change are indispensable for disciplining preference explanations. At the time when Stigler and Becker wrote their paper, such preference measures were not available. However, progress in experimental and econometric methodology over the last 20 years now makes it possible to measure preferences much more precisely. Today, there is, for example, a relatively large literature on risk (*e.g.* Bruhin, Fehr-Duda, and Epper 2010; Starmer 2000), time (*e.g.* Frederick, Loewenstein, O’Donoghue 2002), and social preferences (*e.g.* Fehr and Schmidt 2003, Cooper and Kagel, forthcoming) that provides constraints on the types of preferences that one can reasonably invoke. The experimental and econometric tools available today enable researchers to get much

---

<sup>3</sup> The following example illustrates this. Suppose that a subject is motivated by negative reciprocity, i.e. he responds to hostile acts with hostility (Falk and Fischbacher 2006). In the ultimatum game this means that the subject – if in the role of a responder – rejects low, unfair, proposer offers because the rejection punishes the proposer for his unfairness. Suppose now that the responder is put in a different condition, in which the same low offer is generated by a random mechanism, implying that the proposer is not responsible for the low offer. A responder motivated by negative reciprocity will not reject this offer because the randomly generated low offer is not a signal of unfair proposer behavior. If one assumes a wrong social preference model that neglects the true underlying responder preference, the change in behavior across conditions appears to reflect a change in preferences while in fact, the change in behavior is fully compatible with a stable preference for negative reciprocity.

closer to the structural preference parameters that can predict behavior. Thus, the argument that preference explanations provide “endless degrees of freedom” (Stigler and Becker, p. 89) is no longer valid. The progress in scientific methods renders this argument obsolete.

Moreover, as economists, we tend to overlook the many degrees of freedom provided by our standard framework of explaining behavioural changes by changes in the constraints. Game theory has become an extremely flexible tool that can rationalize a vast amount of different phenomena by tweaking assumptions about the extensive form of the game, such as the structure of moves, the choice variables (*e.g.* prices or quantities), or the asymmetry of information. Typically, there is no direct evidence available to constrain the specification of the extensive form of the game, which could lead one to complain about the “endless degrees of freedom” provided by economic theory to explain empirical patterns.

Exactly this complaint was put forward by John Sutton in his presidential address at the Congress of the European Economic Association in 1989, when he discussed the state of the art at that time in Industrial Economics.<sup>4</sup> Sutton wrote:

“Paradoxically, it is the very success of these game theoretic models in providing a rich menu of candidate 'explanations', which leaves them open to a quite fundamental line of criticism. ... In modelling any particular situation, there is usually considerable scope for designing the 'structure of moves' (the extensive form of the game) in various ways, each of which seems reasonable on *a priori* grounds. ... This richness of possible formulations leads to an often embarrassingly wide range of outcomes supportable as equilibria within some 'reasonable' specification. ... In 'explaining' everything, have we explained nothing? What do these models exclude?” (Sutton, 1990, p. 507)

Thus, taken together, there are neither methodological nor substantive arguments that favour changes in constraints over changes in preferences as explanatory factors. Ultimately, it is an empirical question which would provide the better explanation. However, explanations in terms of changes in preferences will be convincing only if we have clean measures of preference

---

<sup>4</sup> “The consequent elaboration of a richer class of game theoretic models has been remarkably successful in one respect: given any form of behaviour observed in the market, we are now quite likely to have on hand at least one model which 'explains' it – in the sense of deriving that form of behaviour as the outcome of individually rational decisions. New 'explanations' of this kind have been adduced across the entire range of the subject: from predatory pricing to vertical restraints, our tool kit has been enormously enhanced” (Sutton, 1990, p. 506).

changes and if we gain a better understanding of the factors that affect preferences. In this regard, economics, and the social sciences more generally, are still in their infancy. We still know relatively little about how economic, social, and biological factors, and the interactions between them, shape preferences.

## **2. Psychological Mechanisms Underlying the Susceptibility of Preferences to Social Influences**

### *2.1. Context-dependent visual perception*

People often assume that what they see with their eyes is a correct representation of reality. We see a rabbit and believe it is a rabbit, or we see a duck and believe it is a duck. We see two lines of unequal length and believe they are of unequal length. However, vision research has produced many examples indicating that our subjective perception of objects is not uniquely given. Instead, perception is shaped by *context* and *memory*. To illustrate the effect of context, consider Figures 1 – 3.

In the first figure, we perceive square A as darker than square B. However, because A is set against a background of lighter squares, whereas B is set against a background of darker squares, our perception is distorted. It is easy to check by putting the same frame on each square that each is the same colour. In the second figure, the length of the two lines is identical, but the context produces the perception that the left line is shorter than the right one. The third figure shows a duck and a rabbit, but at a given moment in time one can see only a duck or a rabbit. Moreover, the first time one looks at this picture, it is quite difficult to generate a perception shift, *i.e.* to suppress the first impression and see the “hidden” animal.

[Insert Figures 1-3 about here]

### *2.2. Context-dependent preferences – anchoring, framing, and multiple social identities*

A considerable literature in psychology and economics suggests that, like visual perception, preferences are also *context-dependent*; they may be influenced by the elicitation method (Lichtenstein and Slovic 1971 and Hsee 1996), the presentation (framing) of the problem



(Kahneman and Tversky 1979), the “anchor” (Ariely, Loewenstein, and Prelec 2003), and the salience of an aspect of one’s social identity (Hoff and Pandey 2006, 2011; LeBoeuf, Shafir and Bayuk 2010; Benjamin, Choi, and Strickland 2010).<sup>5</sup>

A striking example of the power of anchors to shape preferences is the study of Ariely, Loewenstein, and Prelec. This study elicits the willingness to pay (WTP) for various consumption items (a cordless keyboard, a bottle of wine, Belgian chocolates, etc.). For each item, the subjects are asked whether they are willing to pay more or less than a price based on the last two digits of their social security number. For example, if these digits are 53, the “anchor” price would be \$53. After this anchoring question, the experiment elicits WTP in an incentive-compatible way. Subjects with social security numbers above the median display a considerably higher WTP. One explanation for the malleability of subjects’ WTP is that they are not good at predicting the pleasures and pains produced by a purchase. To address this concern, the study includes a follow-up experiment in which consumers can experience the full extent of this pleasure or pain just before the pricing decision. This experiment exposes subjects to an annoying sound. Then subjects are asked, hypothetically, if they would be willing to listen to the sound again in exchange for a given price. This price is the anchoring device. After this anchoring question, the experiment elicits the subject’s willingness to accept money in exchange for hearing the annoying sound in an incentive-compatible way. Again, a low anchor generates substantially lower bids than a high anchor does.

Hoff and Pandey (2006, 2011) show that by changing what is salient to an individual, slight changes in the social contexts in which individuals find themselves will elicit different preferences regarding effort provision. In the traditional caste order, which still more or less prevails in villages in north India, high-caste individuals are always socially superior to low-caste individuals, and low-caste individuals are marked as “unclean” and forced to live and work

---

<sup>5</sup> While framing and elicitation effects undeniably exist, their prevalence or importance may easily be misinterpreted if beliefs about others and preferences cannot be separately measured. One example of this is given by the study of Ross and Ward (1996) on the effects of framing in a repeated Prisoners’ Dilemma game. When framed as a “Wall Street game,” the defection rate is much higher than when the PD is framed as a “community game.” This effect has often been interpreted as a change in *social preferences* due to framing, whereas it could easily be the result of a change in *beliefs* about other people’s cooperation. It is well-known that many subjects are conditional cooperators (see *e.g.* Fischbacher, Gächter and Fehr 2001), that is, they are more likely to cooperate if they believe that others cooperate with high probability. In fact, recent work by Ellingsen *et al.* (2011) shows that the Wall Street *vs.* community game framing leaves cooperation choices unaffected if the belief effect is ruled out. Thus, the frame appears to affect beliefs about others’ social preferences, rather than the preferences themselves.

in areas segregated from the high-caste individuals. Fables, history, and the examples of continued atrocities against upwardly mobile low-caste individuals make children aware of these ideas at an early age. In this set of experiments, high-caste and low-caste junior high school boys in India solve mazes under piece rate incentives in a session consisting of six subjects. Three different conditions vary the salience of their caste identity. For some subjects, caste identity is not made public in the session. This is the control condition. For other subjects, caste identity is made public in a session consisting of three high-caste and three low-caste individuals. For the third group, the conditions are the same as for the second group except that a session consists of only members of high-status castes or only members of low-status castes. This is the caste-status-segregated condition.

The study finds that the high-caste subjects solve significantly fewer mazes in the caste-status-segregated sessions than in either the unsegregated condition where caste identities are made public, or in the control condition, where caste identities are not made public. Subjects solve mazes in two 15-minute rounds. In round 1, the high-caste subjects in the caste-status-segregated condition solve 18% fewer mazes than do the high-caste subjects in the two other conditions. In round 2, by which time there has been more opportunity for improving one's maze-solving skills, the comparable divergence across conditions is 27%.

Under the piece rate incentive scheme, the output and payoff to an individual are completely independent from the output of the other individuals – individual output thus depends only on the individual's preferences and ability. There is no plausible reason why the *ability* of the high-caste subjects should be impaired by placing them in sessions of only high-caste boys rather than in sessions of three high-caste and three low-caste boys. The phenomenon of "stereotype threat/stereotype susceptibility" would predict that because the high caste is stereotyped as naturally superior, the greater the situational cues to caste identity, the more self-confident the high-caste subjects should feel. With greater self confidence, they should perform better, not worse. The results in this study on the number of subjects who completely fail to learn how to solve a single maze are consistent with the stereotype threat/susceptibility phenomenon: the high-caste subjects are much less likely to fail when their caste identity is made public, and the reverse is true for the low-caste subjects.

A plausible interpretation of the decline in the performance of the high-caste subjects is that the segregation of individuals by caste status is a strong cue to the caste system, in which social status is ascriptive. Given the sense of entitlement of the high caste under the caste order, making caste highly salient activates a mental frame in which a high-caste person has less need to achieve because his status is based on caste and family, not individual performance. Under the evoked mental frame, high-caste subjects see less need to work hard. This interpretation suggests that human preferences are not uniquely determined but leave opportunity for variation, and slight differences in situations can cue different mental frames and lead to the expression of different preferences. In other words, we have learned from the society in which we live a variety of roles that can be situationally evoked, and the conditions in Hoff and Pandey's study act as "frame switches" that evoke different roles. An elaboration of this way of thinking about culture is developed in two important papers in sociology, Swidler (1986) and DiMaggio (1997), which seek to reformulate culture's causal role in shaping action. In the old view in sociology, culture imparts values that are consistent across situations, and the values explain action. In the new view, culture shapes behavior not as a latent variable, but rather as sets of frames that are situationally evoked and that determine which actions seem possible and desirable, given a person's values.

The study by LeBoeuf *et al.* explicitly investigates the hypothesis that people have multiple identities, and that making one identity more salient than others would evoke different norms and values. As a consequence, depending on the identity that is made salient, the subjects are hypothesized to show different behavioural preferences. The authors conjectured, in particular, that for Asian-American subjects, it is possible to trigger their Asian identity associated with norms of cooperation and collectivism, or their American identity that puts less emphasis on cooperation and more emphasis on individualistic values. Different identities are primed by a few survey questions before the experimental task. These were questions like "Where were you born?" or "What is your favourite Chinese (American) holiday?" After the subjects received either the Asian or the American prime, they provided answers to questions that elicited their cooperation preference in a hypothetical prisoners' dilemma (PD) and their degree of individualism in a hypothetical choice situation in a restaurant<sup>6</sup>. The Asian prime led to a much

---

<sup>6</sup> Degree of individualism was elicited as follows: "You and your friends all love the same restaurant, and whenever you go there, you each order the same thing. One day, however, the waiter mentions a new, somewhat exotic dish that

lower defection rate in the PD and to a much lower willingness to make an individualistic food choice in the restaurant.<sup>7</sup>

Benjamin, Choi and Strickland (2010) extend the priming technology applied in the LeBeouf study to show that the Asian prime also renders Asian-American subjects more patient. In contrast to the LeBeouf et al. study, in Benjamin et al. subjects' preferences were elicited in an incentive-compatible way, lending further credibility to the notion that Asian-American subjects have multiple identities. LeBeouf *et al.* also show that the notion of multiple identities is not restricted to ethnic categories; it is possible to prime a "family-oriented" identity, which triggers values related to family obligations, or an "occupation-oriented" identity associated with obligations to one's firm. Thus, just as the visual frames illustrated in Figures 1 and 2 shape perception, making an individual's caste status public or using a simple "background questionnaire" can render certain identities more salient and induce the subjects to align their behaviour with the values and commitments associated with that identity.

The examples mentioned above suggest that individuals may have multiple preference orderings that can be shaped by the social environment and the prevailing institutions. While the notion that an individual may have multiple preference orderings may sound strange, it is no more unusual than the notion that one and the same picture can represent two different – mutually exclusive – animals, as in Figure 3. Very small changes in the visual environment may have a decisive influence of which animal one can see. If one slightly covers, for example, the rabbit's nose in Figure 3, one can see only the duck; if one covers part of the duck's bill, one can only see the rabbit. This means that slight changes in the lighting conditions that change the ease with which the rabbit's nose or the duck's bill can be seen will have a decisive influence on which animal an observer perceives. In the same way, slight changes in the social mindset in which individuals find themselves as a result of priming one or another aspect of an institution

---

sounds quite good to you. To your surprise, your friends do not find the special dish as interesting as you do. In fact, one of them says it sounds "weird." This makes you somewhat nervous about the meal, but still, you find it appealing. The time comes to place your order. What do you think you would do? Order the traditional meal that I always order. Order the new, special meal, regardless of what my friends say."

<sup>7</sup> The prime's impact on cooperation behavior could be due to its effect on subjects' beliefs about others' cooperation or it could have changed preferences. See also footnote 5. Thus, future research should control for beliefs, and subjects should be bear real consequences from their choices in a PD game instead of answering a hypothetical PD question.

may affect the way in which people think about themselves, *i.e.* they may affect which preference ordering becomes salient in a given situation.

In other words, the susceptibility of preferences to elicitation, framing, anchoring, and identity primes implies that legal institutions and the prevailing interaction patterns may influence human behaviour not only by affecting constraints and beliefs, but also by affecting preferences. In this view, social institutions do not just impose constraints and shape beliefs about others' behaviour, but are also preference-elicitation devices, frames, and anchors that may render particular identities, and thus particular values and normative commitments, more salient.

### **3. The Potential Economic and Social Consequences of Endogenous Preferences**

The impact of social practices and institutions on preferences may have far-reaching consequences, but we still know far too little to trace these consequences with any certainty. This section has, therefore, a more speculative nature. Based on evidence that suggests that “society” influences preferences, we speculate about potentially important implications of these findings. The reader may thus view this section as the outline of a research program that finally aims at identifying the hypothesized implications empirically.<sup>8</sup>

#### *3.1. Endogenous Preferences and the Persistence of Inequality between Social Groups*

Take, for example the finding in Hoff, Kshetramade and Fehr (2011), published in this symposium, that suggests that low-caste status has a negative influence on the willingness to sanction social norm violations altruistically. Compared to a member of a high-status caste, a member of a low-status caste is less willing to punish a social norm violation that hurts a member of his own caste “community” (in the sense of the Hindu word *jati*, the endogamous group within which most social relationships occur). This negative influence of low caste status persists even if we control for wealth, education, and the strength of the prevailing social norms. Both low and high castes consider it as a clear violation of a social norm if an individual does not reciprocate a favour and, thus, both expect harsh punishment of the norm violation by third parties. Nevertheless, third parties from low castes show a markedly lower propensity to punish

---

<sup>8</sup> The interested reader may also consult Etzioni (1985) and Saez-Marti and Zilibotti (2008) who describe further implications of endogenous preferences.

non-reciprocation regardless of whether the norm violator or the victim of the violation has low- or high-caste status.

Since contracts involve the mutual reciprocation of favours, an immediate potential implication of this finding is that low castes may be generally less willing to enforce contracts through altruistic sanctioning. Previous evidence (Fehr, Kirchsteiger and Gächter 1997) indicates that altruistic sanctioning is a powerful means of enforcing contracts. A lower willingness to sanction contract violations may thus render the low castes less able to exploit gains from trade because they are less protected – via lower altruistic punishment – from contract violations. In addition, if the lower willingness to punish contract violations is more generally associated with a lower propensity to punish free-riders in collective action, then the low castes would be less able to discipline free-riders and thus to organize collective action. This reduced ability for collective action might have consequences for their ability to improve their status through political means. Taken together, this chain of arguments might mean that the caste system exerts a self-perpetuating influence on social preferences that renders the low castes less able to change the caste system. The endogeneity of the willingness to sanction altruistically thus may be part of a vicious circle that contributes to societal rigidity.

### *3.2. The Potential Role of Endogenous Preferences in Group Conflicts*

One reason why violent conflicts between groups and societies are often so difficult to end is that in the process of war, preferences may change. Voors et al. (forthcoming) show that the intensity of civil war has a significant impact on time, risk and social preferences. Individuals in Burundi who happen to live in regions with higher levels of violence are more risk-seeking and less patient in a preference elicitation experiment. In addition, the individuals in those violent regions are more altruistic towards each other. These findings are robust to controls for a wide range of household and community characteristics, including the application of plausible instrumental variables. Likewise, Gneezy and Fessler (2011) provide experimental evidence that during the recent war between Israel and Hezbollah, individuals were more willing to sanction others for greedy behaviour, or to reciprocate generous acts, than these populations had been nine months before the war, or were one year after the war. Bauer *et al.* (2011) conducted an experiment with children that shows that those children who were more affected by warfare during the conflict between Russia and Georgia display more altruism towards in-group members. Note that these

findings establish more than the mere existence of in-group favoritism; they suggest that inter-group conflicts may increase it.

The effect of group conflicts on time, risk, and social preferences may affect the dynamics of group conflicts. For example, if a war increases in-group altruism, people will be more likely to contribute to collective action against the enemy, which may perpetuate the inter-group conflict.<sup>9</sup> Likewise, if a war increases the willingness to take individual risks, individuals will be more willing to engage in risky behaviors against the enemy. Finally, if wars make people impatient, this impatience may also extend to the perceived future returns from peace. If these returns are discounted more steeply, people will be less willing to support peace agreements between the conflicting parties. Thus, the potential impact of inter-group conflict on preferences introduces a whole new dimension of effects that may help explain why reaching a peace agreement in entrenched conflicts is so difficult.<sup>10</sup>

### *3.3. Endogenous Preferences and the Role of the Law: Entitlements and Default Rules*

Preferences may also be directly affected by legal changes, such as an increase or a decrease in minimum wage laws. Standard economic models of the labour market assume that workers' reservation wages are not *directly* affected by changes in the legal minimum wage. Of course, for a *given* reservation wage, variations in the minimum wage may affect the willingness of workers to accept a job, but the legal minimum wage is assumed to have no effect on the reservation wage. In a recent laboratory minimum wage experiment, Falk, Fehr and Zehnder (2006) show that the subjects' reservation wages differ depending on whether they have previously experienced a situation with a minimum wage or not. Subjects who had experienced a minimum wage previously, but for whom the minimum wage was abolished, displayed much higher reservation wages compared to subjects who had never experienced a minimum wage. Thus it

---

<sup>9</sup> The findings of Abbink *et al.* (2010) also suggest that the presence of intergroup conflict may render subjects so willing to punish those who don't contribute their fair share to the group's collective action that the groups collectively spend more on the "prize" over which they are fighting than it is worth to them.

<sup>10</sup> There may be an additional effect of conflict on preferences through the effect on shared understandings. The capacity for collective action is a matter not just of groups' sharing interests, but also of a shared understanding of the elements of a problem and its solution. In this view, the failure to act collectively can stem from the groups' failure to ascribe a common meaning to a given situation. This problem may be especially severe in countries characterized by endemic conflict, where institutions that might otherwise provide frames for a common understanding of a situation have been weakened or destroyed (Gauri, Woolcock, and Desai, 2011).

seems that minimum wage laws create a kind of entitlement effect that affects subjects' preferences, because they are less willing to work for a given wage after the experience of a legal minimum wage.

These findings suggest that laws have a much deeper impact on human behaviour than just changing constraints and beliefs. They may also affect what is considered as normatively legitimate, which may directly affect economic outcomes. In Falk, Fehr and Zehnder, the impact of history on subjects' reservation wages changes actual wages and employment in the experiment. There are hysteresis effects not, as in the standard model, through *e.g.* fixed costs, but instead through changes in preferences.

An anomaly that has attracted a great deal of attention is the effect of default options on decisions. Default options appear to have very large consequences for choices that people make over retirement savings, insurance plans, and organ donation. For instance, in Austria, consent to being an organ donor is the default option, so that not consenting requires an active decision to opt out. Here the effective consent rate is 99.98% (Johnson and Goldstein 2003). Just over the border in Germany, consent to being an organ donor requires opting out of the default option, and the proportion of individuals who are organ donors is 12%.

Several factors consistent with the standard economics view that individuals have fixed preferences could contribute to this vast difference in consent rates: Filling out a form to opt out may have a high disutility. The default option may suggest that the norm is to contribute and individuals may value meeting others' expectations, or they may implicitly reciprocate a favor if they believe that most others will also be organ donors. Perhaps differences in education, transplant infrastructure, or religion influence contributions. But another possible explanation, outside the classical economics view, is that preferences for organ donation are constructed at the time of decision. In that view, choosing the default option is favored because it avoids the cost of making an active decision.

A test that can provide evidence for the constructed preference view is that individuals who have not made up their mind, but not other individuals, can be nudged to contribute to the public good by a manipulation that makes them think about it. In this symposium, Stutzer, Goette, and Zehnder (2011) test this prediction in a field study of blood donations by the Swiss Red Cross in Zurich. All potential donors are asked in a questionnaire to indicate YES or NO to the question



“Do you feel sufficiently informed about the importance of donating blood?” The study randomly assigns individuals to three groups: in the control group there is no intervention; in the “strong active decision group” (hereafter strong AD), individuals are asked to either agree to participate at a specific date and time, or alternatively to refuse to participate in the blood drive by checking a box. The third group, the “weak AD group,” is the same as the strong AD group except that an individual is given the *additional* option to check a box indicating that he does not want to make a decision now about donating blood. Under standard assumptions, where individuals are assumed to know their preferences, the AD interventions should have no effect. Under the constructive preferences view, the AD interventions would have a larger effect on the subpopulation that felt they were not sufficiently informed about the importance of blood donation.

The findings support the hypothesis that preferences in many domains do not exist for many individuals. On the one hand, individuals who have not made up their mind about blood donations can be nudged to donate by a subtle manipulation that makes them think about donating blood. On the other hand, the treatment is ineffective for those who have already made up their mind. The difference in how individuals in the two groups respond to the two treatments is significant. Despite the subtlety of the treatment, it leads to a statistically significant effect in the predicted subgroup of subjects. Moreover, the effect is quantitatively large—the strong AD intervention increases the proportion of blood donors by 8 percentage points, which represents a doubling of the sample average of 8 percent. This paper demonstrates a precise mechanism through which default options work: Some individuals have not formed a preference, and those individuals tend to choose the option that permits them to avoid making an active decision. If there is no such option—that is, if individuals have to check a box to explicitly agree or explicitly refuse to give blood—then more individuals will construct a preference. As long as some of those who are forced to construct a preference wish to give blood, the AD treatment will increase the number of individuals who freely show up (some days later) to donate blood. Stutzer *et al.* provides new evidence against the classical economics view that individuals have fixed preferences that are easily accessible to them.

### 3.4. *The Size of the Welfare State, the Demand for Social Insurance, and Parents' Incentive to Instill Work Norms in their Children*

A longstanding puzzle is the considerable difference between Western Europe and the US in the size of the welfare state. Several reasons have been put forward as an explanation for this difference; among them are differences in political institutions, and ethnic heterogeneity (Alesina and Glaeser 2004) or self-fulfilling beliefs and multiple equilibria (Piketty 1995, Bénabou and Tirole 2005). However, it may also be the case that the willingness to migrate to the US has historically been associated with different ideologies and these ideologies may have affected the population's preferences. In other words, the US population may have a different demand for social insurance and this may be another reason for the differences in the welfare state. A recent paper (Naef et al. 2011) provides evidence for the hypothesis that the US population is less risk averse than the German population. Whether this preference difference holds more generally between the US and Europe is, however, not known.

Eugster et al. (2011), in this symposium, study the impact of differences across language groups in the demand for social insurance across the cultural boundary defined by German and "Latin" (*i.e.* French, Italian, and Romansh) language groups in Switzerland. The authors use a regression discontinuity design to identify cultural differences in the demand for social insurance. The Swiss case is well-suited for this research design because on both sides of the language border exactly the same legal system prevails. This means that the supply of social insurance (*e.g.* unemployment insurance, the retirement system, maternity leave, etc.) is identical. In addition, the wealth distribution, the probability of becoming unemployed, and other risks are similar on both sides of the border. Eugster *et al.* use data from referenda over the period 1980-2009 to document a persistent difference in the demand for social insurance – the German group expresses a much lower demand than the Latin group. These two groups are at opposite ends of scale defined by average responses in countries to the question whether government should do more to redistribute income. Eugster *et al.*'s findings suggest that even among groups with the same economic fundamentals, there can be large and long-lasting differences in the demand for social insurance.

A neat example of how the welfare state may shape preferences with regard to labor supply is Lindbeck and Nyberg (2006). They explicitly examine the influence of the welfare state on parents' incentive to instill preferences for hard work ("work norms" for short) in their children.

Altruistic parents have a reason to care about their children's preferences for work because if the grown-up children suffer from economic hardship, the parents will want to help them. The stronger the work norms of the children, the less will the parents have to help them. Altruistic parents have, therefore, an economic reason to instill strong work norms in their children. However, if the welfare state takes over the obligation to help children who happen to suffer economic hardships, the parents have less incentive to instill work norms in their children. Lindbeck and Nyberg develop a theory that formalizes these intuitions. Interestingly, data from the World Values Survey in 1998 are consistent with their account. The seven countries that rank lowest in terms of the share of respondents who consider a willingness to work hard important to learn at home are advanced European welfare states. Denmark, where only 2 percent of the respondents hold this view, ranks last out of 42 countries, followed by Sweden, Finland, Norway, the Netherlands, Austria, and West Germany.<sup>11</sup>

#### **4. Conclusion**

Over two decades ago, when economists were laying the groundwork for an economics in which information was endogenous, Stiglitz (1985) argued that the standard theory “is not robust to slight alterations in the informational assumptions. It is but a special—and not very plausible—‘example’ among the possible set of informational assumptions which can be employed to characterise an economy” (p. 21). Here we have suggested that the assumption of exogenous preferences—unaffected by the outcomes in society—is but a special and not very plausible example among the possible set of assumptions about preferences that can be employed in explaining economic outcomes.

A considerable body of research suggests that institutions may have effects that go far beyond the incentives that they provide – they may create framing and priming effects that influence the accessibility of memories and the perception, interpretation and, hence, the meaning of facts. It is possible that exposure to certain institutions and environments influences whether in a figure such as Figure 2, some societies have the ability to see that the two lines are

---

<sup>11</sup> Another interesting paper on the importance of endogenous preferences is Doepke and Zilibotti (2008). They develop a theory of preference formation that is a key component in explaining the reversal of fortunes between the landowners and the emerging bourgeoisie at the outset of the industrial revolution. In their approach, economic incentives induce parents to instill particular preferences in their children (e.g. a strong work ethic and a low discount rate in the case of the middle classes at the outset of the industrial revolution); these preferences in turn have important effects on economic development and social stratification.

equal, while people in other societies lack this ability. The cross-cultural experiments by Segall et al. (1966) document striking cultural differences in the susceptibility of people to the Müller-Lyer illusion. In a few societies, the illusion was completely absent while among European and US subjects the illusion was strongest. Segall *et al.* (1966) suggest that the exposure to “carpentered corners” of modern environments may have favoured certain optical calibrations and visual habits that perpetuate the illusion. Thus, in line with Henrich, Heine, and Norenzayan (2010, p. 64), one may ask the question: If “even a process as apparently basic as visual perception can show substantial variation across populations, .... what kind of psychological processes can we be sure will not vary?”

The recent work in psychology and economics has implications that are only beginning to be explored for the examination of the question how different aspects of society affect preferences, but which promise to undermine long-held standard assumptions in economics. If historical institutions shape preferences, which shape the choice of public policies and their effects, then societies almost surely have different optimal paths of development. If individuals understood these processes, they could design policies to nurture or dampen desirable preferences and in turn shape the paths of economic change. At a minimum, these findings should give pause to those who accept WTAs, WTPs, and market prices as stalwart guides to welfare.<sup>12</sup>

In a plea for bringing cultural elements into economic models, McCloskey (1998) imagines the following riposte by an economist defending the standard paradigm in which there is no role for culture: “Give me a break: I’m not in the business of explaining all behaviour. I propose merely to explain some portion, and in many cases a large portion.” (p. 308)

If culture had just a “level” effect on the performance of economies, something like the initial “imprinting” of preferences with anchors that was done in Ariely *et al.* (2003), but an imprinting that was done just once and so could be treated as a fixed effect, then this objection would stand on solid ground. Economists would not be able to explain certain level effects—leave that to the Department of Psychology or Sociology, as McCloskey’s imagined interlocutor says—but they could still explain the behavioural changes in response to changes in constraints. A growing literature provides some evidence, on the contrary, that preferences change in

---

<sup>12</sup> For an early discussion of the consequences of endogenous preferences for welfare economics see Gintis (1972).

response to changes in social institutions and that such changes may be central to the process of economic and societal change—in the degree and persistence of group inequality, the tenacity of hostilities between warring groups, the pressures on the European welfare state, to name just a few. We look forward to seeing the new areas of research that the perspective of endogenous preferences will open up.

## References

- Abbink, K., Brandts, J., Herrmann, B. and Orzen, H. (2010). 'Intergroup conflict and intra-group punishment in an experimental contest game', *American Economic Review*, vol. **100**(1), pp. 420-477.
- Alesina, A. and Glaeser, E. L. (2004). *Fighting Poverty in the US and Europe – A World of Difference*, Oxford: Oxford University Press.
- Ariely, D., Loewenstein, G. and Prelec, D. (2003). 'Coherent arbitrariness: Stable demand curves without stable preferences', *Quarterly Journal of Economics*, vol. **118**(1), pp. 73-105.
- Bauer, M., Cassar, A., Chytilova, J. and Henrich J. (2011). 'Warfare increases egalitarian and parochial motivations in children', mimeo, Charles University Prague.
- Becker, G. and Stigler, G.J. (1977). '*De gustibus non est disputandum*', *American Economic Review*, vol. **67** (2), pp. 76–90.
- Bénabou, R. and Tirole, J. (2005) 'Belief in a just world and redistributive politics', *Quarterly Journal of Economics*, vol. **117**, pp. 871-915.
- Benjamin, D.J., Choi, J.J. and Strickland, A.J. (2010). 'Social identity and preferences', *American Economic Review*, vol. **100**(4), pp. 1913-1928.
- Bowles, S. (1998). 'Endogenous preferences: The cultural consequences of markets and other economic institutions', *Journal of Economic Literature*, vol. **36** (1), pp. 75-111.
- Bruhin, A., Fehr-Duda, H. and Epper, T. (2010). 'Risk and rationality: uncovering heterogeneity in probability distortion', *Econometrica*, vol. **78**(4), pp. 1375 – 1412.
- Cooper, D. J. and Kagel, J. (forthcoming). '*Other Regarding Preferences: A Survey of Experimental Results*'. in J. Kagel and A. Roth (eds.), *The Handbook of Experimental Economics*, Vol. 2, Princeton: Princeton University Press.
- DiMaggio, P. (1997). 'Culture and Cognition', *Annual Review of Sociology*, vol. **23**, pp. 263-287.
- Doepke, M. and Zilibotti, F. (2008). 'Occupational Choice and the Spirit of Capitalism'. *Quarterly Journal of Economics* 123(2), pp. 747-793.

- Ellingsen, T., Johannesson, M., Mollerstrom, J., Munkhammar, S., (2011). 'Social framing effects: Why do situational labels affect cooperation?', mimeo, Stockholm School of Economics. Available at: <http://www2.hhs.se/personal/ellingsen/pdf/FramingEconQ1.pdf>.
- Etzioni, A. (1985). 'Opening the preferences: a socio-economic research agenda', *Journal of Behavioral Economics*, **14**(1), pp. 183-198.
- Eugster, B., Lalive, R., Steinhauer, A. and Zweimuller, J. (2011). 'The demand for social insurance: does culture matter?' *Economic Journal*, this issue.
- Falk, A., Fehr, E. and Zehnder, C. (2006). 'Fairness perceptions and reservation wages—the behavioral effects of minimum wage laws', *Quarterly Journal of Economics*, vol. **121** (4), pp. 1347-1381.
- Falk, A. and Fischbacher, U. (2006). 'A theory of reciprocity'. *Games and Economic Behavior*, vol. **54**(2), pp. 293-315.
- Fehr, E., Gächter, S. and Kirchsteiger, G. (1997). 'Reciprocity as a contract enforcement device: experimental evidence', *Econometrica*, vol. **65**(4), pp. 833-860.
- Fehr, E. and Schmidt, K. M. Schmidt (2003). 'Theories of fairness and reciprocity—Evidence and economic applications,' in M. Dewatripont, L.P. Hansen, and S. Turnovsky, (Eds.) *Advances in Economics and Econometrics*. vol. 1, Cambridge University Press, pp. 208-257.
- Fischbacher, U., Gächter, S. and Fehr, E. (2001). 'Are people conditionally cooperative? Evidence from a public goods experiment', *Economics Letters*, vol. **71**(3), pp. 397-404.
- Frederick, S., Loewenstein, G., O'Donoghue, T. (2002). 'Time discounting and time preference: a critical review', *Journal of Economic Literature*, vol. **40**(2), pp. 351-401.
- Gauri, V., Woolcock, M., and Desia, D. (2011). 'Intersubjective meaning and collective action in 'fragile' societies', World Bank Policy Research Working Paper 5707, Washington DC.
- Gintis, H. (1972). 'A radical analysis of welfare economics and individual development', *Quarterly Journal of Economics*, vol. **86**(4), pp. 572-99.
- Gneezy A. and Fessler, D. T. (2011). 'Conflict, sticks and carrots: war increases prosocial punishments and rewards', *Proceedings of the Royal Society B*, published online before print. June 8, 2011, doi:10.1098/rspb.2011.080.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R. (2001). 'In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies'. *American Economic Review*, 91(2), 73-78.

- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barret, C., Bolynatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesogorol, C., Marlowe, F., Tracer, D. and Ziker J. (2010). 'Markets, religion, community size, and the evolution of fairness and punishment', *Science*, vol. **327**, pp. 1480-1484.
- Henrich, J., Heine, St. J. and Norenzayan, A. (2010). 'The weirdest people in the world', *Behavioral and Brain Sciences*, vol. **33**, pp. 61-135.
- Hoff, K., Fehr, E. and Kshetramade, M. (2011). 'Caste and punishment: the legacy of caste culture in norm enforcement', *Economic Journal*, this issue.
- Hoff, K. and Pandey, P. (2006). 'Discrimination, social identity, and durable inequalities', *American Economic Review, Papers & Proceedings*, vol. **96**(2), pp. 206-211.
- Hoff, K. and Pandey, P. (2011). 'Making up people: Experimental evidence on identity and development from caste India', mimeo, World Bank.
- Hsee, C. K. (1996). 'The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives', *Organizational Behavior and Human Decision Processes*, **67**(3), pp. 247-257.
- Johnson, E. J. and Goldstein, D.G. (2003). 'Do defaults save lives?' *Science*, vol. **302**, pp. 1338-1339.
- Kahneman, D. and Tversky, A. (1979). 'Prospect theory: An analysis of decision under risk', *Econometrica*, vol. **57**, pp. 263-291.
- LeBoeuf, R. A., Shafir, E., and Bayuk, J. B. (2010). 'The conflicting choices of alternating selves', *Organizational Behavior and Human Decision Processes*, vol. **111**(1), pp. 48-61.
- Lichtenstein, S. and Slovic, P. (2006). *The Construction of Preference*, Cambridge,U.K.: Cambridge University Press.
- Lichtenstein, S. and Slovic, P. (1971). 'Reversals of preference between bids and choices in gambling situations', *Journal of Experimental Psychology*, vol. **89**, pp. 46-55.
- Lindbeck, A. and Nyberg, S. (2006). 'Raising children to work hard: altruism, work norms, and social insurance', *The Quarterly Journal of Economics*, vol. **121**(4), pp. 1473-1503.
- McCloskey, D. N. (1998). 'Bourgeois virtue and the history of P and S.' *Journal of Economic History*, vol. **58**, pp. 297-317.
- Naef, M., Fehr, E., Fischbacher, U., Schupp, J. and Wagner G. (2011). 'Decomposing trust – explaining national and ethnic trust differences'. Working Paper, Department of Economics, University of Zurich.
- Piketty, T. (1995) 'Social mobility and redistributive politics.' *Quarterly Journal of Economics*, vol. **110**, pp. 551-583.

- Ross, L. and Ward, A., (1996). 'Naive realism in everyday life: implications for social conflict and misunderstanding'. In Reed, E.S., et al. (Eds.), *Values and Knowledge*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 103–135
- Saez-Marti, M., and Zilibotti, F. (2008). 'Preferences as Human Capital: Rational Choice Theories of Endogenous Preferences and Socioeconomic Changes'. *Finnish Economic Papers*, 2008, Vol. 21(2), pp. 81-94.
- Segall, M.H., Campbell, D. T. and Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*, Indianapolis: Bobbs-Merrill Co.
- Starmer, C. (2000). 'Developments of non-expected utility theory – the hunt for a descriptive theory of choice under risk', *Journal of Economic Literature*, vol. **37**, pp. 332 – 382.
- Stiglitz, J. E. (1985). 'Information and economic analysis: a perspective.' *The Economic Journal*, vol. **95**, pp. 21-41.
- Stutzer, A., Goette, L. and Zehnder, M. (2011). 'Active decisions and prosocial behaviour: a field experiment on blood donation.' *The Economic Journal*, in press.
- Sutton, J. (1990). 'Explaining everything, explaining nothing? Game theoretic models in industrial organisation'. *European Economic Review*.
- Swidler, A. (1986). 'Culture in action: symbols and strategies'. *American Sociological Review*, vol. **51**, pp. 273-286.
- Van den Berghe, P. (1981). *The Ethnic Phenomenon*, NY: Elsevier.
- Voors, M., Nillesen, M., Verwimp, P., Bulte, E., Lensink, R. and van Soest, D.(forthcoming). 'Does conflict affect preferences? Results from field experiments in Burundi. *American Economic Review*.



Figure 1. Framing effects disguise the fact that the colors in squares labeled A and B are the same.

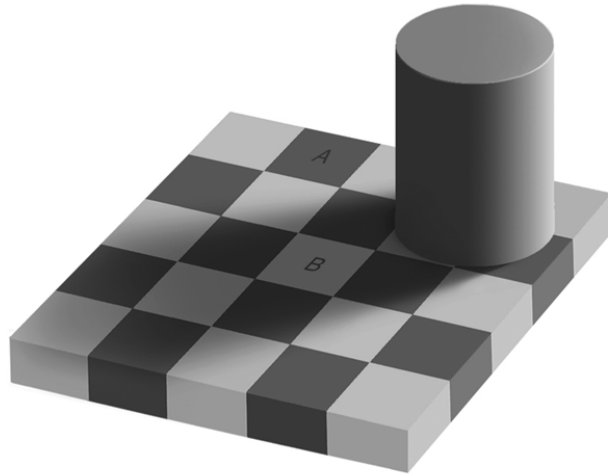


Figure 2. Muller-Lyer illusion disguises the fact that the length of the two lines is the same.



Figure 3. At each moment in time, we can see only the bird or the rabbit.

