TOWARD AN ECONOMIC THEORY OF DYSFUNCTIONAL IDENTITY

By

Hanming Fang and Glenn C. Loury

October 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1483

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Toward An Economic Theory of Dysfunctional Identity[*]

Hanming Fang[†] Glenn C. Loury[‡]

September 27, 2004

Abstract

We advance a novel choice-theoretic model of "identity" based on the notions of *categories* and *narratives*. Identity is conceived as a matter of "reflexive perception" – how people understand themselves. Choosing an identity is equivalent to making a generalization about one's past that highlights the most salient aspects of experience. When many individuals make a common choice in this regard, they embrace a *collective identity* which is *dysfunctional* if it is Pareto dominated by an alternative self-classificatory schema. Using a simple multi-stage risk sharing game, we explore conditions under which *dysfunctional collective identities* might be expected to emerge.

Keywords: Identity; Dysfunctional collective identity.

JEL Classification Codes: Z1, Z13.

[*] The authors thank Chris Barrett, Stephen Coate, Roland Fryer, Antonio Merlo, Stephen Morris and Andrew Postlewaite for helpful discussions. We remain solely responsible for any errors. We gratefully acknowledge the financial support of The Andrew W. Mellon Foundation and The Pew Charitable Trusts.

[†] Department of Economics, Yale University, P.O. Box 208264, New Haven, CT 06520-8264. Email: hanming.fang@yale.edu

[‡] Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215. Email: gloury@bu.edu

1 Introduction

Rigorous thinking about the nature and sources of human identity, and about the links between identity and culture, is vitally important for understanding a variety of significant social problems. Students of the subject have pondered why people embrace one identity rather than another, and how their convictions in this regard affect their economic performance. Numerous, conflicting conceptions of identity can be found in the literature. Psychologists draw a fundamental distinction between *social* identity, which deals with how an individual is perceived and categorized by others, and *personal* identity (sometimes called "ego identity"), which invokes a person's answer to the question, "Who am I?"[1] Goffman (1963) describes an individual's social identity as "the categories and attributes anticipated by others during routines of social intercourse in established settings."[2] Social psychologists also use the concept of *collective identity* to ask how a group of distinct individuals might come to embrace a common answer to the "Who am I?" question, and what follows from their having done so.[3] Thinking about *collective* identity leads naturally to a reflection on how social interaction influences the formation and maintenance of personal identities which, in turn, leads naturally into a discussion of "culture."

This essay explores some connections between identity, culture and economic functioning. In essence, we will do three things here: propose what we believe to be a novel definition of identity; make precise a sense in which the collective identity of a group of people can be said to be *dysfunctional*; and, describe a specific mechanism of social interaction through which rational individuals could nevertheless choose to embrace a way of thinking about themselves that inhibits their economic functioning.

We are motivated in this pursuit by the commonsense observation that – whether looking within or between countries – economic backwardness, multi-generational poverty, and chronic underdevelopment seem to be connected in some way to the "culture" of those who are disadvantaged, and especially to what may be regarded as their "dysfunctional" notions about identity. Thus,

[1] See, for example, the useful survey on "The Self" in the *Handbook of Social Psychology* (Baumeister 1998).

[2] Goffman (1963) uses yet another distinction – between "virtual" and "actual" social identities. The former is a social artifact, an identity constructed "from the outside" via social imputations based on a person's physical presentation. While the latter is relatively objective, an identity constructed "from the inside" via the accumulation of facts specific to a person's biography. Goffman's analysis of "stigma" is all about the interesting drama that unfolds when virtual and actual identities diverge systematically in the social experience of a given individual.

[3] Sidanius and Pratto (2001) and Aronson et al. (2003) are interesting illustrations of how the "collective identity" concept has been used in the social psychology literature.

backward groups within many societies (inner-city blacks in the US, low caste people in India, or gypsies [i.e., Roma] in Europe) have been said to languish because they embrace a "culture of poverty," (Banfield 1970), or because they are obsessed with their own victimhood (McWhorter 2000), or due to their adopting an "oppositional identity" (Ogbu 2003). These culture-identity orientations are said to promote economically self-limiting behaviors (regarding education, savings, or occupational choice), and to cause disadvantaged people to "dis-identify" with success in their respective societies.[4] Often, such conjectures about the root causes of "backwardness" are stated imprecisely and supported only by anecdotal evidence. It thus seems desirable to have a more formal way to talk about how an "identity" could be "dysfunctional." We are taking a small step in that direction with this paper. It is also a commonplace that authors trumpeting the cultural roots of economic backwardness treat "culture" as exogenous. One goal of this exploratory analysis is to consider how strategic interactions among agents in particular environments might incline them to adopt one or another common stance on certain identity questions. In this limited way, at least, we are striving to make "culture" endogenous.

Given the *a priori* plausibility of their connections to growth and inequality, economists have recently made some interesting attempts to model concepts like "identity" and "culture."[5] Generally speaking, this literature takes what might be called a utilitarian approach to the subject. That is, choices are to be utility-maximizing, but a non-standard utility function is posited – one that has been augmented to incorporate the value of conforming to the norms and expectations associated with a decision-maker's social position. A leading example of this approach is found in the work of Akerlof and Kranton (2000), who model "identity" as a combination of "role" and "prescription." Society is partitioned into a set of "types" – occupants of the various roles. These types have identity-influenced preferences which are biased in favor of certain actions – the ones most consistent with the prescriptions attached to their roles. Both roles and prescriptions are exogenous, given by history. Akerlof and Kranton (2002) go on to delineate a set of role/prescription pairs characteristic of a particular venue of social interactions (secondary schools), and to study how the emergent behaviors of role occupants operating in that venue reflect the prescriptive ac-

[4] One popular version of this hypothesis is the suspicion that native-born black Americans fare poorly in school because many think that the doing of academic work is "acting white" [see, e.g., Ronald Ferguson's chapter in the Loury, Teles and Modood volume (in press).]

[5] See, for example, North (1981), Grief (1994), Bernheim (1994), Akerlof (1997), Akerlof and Kranton (2000, 2002), and Fryer (2003). Fryer's work on 'cultural capital,' which also studies an infinitely repeated game as a laboratory for investigating the economic consequences of 'culture,' is the most similar to our own.

tions associated with their roles. One limitation of this approach is that, while it explores the implications of individuals having adopted certain identities, it offers no account of how and why people come to have the identities they have. Thus, it cannot guide an assessment of the *efficiency* of people's identity choices.

In contrast, our proposed theory is driven by cognitive, not utilitarian considerations. Building on ideas about racial classification, social cognition and identity introduced in Loury (2002, chapter 2), and following the categorical approach to cognition pioneered in Fryer and Jackson (2003), we go on to consider the problem of *auto-cognition* – how people see themselves. We ground our approach to identity (and thus, to "culture") in the elemental notions of "categories" and "narratives." Our core idea is that, at its root, personal identity is all about self-perception and self-representation. We are interested in choices about identity made by rational agents anticipating subsequent interaction, who expect their payoffs from this interaction to vary with their identity commitments. Technically, we study a two-stage game in which identity choices are made in the first stage, and agents engage (more or less remuneratively) in some economic interactions in the second stage. Within this framework, we say that a *collective identity* has been adopted when, in subgame perfect equilibrium, individuals make the same first stage choices. We are particularly interested in showing how a group of people might come to embrace an inefficient, or *dysfunctional* identity.

Using the psychologist's terminology, then, ours is a paper about *personal not social* identity – albeit in a multi-agent, interactive setting. We formally explore how people with ongoing economic relations might arrive at an answer to the "Who am I?" question. We will say that a person's answer to this question constitutes a "narrative" about personal history – that is, a summing-up of all the events a person has experienced. Yet, for people to tell us who they are, their elaborate stories must be projected onto simpler categories of self-description. A personal history is, necessarily, a *very* complex object. To convey it, an agent must project her richly variegated experiences onto a relatively few descriptors using the limited cognitive resources available.

In the model to be presented here, agents need to "talk" about their personal experiences before realizing potential gains from trade. How they elect to represent themselves to one another affects the productivity of their subsequent economic interactions. Because cognitive resources are limited, they make their representations in a simplified form. An agent's "identity" is the specific method she uses to implement such acts of selective self-representation. A group's "collective identity" is any self-representional mode adopted in common by (most of) the agents in that group. *So, for* *us, identity choice amounts to a decision about how to articulate a rich life history while using only the limited vocabulary available to a person for conveying who she might be. It is, in other words, the embrace of a way to make selective generalizations about personal experience.* Such generalizing acts unavoidably highlight and retain for future reference only that which is most salient. Our "categories" reflect the range of things an agent might take to be salient about herself. Our "narratives" are what results when a complex personal history is mapped onto the categories.[6] These categories, and the narratives to which they lead, are the building blocks of our theory of collective identity. This is potentially a powerful approach, we think, because people who embrace a common identity are predicted to recall their experiences in similar ways, to sort their historical data among the same bins, so to speak. This implication would appear to be testable by direct experimental methods.[7]

To illustrate, consider some hypothetical identity narratives: "I'm an immigrant who came up the hard way;" "I'm a child of the 1960's, and proud of it;" "I'm a working class white male angry at the world for not feeling my pain;" "I'm a tough-minded professional woman determined not to take a back seat to any less qualified man;" "I'm an intellectual who grew up in poverty, unlike those silver-spoon-fed intellectuals who love to talk about the poor but know nothing of them;" "I'm a black man who likes to have sex with other men, but I'm not a 'sissy,' and neither am I 'gay'." (Denizet-Lewis 2003). Each of these hypothetical people – in responding to the question, "Who am I?" – offers us a selective account. Having embraced certain categories of self-representation, they offer a "narrative" about personal experience using their chosen categories. These narratives are their ways of perceiving and representing themselves – their "modalities of self-awareness," if you will.

The key intuition that we strive to capture in our model is that identity choice is a social event, not merely the expression of individuals' values or preferences. In particular, people who interact frequently may end-up embracing similar categories of self-representation because they think this leaves them better placed to manage their collective action problems. When this is so, different contexts of social interaction can foster different equilibrium identity configurations, and agents

[6] All of this is very much in the spirit of Fryer and Jackson (2003). To reduce a person's full experience to a relatively few descriptors is akin to associating an object's "attributes" with the "prototypes" discussed by Fryer and Jackson.

[7] Some experimental work in this spirit has already been undertaken. Hoff and Pandey (2003) study the effects of caste identity on the cognitive performance of youngsters in an Indian village. Burns (2004) studies the impact of racial identity on trust in post-Apartheid South Africa.

interacting within relatively closed social networks may be inclined to embrace the same or similar identities. But, not all common categorical maps (collective identities) are created equal. Some may be superior to others, in terms of the quality of the interactions to which they give rise. In what follows we show how a "bad" (dysfunctional, self-destructive, victim-based, alienated, oppositional, anti-system) collective identity can be sustained in equilibrium for one group of people and not another, notwithstanding the fact that the "values" of people in the two groups are similar. And, we illustrate why it can be difficult to shift such a problematic pattern of personal identifications using only a marginal intervention: Beneficial tacit arrangements may have evolved among the agents, the viability of which turns on their embrace in common of the prevailing identity convention.

We will say that a *dysfunctional collective identity* has been affirmed when an alternative configuration of self-representations exists that would leave everyone better off, and yet no agent wants to embrace any alternative so long as the others with whom she routinely interacts are expected to adhere to the dysfunctional scheme.

In making the arguments to follow, we are inspired in a general way by the distinguished cultural anthropologist, Mary Douglas (2004). Her brilliant essay, "Traditional Culture – Let's Hear No More About It," includes (among many gems!) the following observation, which could readily serve as an epigraph for this paper:

> "Cultural solutions to coordination problems cost time and resources, and ... need to be grounded in regular personal interaction. Here are two partners who habitually work together, they rely on each other over their lifetimes and help each other in crises, often at personal cost. How does culture enable them to maintain their impressive solidarity? By organizing things so that the benefits pile up on the side of trust. This involves investing personal and political relations with value, such as family, or monarchy; it uses shame to put individuals under heavy obligations of reciprocity; it builds sanctions around the idea of honor and probity; it requires proofs of loyalty to kin, such as wildly ostentatious weddings and funerals to which all kinsfolk must be invited. It controls envy by redistributive institutions which disperse private accumulations and prevent great disparities of wealth. All of this reduces incentives, which is admittedly incompatible with development."

2 The Model

2.1 The Basic Set-up

The formal model we are about to study has three essential features: Agents can gain from trading with one another to an extent that depends on what they commonly know about the state of the world. Each agent has some private information about that state. And, by deciding *ex ante* what kind of "face to show to the world," agents determine what features of their private information become public. In this context we propose to study the emergence of dysfunctional collective identities.

More specifically, we consider a simple two-agent, two-stage game of identity choice and repeated risk sharing. Let the agents be indexed by $i = 1, 2$. In the first stage of play each agent makes a once-for-all choice of "identity." In every one of the infinite sequence of periods that constitutes the second stage, the agents receive random income endowments which they might agree to share with one another. We focus initially on what happens in the second stage. Let $y \in Y$ be an endowment realization. We assume that Y is a finite[8] set of non-negative real numbers representing the possible levels of receipt in each period of some perishable consumption good. Because endowments cannot be stored, the sum of agents' consumptions in any period cannot exceed the sum of that period's receipts. Moreover, agents will consume all net resources available to them in each period. To keep things simple, suppose that endowments are independent and identically distributed, both across agents and across periods. Let $p(y)$ be the probability that the endowment y is realized, so: $p(y) > 0$, and $\sum_{y \in Y} p(y) = 1$.

Thus, we have a dynamic game in two stages, with the second stage extending over an infinite sequence of periods. We assume that the agents play non-cooperatively, and that their first stage identity choices are common knowledge when they enter the second stage. The time line of the model is as follows: Before all interactions start, both agents choose their identities. After observing each other's choices in this regard, they engage in an infinitely repeated risk sharing interaction. We adopt subgame perfection as an equilibrium concept. When agents make a common choice in the first stage of an equilibrium path of play, we think of this as their *collective identity*.

Agents derive utility from consumption over the course of the second stage. They are risk

[8] We make Y finite here to ease the exposition of the general case. Nothing of consequence turns on this. Later in the paper, when we study the special case where $|X| = 2$, it is convenient to let Y be an interval of real numbers (permitting use of the calculus.)

averse, and their identical preferences are additively separable across periods. Indeed, we assume that they are expected discounted utility maximizers in the second stage, and that they discount the future at a common, constant rate, $\delta < 1$. We denote the utility function by $u : R_+ \rightarrow R$, and assume that $u(\cdot)$ is continuous, three-times differentiable, and satisfies (on the relevant range): $u' > 0$, and $u'' < 0$. (We shall see that the sign of u''' figures significantly in the analysis.) This is all we shall have to say about the agents' "tastes" or "values" in this paper on identity. Note that, in our formulation, agents do not derive utility from their "identities" as such.

Given these preferences, consumption fluctuations are undesirable. So, gains from trade are available to the agents if they can arrange to make interpersonal income transfers in an ongoing manner. This is their collective action problem. Because their second stage interactions are repeated, by making future dealings contingent on current behavior agents can exert leverage to enforce compliance with a variety of alternative transfer arrangements. We might even want to think of them as embracing some *custom* or *tradition* in regard to their risk sharing behavior. Whatever the interpretation, a *risk sharing arrangement* is defined to be any agreement obligating the agents to make and receive interpersonal transfers to and from one another in some specified manner. We will study some ways that agents' choices about identity affect their risk sharing prospects.

Before doing so, let us discuss how identity is to be represented in the model. Imagine that the endowment realizations are private information in each period, but that a set of "indicators" is available through use of which an agent can publicly signal her endowment. Let $x \in X$ denote a possible signal. The set of all available signals, X, is a finite collection of indicators, with $|X| << |Y|$. (That is, to capture our view that there are many fewer indicators than there are income states, we think of X as being a much smaller set than Y.) Moreover, while the $y \in Y$ are simply numbers – reflecting various levels of the endowment, the $x \in X$ can be more abstract objects – reflecting, for instance, various modes of self-presentation, alternative facial expressions, distinct demeanors or different verbal cues. To capture our position that it is practically infeasible for an agent to fully describe all aspects of her experience, we require that in every second stage period each agent makes a public "representation" about her income, $y \in Y$, by "announcing" an indicator, $x \in X$.

We stress that the making of these announcements is not, strictly speaking, a strategic act. We have in mind a situation where, once agents enter the second stage, the signals they emit about their endowments are given-off involuntarily, according to some formula or "code" that was adopted by the agent in the first stage of play. It is true that in our model the modes of self-presentation ultimately settled-upon by agents do, indeed, emerge from their strategic interactions in the first stage. But, when acting out these behavioral commitments in the second stage the agents ought not to be thought of as engaging in goal-oriented behavior. Rather than trying to take advantage of a risk sharing arrangement by giving a misleading report, we envision these agents encountering one another during the normal course of their social interactions and, in the context of such encounters, being unable to avoid bearing imperfect witness to their current period's endowment realization. An agent's "identity" in this world is simply her chosen modality for reacting in public to her private (income) experiences.

Accordingly, a function mapping the set of incomes *onto* the set of indicators, $C : Y \to X$, is to be called a *code*. In the first stage of play, agents simultaneously commit themselves to a code. That is, they adopt what might be called a "mode of self-presentation" which determines how they publicly react to their privately observed income realizations throughout the second stage. One can think of the agents as using these indicators to construct a "narrative" about their (income) experience. This is what "identity" means in our model. Their behaviors in this regard bind the agents to noisily signal their respective income realizations to one another in a particular manner.

Finally, and this is the key step in our analysis, we posit that any (implicit) income-sharing arrangement adopted by the agents in the second stage must be implemented solely in terms of these "income narratives." That is, consumption smoothing transfers between them can depend only on what is common knowledge between them – namely, their indicators, not their endowment realizations. So, resources move from the one with signal x to the one with signal $\tilde{x}$, but never from the one with income y to the one with income $\tilde{y}$. Intuitively – given the stationary, symmetric, i.i.d. environment that has been assumed – an ideal second stage risk sharing arrangement would move resources in each period from the higher-income (y_1, say) to the lower-income (y_2) agent, according to the formula: $t = \hat{T}(y_1, y_2)$, under which transfers come as close as possible to equalizing consumptions, subject to the constraint that the higher-income agent always has an incentive to make the transfer. However, this ideal arrangement is not feasible in our world because the endowment realizations (y_1, y_2) are not publicly observed. Instead, resource flows between agents must be a function of their announced indicators: $t = T(x_1, x_2)$. And, because the indicators are noisy signals of the incomes, a code-constrained income transfer agreement $T(\cdot, \cdot)$ can never perform as well as the full-information ideal, $\hat{T}(\cdot, \cdot)$.[9] How well the code-mediated sharing arrangements

[9] As we shall see, the ideal code-mediated arrangement moves resources in each period from the "higher-indicator" (x_1, say) to the "lower-indicator") (x_2) agent, according to the formula: $t = T(x_1, x_2)$, where transfers attempt to

actually do perform depends, in a manner to be investigated thoroughly in what follows, on the identity codes embraced by the agents at the first stage.[10]

We do not wish to allow our agents to adopt every conceivable code. In what follows, we assume that only the codes satisfying a property we call *monotonicity* can be considered. For reasons that will become clear, it is desirable that the signalling process preserve the natural order of the endowments. Yet, although the set of possible endowment realizations can be ordered in the natural way, there is in general no meaningful sense in which one indicator (e.g., a facial expression or tone of voice) is "larger" than another. Monotonicity is the requirement that the sets $\{C^{-1}(x) : x \in X\}$ respect the natural ordering on Y in the following sense:

Definition 1 *A code $C : Y \to X$ is **monotonic** if, for every $\{y, y', y''\} \subset Y : C(y) = C(y')$ and $\min\{y, y'\} < y'' < \max\{y, y'\}$ implies $C(y'') = C(y)$.*

So, if a code is monotonic then there is a way to assign numbers to indicators such that higher numbers invariably connote higher endowments.

Let C_i denote the first stage choice of a (monotonic)[11] code by agent i. We will refer to the pair (C_1, C_2) as a *code configuration.* As discussed, a *risk sharing arrangement* is a way to transfer resources between agents that depends on what they have to "say" to each other about their incomes, not the incomes themselves. And, such an arrangement is *feasible under a given code configuration* if it can be supported as a subgame perfect equilibrium continuation for the infinitely repeated interactions in the second stage. Given that codes are fixed once-and-for-all at the start of the second stage, that the maximal punishment available for a deviation from any proposed arrangement is (obviously) a reversion to autarky, and that random endowments are i.i.d. across agents and periods, no generality is lost by restricting attention to period-stationary risk sharing arrangements.[12] In light of the assumed discounting, if no one-shot deviation from a proposed

equalize conditional expected marginal utilities of consumption, subject to incentive constraints. [The notions of "higher" and "lower" indicators are sensible for "monotonic" codes, per the definition below.]

[10] One might think that cooperative risk sharing would be easier to sustain in an equilibrium continuation of the second stage, infinitely repeated interactions, if the agents have adopted the same codes in the first stage. This conjecture is basically correct, and in what follows it is verified in the context of our model.

[11] While non-monotonic codes are conceivable, it is intuitively obvious that, given the nature of the subsequent income sharing problem, they are informationally inefficient when compared to some alternative monotonic code that uses "the same" (in a sense that could be made precise) cognitive resources. So, practically speaking, we do not see this monotonicity condition as entailing any real loss of generality.

[12] That is, we consider only those arrangements where transfers depend, in the same manner each period, on that

arrangement is beneficial, taking the ensuing punishment into account, then neither can any finite or infinite sequence of deviations be beneficial. Now, let $t \in \Re$ denote a (possibly negative) transfer from agent one to agent two.[13] Reflecting the discussion to this point, we introduce the following formal definitions:

Definition 2 *A* ***risk sharing arrangement*** *is a period-stationary function,* $T : X^2 \rightarrow \Re$, *such that whenever the agents' signals are* (x_1, x_2), *the income transfer between agents is given by:* $t = T(x_1, x_2)$.

Definition 3 *A risk sharing arrangement* T *is* ***feasible under a given code configuration*** *if, for both agents* $i = 1, 2$, *in every second stage period and for all possible income realizations* $(y_1, y_2) \in Y \times Y$, *no net gain is anticipated for a one-shot deviation from the arrangement that is followed by a reversion to autarky.*

It is worth a moment's reflection at this point on what the model's primitives are supposed to be capturing about the contexts where identity choice occurs. There are four primitives here: the utility function, the discount factor, the set of available signals, and the distribution of random endowments. Using the linguistic conventions of economics, the first two reflect the agents' "tastes," and the last two their "opportunities." As may already be clear, what is most important about the utility function is its degree of risk aversion, and how this varies with the level of consumption. The more risk averse are the agents, the greater is their stake in second stage interactions. The rate of discount in repeated game models usually reflects factors like relationship stability and elapsed time between encounters. Here it is more natural to think of this parameter as capturing the density, or the degree of closure, of the social network mediating agents' second stage interactions. That is, $\delta \approx 1$ can be interpreted to mean that their encounters are quite frequent because their network is quite dense. Finally, the importance of identity choice varies inversely with the extent to which the available signals can serve as good proxies for the actual endowments. If the set X is quite "small" relative to the set Y, and if the endowments are very noisy, then "wrong" identity choices will have grave consequences. We illustrate the significance of these parameters later in the paper when we present a numerical comparative statics analysis of the following example.

period's indicators alone.

[13] When $t > 0$ we will speak of agent one "giving" and agent two "receiving" a transfer of size $|t|$, and conversely when $t < 0$.

2.2 An Example: The Case $|X| = 2$.

To illustrate these ideas, and for the sake of concreteness, we now introduce a simple example to which we shall have occasion to refer throughout this paper. This example posits that only two indicators are available: $X = \{B, G\}$. So, each second stage period involves the agents involuntarily signalling to one another, in effect, whether that period's endowment realization has been "good" or "bad." Subsequent transfers between the agents must be based on these binary signals.

This special case is already sufficiently rich to capture the key trade-off at work in our model. With $|X| = 2$, to choose a code, C, is necessarily to partition the endowment space into realizations with "good" and with "bad" signals: $Y = C^{-1}(B) \cup C^{-1}(G)$. Moreover, monotonic codes are always of the following threshold form: For some $y^* \in Y$, $C(y) = B$ if and only if $y \leq y^*$. To choose a code is thus to decide both about *the frequency of* and *the disparity between* good and bad endowment states. There are good reasons to think that the decentralized choices of self-interested agents in this regard will generally not be Pareto efficient. That is, there are good reasons to suppose that the identity configurations emergent in decentralized equilibrium will generally be dysfunctional.

To see the key trade-off at work here, the following two observations are useful: First, notice that the more widely disparate are the agents' endowment states associated with a given indicator pair, the more profitable are their risk sharing trades conditional on those signals. Secondly, observe that the more frequent are the encounters between unequally endowed agents, the greater are their opportunities to engage in profitable risk sharing. Hence, two traits of a code configuration – which we refer to as "mismatch frequency" and "endowment disparity" – are socially desirable. When $|X| = 2$, both traits are simultaneously determined by the choice for each agent of a dividing line between "good" and "bad" endowments, y_i^*. Therefore, in the neighborhood of an optimal choice, one of these desiderata is being traded-off against the other at the margin.

2.2.1 Three Endowment Realizations

To begin a more detailed discussion of the case $|X| = 2$, suppose further that only three endowment realizations are possible: $y \in Y = \{l, m, h\}$, $l < m < h$. In this circumstance, we will denote the endowment probabilities $p(y)$ by p_l, p_m and p_h respectively, where $\sum_{k \in \{l,m,h\}} p_k = 1$. A code is simply a map, $C : \{l, m, h\} \to \{G, B\}$. Under monotonicity, and without further loss of generality, we can restrict attention to the codes, C^P (for "pessimistic") and C^O (for "optimistic"), where:

$$\begin{aligned} C^P(l) &= B,\ C^P(m) = B,\ C^P(h) = G; \\ C^O(l) &= B,\ C^O(m) = G,\ C^O(h) = G. \end{aligned}$$

Thus, only three code configurations are possible in this two-person society: both are "pessimists" $\langle C^P, C^P \rangle$; both "optimists" $\langle C^O, C^O \rangle$; or the codes are mixed $\langle C^P, C^O \rangle$ or $\langle C^O, C^P \rangle$. In each second stage period the agents' incomes $y_i \in \{l, m, h\}$ are mapped to their signals $x_i \in \{B, G\}$ via one of the two codes, so:

$$x_i = C_i(y_i),\ \text{for } C_i \in \{C^P, C^O\},\ i \in \{1, 2\}.$$

Risk sharing transfers are then carried out in each period according to some period-stationary function of the announced indicators, $T(x_1, x_2)$.

Given this set-up, the analysis might proceed in two steps: For each code configuration, we would derive the agents' discounted sums of expected utility associated with some feasible transfer arrangement chosen by them in the second stage continuation. Then, we would study first stage code choice as equilibrium behavior in the symmetric, simultaneous move, 2×2 game where actions are the alternative codes $\{C^P, C^O\}$, and payoffs are the agents' respective welfare levels in the implied continuations. Obviously, in this example and in general, many feasible continuations are possible for each configuration since there exist many subgame perfect equilibria of the second stage's repeated interaction. So, to pursue this two-step program we would need to associate a *unique* second stage welfare level for the agents with each configuration, thereby specifying how the expected utility surplus (relative to autarky) generated by the prospect of risk sharing is to be divided among agents.[14] Once we have done this, the 2×2 first stage game would be well-defined.

Accordingly, throughout this paper *we posit that the agents adopt as a second stage continuation that feasible risk sharing arrangement which maximizes the sum of their expected discounted utilities.*[15]

[14] Sometimes we shall be interested only in the question of whether, for a given configuration, there exists any surplus whatsoever in the second stage, in which case the issue of surplus division does not arise.

[15] To be sure, other methods of surplus-splitting can be imagined – Nash bargaining, for instance. But our assumption here seems quite plausible. For, if both agents have chosen the same code, the utility possibility frontier for the second stage continuation is symmetric about the 45° line, in which case our selection method coincides with the Nash bargaining outcome. On the other hand, given the *ex ante* symmetry of this strategic situation, it makes sense to think that each agent is "equally likely" to end up on either side of a mixed configuration. So, rational agents

With this convention about surplus division in hand, we can then characterize first stage play with a reduced normal form game given by the following matrix:[16]

		Agent 2	
		C^P	C^O
Agent 1	C^P	V_P^*, V_P^*	V_M^{P*}, V_M^{O*}
	C^O	V_M^{O*}, V_M^{P*}	V_O^*, V_O^*

Our interpretation of this 3×2 example is as follows: the signals reflect either a "good" or a "bad" outcome, while the endowments can be either "high," "medium" or "low." So, given the requirement of monotonicity, an agent's choice of "identity" amounts to a choice about how to react to an intermediate income realization (whether to code it as a "good" or a "bad" event.) One way to talk about this is that, in effect, the agents must choose between being "pessimists" or "optimists." Alternatively, we could envision them as deciding whether, in the event of a middling endowment realization, to view themselves as a "victim" – that is, as someone who needs a helping hand but who is not in position to lend one.[17] Whatever the interpretation, we can ask whether the "optimistic" configuration $\left\langle C^O, C^O \right\rangle$ is better than the "pessimistic" one $\left\langle C^P, C^P \right\rangle$, in terms of the potential gains from second stage risk sharing that it engenders. And, we can inquire whether a mixed configuration – $\left\langle C^P, C^O \right\rangle$, say – is inferior to either "collective identity."[18]

Thus, in this example where only two choices of code are possible, we are able to discuss our ideas about dysfunctional collective identities using the basic notions of elementary game theory. If the normal form depicted above is a coordination game (i.e., if $V_P^* > V_M^{O*}$ and $V_O^* > V_M^{P*}$), then strategic forces favor the adoption of *some* collective identity and multiple, Pareto-ranked equilibria exist. Avoiding a dysfunctional identity then becomes a coordination problem for the agents.[19] Alternatively, if this game is a Prisoners' Dilemma (i.e., if $V_M^{P*} > V_O^* > V_P^* > V_M^{O*}$, for

viewing the surplus division problem from behind a 'veil of ignorance' well might agree to adopt the equilibrium selection method we have proposed.

[16] Here we are using the obvious notation: V_M^{O*} is the payoff to the optimistic agent under a mixed configuration, while V_P^* is either agent's payoff under a pessimistic configuration, etc.

[17] On this interpretation the example permits us to ask, in the habit if not in the spirit of McWhorter (2000), whether an expansive sense of one's victimization constitutes a "dysfunctional collective identity!"

[18] Stating this more provocatively, the example permits us to investigate whether the agents spread their joint income risks more effectively when they embrace a common "narrative of victimization!"

[19] As the literature on finitely repeated games makes clear (e.g., Benoit and Krishna 1985), in principle this coordination problem could be easily "solved." In our two-stage setup, given that autarky is always an equilibrium

instance, so that, although a pessimistic configuration is Pareto inferior to an optimistic one, it is nevertheless a dominant strategy for the agents to be pessimistic), then the two-stage strategic interaction has a "tragedy of the commons" quality about it, and the adoption by rational agents of a dysfunctional identity is all but guaranteed! (In section 4 below we use numerical analysis to further explore this case, exhibiting conditions on the primitives of the model under which dysfunctional collective identities are likely (or, bound) to emerge.)

2.2.2 A Continuum of Endowment Realizations

We can readily extend this 3×2 example. The assumption of three discrete income realizations, though allowing a colorful interpretation, is incidental to the analysis. When Y is an interval of real numbers and $|X| = 2$, the reduced-form game involves the agents simultaneously choosing thresholds (y_1^*, y_2^*) in the first stage, and reporting a "bad" outcome whenever their endowments are at or below the chosen thresholds.[20] This continuum specification is useful because, since the set of alternative thresholds is a bounded interval, and the agents' payoffs are differentiable functions of the threshold pair (assuming a well-behaved endowment distribution), we can use calculus to study the agents' strategic interaction in the first stage. (In section 3.4 below we explicitly solve this continuum example, adopting a quadratic utility function and letting the discount factor approach one.)

Now, suppose agent one has a lower threshold than agent two: $y_1^* < y_2^*$. Furthermore, let

continuation at the second stage, coordination on the efficient equilibrium in the reduced normal form could be enforced by threatening the autarkic risk sharing continuation if either agent embraces the "wrong" identity. Exploiting this insight, one might argue that a dysfunctional identity ought not to emerge in cases where the reduced normal form is a coordination game if the agents use all of the strategic resources available to them.

We do not find this argument convincing. The prospect of renegotiation seriously undermines the credibility of any such threat. ("If you turn out to be the 'wrong' kind of person, then I won't have anything to do with you" is a threat lacking credibility in most social networks!) And while the same claim could be made about reversion to autarky as a threat supporting any risk sharing *within* the second stage, we think that the renegotiation of a risk sharing arrangement after a deviation on identity choices has been observed, but before any risk sharing has actually taken place, is a much easier thing to envision than the renegotiation of such an agreement after its own terms have just been violated. This admittedly informal reasoning nevertheless suggests that we might plausibly impose "renegotiation-proofness" between *stages,* but not between *periods* within the second stage, which would leave the agents still facing the coordination difficulty that we are discussing here.

[20] It is natural here, in keeping with the intuition from the 3×2 case, to associate a higher threshold y_i with a "more pessimistic" identity choice by agent i (or, with the agent adopting a "more expansive sense of her victimization"), since a higher threshold makes it less likely that a "good" signal is announced.

$\pi_i = \int_{\{y \in Y,\ y \leq y_i^*\}} p(y)dy$ be the probability that agent i announces "B." So, $\pi_1 < \pi_2$.[21] Then, since the utility function is strictly concave, the endowment disparity between the agents conditional on the event $E = \{y_1 \leq y_1^*\} \cap \{y_2 > y_2^*\}$ permits a transfer (from agent two to agent one) with relatively low utility cost to the giver and high utility benefit for the receiver. The more widely disparate are y_1^* and y_2^*, the greater is the social surplus from such a transfer. However, the mismatch frequency for this event is $\Pr\{E\} = \pi_1(1 - \pi_2)$. Thus (as mentioned above) encounters of this kind, though more profitable, occur less often as y_1^* and y_2^* become more widely disparate (since π_1 falls and/or π_2 rises.) Moreover, as y_1^* and y_2^* grow further apart, trading opportunities deteriorate in the other "mismatch event," $E' = \{y_1 > y_1^*\} \cap \{y_2 \leq y_2^*\}$ (because the endowment ranges conditional on this event overlap more.)[22] Thus, at the socially optimal code configuration in this continuum case, the disparity between y_1^* and y_2^* will be such that the benefit of more profitable transfers conditional on E is just balanced by the cost of less profitable transfers conditional on E', plus the cost that E occurs less frequently.[23]

This continuum example can also be used to illustrate why inefficient collective identity choices are to be expected: *The private evaluation of benefits and costs associated with alternative code configurations is likely to differ from this social assessment.* Two countervailing factors can cause private and social valuations to differ in our model:

(i) When contemplating the choice of a higher threshold in the first stage of play, an individual (agent one, say), takes into account that the second stage transfer policy will become marginally less attractive for her (because raising her threshold makes her endowment distribution more favorable conditional on either signal, thereby lowering the transfer she receives, or raising the transfer she gives, at every indicator pair.) *But this private cost to agent one is not a social cost.* Invoking the Envelope Theorem, we know that in the neighborhood of an optimal configuration the net social impact of an induced shift in the transfer arrangement is zero. So, due to this pecuniary externality, agent one may tend to set y_1^* *below* its socially optimal level.

(ii) On the other hand, since agent one's likelihood of giving a transfer declines as y_1^* rises,

[21] Hereafter, if Y is an interval of real numbers we take $p : Y \to \Re_+$ to be a probability density function, with

$$\int_Y p(y)dy = 1 \text{ and } \int_Y yp(y)dy < \infty.$$

[22] Notice that $\Pr\{E'\} = \pi_2(1 - \pi_1) > \Pr\{E\}$, since $\pi_2 > \pi_1$.

[23] Of course, if the agents embrace a collective identity then they share a common threshold, and the disparity between y_1^* and y_2^* is zero.

raising her threshold has a negative effect on her trading partner.[24] *But this social cost is not a private cost to agent one.* When choosing their thresholds, each agent ignores this impact on the other agent. So, due to this external diseconomy, agent one may tend to set y_1^* *above* its socially optimal level.

In general, how the equilibrium and the socially optimal configurations compare depends on the relative magnitude of these two wedges between private and social valuation. In particular, the symmetric equilibrium threshold will exceed the socially optimal level if, when considering a marginal increase in y_1^*, the external diseconomy on agent two due to agent one's lowered frequency of giving a transfer [specified in (ii) above] exceeds the pecuniary externality on agent one due to the induced decline in her net transfer receipts [specified in (i).] But, using the Envelope Theorem again, any induced negative impact on agent one is just offset by an induced positive impact on her trading partner. We conclude that the equilibrium threshold exceeds the socially optimal one if the direct plus the induced impact on agent two of a marginal increase in agent one's threshold is negative.

We can make this point somewhat more formally, while introducing some notation specific to this example that will prove useful later. Thus, with $X = \{B, G\}$ and Y an interval on the non-negative real line, denote by $U(y_1, y_2)$ player one's payoff at the threshold pair, (y_1, y_2). Let $W(y) = U(y, y)$; let U_i be the partial derivatives of U with respect to y_i, $i = 1, 2$; let $y^e = y_1^* = y_2^*$ be the agents' common threshold in a symmetric equilibrium, and let y^o be the socially optimal (i.e., the sum-of-discounted–utility-maximizing) common threshold. Then, we have the first-order conditions: $U_1(y^e, y^e) = 0$, and $W'(y^o) = U_1(y^o, y^o) + U_2(y^o, y^o) = 0$. It follows that $U_2(y^e, y^e) \lesseqqgtr 0$ implies $W'(y^e) \lesseqqgtr 0$ which, in turn, implies $y^e \gtreqqless y^o$ (assuming the relevant second-order condition.)

We conclude (in the context of this extended example) that the *symmetric equilibrium identity configuration is a spoiled collective identity involving too much pessimism (too much optimism) whenever the net effect of raising one agent's threshold marginally from its equilibrium level is to reduce (increase) the payoff of the other agent*! Thus, the case $|X| = 2$ affords us a tractable context within which to demonstrate the kinship of the "identity coordination problem" being posed here with the classical "tragedy of the commons."

[24] With $y_1^* < y_2^*$, the likelihood of agent one (resp., agent 2) giving a transfer is $\pi_2(1 - \pi_1)$ [resp., $1 - \pi_2(1 - \pi_1)$].

3 Analysis

3.1 Notation and Preliminaries

We return now to a discussion of the general model. Imagine that the agents enter the second stage having adopted the (monotonic) code configuration $\langle C_1, C_2 \rangle$. We begin by describing the feasible risk sharing arrangements available to these agents, and the expected discounted utility surpluses (relative to autarky) that accrue to them from adopting any particular arrangement. Ultimately, we will provide (in Theorem 1) an explicit characterization of the discount factors for which a feasible arrangement can be found that generates positive surplus for both agents. Toward this end, we require some more notation.

For $y \in Y$ an endowment level and $t \in \Re$ a (possibly negative) net transfer, denote the utility change for someone with endowment y who receives a net transfer of t by:

$$\Delta u(y,t) \equiv u(y+t) - u(y).$$

Given the distribution of endowments, any code choice induces a distribution of indicators. For $x \in X$, let $q_i(x)$ denote the probability that agent i announces indicator x under code C_i. Then:

$$q_i(x) \equiv \sum_{y \in C_i^{-1}(x)} p(y),\ \ i = 1,2.$$

Moreover, for $x \in X$ and $t \in \Re$, consider the conditional expected utility gain over autarky for agent i, given that her indicator realization is x and that her net transfer is to be t. We denote this conditional expected payoff by $v_i(x,t)$, where:

$$v_i(x,t) \equiv E\left[\triangle u(y,t) \middle|\, y \in C_i^{-1}(x)\right] = \sum_{y \in C_i^{-1}(x)} \frac{p(y) \triangle u(y,t)}{q_i(x)}, i = 1,2. \tag{1}$$

Analogously, we write $v_i'(x,t)$ to represent the conditional expected *marginal* utility (i.e., the "conditional shadow price") for agent i at indicator x given transfer t. So:

$$v_i'(x,t) \equiv E\left[u'(y+t) \middle|\, y \in C_i^{-1}(x)\right] = \frac{\partial v_i}{\partial t}(x,t),\ \ i = 1,2. \tag{2}$$

In what follows we shall be particularly interested in the conditional shadow prices when transfers are zero, $v_i'(x,0)$.

Now, it is obvious that $\{C_i^{-1}(x) : x \in X\}$ is a finite, pairwise-disjoint family of sets that covers Y – i.e., a partition of Y. It is also obvious that when the code C_i is monotonic, these sets are "intervals," in the sense that if $x \neq x'$, then either $\min\{C_i^{-1}(x)\} > \max\{C_i^{-1}(x')\}$ or $\max\{C_i^{-1}(x)\} < \min\{C_i^{-1}(x')\}$.[25] Given the strict concavity of $u(\cdot)$, we conclude that, for every $x \neq x' \in X$, either:

$$\begin{aligned} v_i'(x,t) &> v_i'(x',t) \text{ for all } t \in \Re, \text{ or} \\ v_i'(x,t) &< v_i'(x',t) \text{ for all } t \in \Re. \end{aligned}$$

That is, a monotonic code always induces a complete, strict ordering of the marginal valuation schedules, conditional on the elements of X.

Intuitively, given an indicator pair (x_1, x_2) and a level of transfer t, the ratio $\frac{v_1'(x_1,-t)}{v_2'(x_2,t)}$ is the "conditional marginal rate of (utility) substitution" between the agents via adjustments to the transfer at the indicator pair (x_1, x_2). In particular, efficient risk sharing should entail an effort to equalize these substitution rates, moving resources from agent one to agent two ($t > 0$) when $\frac{v_1'(x_1,0)}{v_2'(x_2,0)}$ is "low", and from agent two to agent one ($t < 0$) when it is "high." Accordingly, we use $(\tilde{x}_1, \tilde{x}_2)$ and $(\hat{x}_1, \hat{x}_2)$ to denote the special indicator pairs at which the zero-transfer conditional substitution rate takes its largest and its smallest values:

$$\frac{v_1'(\tilde{x}_1, 0)}{v_2'(\tilde{x}_2, 0)} \leq \frac{v_1'(x_1, 0)}{v_2'(x_2, 0)} \leq \frac{v_1'(\hat{x}_1, 0)}{v_2'(\hat{x}_2, 0)}, \text{ for all } x_1 \in X,\ x_2 \in X. \tag{3}$$

Evidently, $v_1'(x, 0)$ is minimized over X at $\tilde{x}_1$ and maximized at $\hat{x}_1$; while, $v_2'(x, 0)$ is maximized at $\tilde{x}_2$ and minimized at $\hat{x}_2$. So, again thinking at an intuitive level, the joint indicator realization $(\tilde{x}_1, \tilde{x}_2)$ is the event commonly known to the agents that is most favorable for a marginal transfer of resources from agent one to agent two (in the sense that the cost to agent one from the transfer is least and the gain from it for agent two is greatest at this event.) Likewise, the realization $(\hat{x}_1, \hat{x}_2)$ is the event that favors most a marginal transfer of resources from agent two to agent one.

To see how these notions will prove useful, imagine that initially no transfers are taking place and consider the problem of determining whether there is a *marginal transfer arrangement* (i.e., one "near" zero) which leaves both agents better off than under autarky. Clearly, a transfer from agent one at $(\tilde{x}_1, \tilde{x}_2)$ that is offset with a transfer in the other direction at $(\hat{x}_1, \hat{x}_2)$ gives the agents their best chance to achieve a Pareto improvement via a marginal arrangement. This is so for two reasons: the size of transfer needed to produce a given increase in the recipient's welfare is least at these realizations; and, the loss in welfare due to making a transfer of given size is also least at

[25] So, the choice of a monotonic code amounts to deciding upon a way to partition the range of incomes into "connected" subsets of Y, with there being as many cells in the partition as there are elements of X.

these realization. Of course, making both transfers would need to be consistent with incentives if the agents are to achieve a Pareto improvement in this way. But, because the least well-endowed (i.e., most incentive-constrained) giver's cost of a marginal transfer is least at these indicator pairs, if the incentive conditions cannot be satisfied at these realizations then they cannot be satisfied elsewhere. We summarize the discussion to this point in the following Lemma:

Lemma 1 *There is a strict Pareto improving, feasible marginal transfer arrangement only if an arrangement of this kind exists which also satisfies:* $T(\tilde{x}_1, \tilde{x}_2) > 0, T(\hat{x}_1, \hat{x}_2) < 0,$ *and* $T(x_1, x_2) = 0$ *otherwise.*

3.2 Value Functions and Incentive Constraints

Let $V_i(T)$ denote the expected discounted utility surplus (relative to autarky) over the course of the second stage enjoyed by agent i under arrangement T. Given an arrangement, when the announced indicators are (x_1, x_2) agent one consumes $y_1 - T(x_1, x_2)$ and agent two consumes $y_2 + T(x_1, x_2)$. So, exploiting the stationarity and using (1), we can write:

$$V_1(T) = (1-\delta)^{-1} \sum_{x_1 \in X} \sum_{x_2 \in X} q_1(x_1) q_2(x_2) v_1\left(x_1, -T(x_1, x_2)\right), \tag{4}$$

and

$$V_2(T) = (1-\delta)^{-1} \sum_{x_1 \in X} \sum_{x_2 \in X} q_1(x_1) q_2(x_2) v_2\left(x_2, T(x_1, x_2)\right). \tag{5}$$

In view of our assumption of risk aversion, it is obvious that the $V_i(\cdot)$ are concave functions of the elements of T.[26]

A transfer arrangement is feasible under a given code configuration if neither agent ever expects to gain by a one-shot deviation from it that is followed by reversion to autarky. So, if under a feasible arrangement T agent i has endowment y and is required to make a transfer to the other agent of magnitude $|t|$, then it must be that:

$$u(y) - u(y - |t|) \leq \delta V_i(T).$$

Hence, at every indicator pairs (x_1, x_2) the incentive requirements for feasibility are as follows:

$$\begin{aligned} \text{Agent one} \quad &: \quad -\Delta u\left(y, -T(x_1, x_2)\right) \leq \delta V_1(T), \text{ for all } y \in C_1^{-1}(x_1); \\ \text{Agent two} \quad &: \quad -\Delta u\left(y, T(x_1, x_2)\right) \leq \delta V_2(T), \text{ for all } y \in C_2^{-1}(x_2). \end{aligned}$$

[26] We stress that the functions $q_i(x_i)$, $v_i(x_i, t)$ and $V_i(T)$ very much depend on the code C_i, and that our notation suppresses that dependence.

Since $\Delta u(y,t) \gtreqless 0$ as $t \gtreqless 0$, agent two's inequality above holds trivially when $T(x_1,x_2) > 0$, as does agent one's when $T(x_1,x_2) < 0$. Moreover, since $u(\cdot)$ is a strictly concave function it is clear that if agent i is asked to make a transfer when her indicator is x_i, then the most favorable relevant circumstance for a profitable deviation occurs when $y_i = \min\{C_i^{-1}(x_i)\}$. So, we may write the $2\left|X\right|^2$ incentive conditions (for each agent $i = 1,2$ and at each indicator pair $(x_1,x_2) \in X^2$) as follows:

$$\delta V_1(T) + \Delta u\left(\min\{C_1^{-1}(x_1)\}, -T(x_1,x_2)\right) \geq 0; \tag{6}$$

and, likewise:

$$\delta V_2(T) + \Delta u\left(\min\{C_2^{-1}(x_2)\}, T(x_1,x_2)\right) \geq 0. \tag{7}$$

Given $\langle C_1, C_2\rangle$ and δ, and for the value functions as specified in equations (4) and (5), let $\Im(C_1,C_2;\delta)$ be the set of feasible transfer arrangements under code configuration $\langle C_1, C_2\rangle$ and discount factor δ:

$$\Im(C_1,C_2;\delta) \equiv \left\{T : X^2 \to \Re \middle| T \text{ satisfies (6) and (7)}\right\}.$$

Notice that the LHS of inequalities (6) and (7) are concave functions of the $|X|^2$ real numbers, $\{T(x_1,x_2)\}$. Therefore, the set of feasible transfer, $\Im(C_1,C_2;\delta)$, arrangements is convex. Hence, the attainable payoffs for the two players in the second stage subgame – given a code configuration and discount factor – form a convex subset of $\Re^2$. Since autarky (no transfers and zero surplus for both players) is surely feasible, we have the following result:

Lemma 2 *A feasible transfer arrangement generating a positive surplus for both agents exists if and only if a strictly Pareto improving marginal transfer arrangement exists.*

3.3 Gains from Trade in the Second Stage

3.3.1 A General Result

Given a code configuration $\langle C_1, C_2\rangle$, we are interested in determining the range of discount factors over which it is possible for both agents to realize a positive surplus from risk sharing in the second stage (relative to autarky). It is clear that for δ small enough any non-zero risk sharing arrangement is infeasible. Moreover, any arrangement that would yield a positive surplus for both agents becomes feasible when δ is close enough to one. Intuition therefore suggests that for every configuration $\langle C_1, C_2\rangle$ there is a cut-off level for the discount factor, $\bar{\delta}(C_1,C_2)$, such that no gains from second stage trade are possible when $\delta \leq \bar{\delta}(C_1,C_2)$. This is, indeed, the case. Theorem 1 establishes this fact and provides an explicit characterization of $\bar{\delta}(\cdot,\cdot)$. The significance of this result is that it gives us a way to assess the economic efficacy at relatively low discount factors of alternative identity configurations: The lower is $\bar{\delta}(C_1, C_2)$, the wider is the range of environments under which a positive surplus can be realized, and so (in this specific sense) the greater is the scope for risk sharing afforded by the configuration.

To state and prove the theorem we (unfortunately) need one more bit of notation. For agent i with code C_i and signal $x_i \in X$, consider the most that it could cost that agent to surrender a marginal unit of consumption starting from a situation of zero transfers, and denote this number by ϕ_i. That is:

$$\phi_i(x_i) \equiv u'\left(\min\{C_i^{-1}(x_i)\}\right),\ i = 1, 2.$$

So, $\phi_i(x_i)$ is the marginal utility of the agent i who has received the lowest endowment level consistent with announcing indicator x_i. (Obviously, $\phi_i(x_i) > v_i'(x_i, 0)$.) We can now state our result.

Theorem 1 *Given code configuration $\langle C_1, C_2\rangle$ and discount factor δ, there exists a transfer arrangement $T \in \Im(C_1, C_2; \delta)$ for which $V_i(T) > 0$, $i = 1, 2$, if and only if $\delta > \bar{\delta}(C_1, C_2)$, where $\bar{\delta}(C_1, C_2)$ is the unique solution in the unit interval of :*

$$\frac{v_2'(\tilde{x}_2,0)}{v_1'(\tilde{x}_1,0)}\cdot\left[1+\frac{(\frac{1-\delta}{\delta})\phi_1(\tilde{x}_1)}{q_1(\tilde{x}_1)q_2(\tilde{x}_2)v_1'(\tilde{x}_1,0)}\right]^{-1}=\frac{v_2'(\hat{x}_2,0)}{v_1'(\hat{x}_1,0)}\cdot\left[1+\frac{(\frac{1-\delta}{\delta})\phi_2(\hat{x}_2)}{q_1(\hat{x}_1)q_2(\hat{x}_2)v_2'(\hat{x}_2,0)}\right], \qquad (8)$$

and where $(\tilde{x}_1, \tilde{x}_2)$ and $(\hat{x}_1, \hat{x}_2)$ are the special indicator pairs defined above in (3), at which the per-period utility substitution rate $\frac{v_2'(x_2,0)}{v_1'(x_1,0)}$ takes, respectively, its highest and lowest values on X^2.

Theorem 1 has an intuitive interpretation: It can be understood to say that there are no gains from trade if, with zero transfers taking place, agent 2's marginal welfare cost per unit of agent 1's welfare benefit at $(\hat{x}_1, \hat{x}_2)$ (where it is most favorable for agent 2 to give) exceeds her marginal benefit per unit of agent $1'$s cost at $(\tilde{x}_1, \tilde{x}_2)$ (where it is most favorable for agent 1 to give). To see this, notice that the LHS of equation (8) gives the marginal rate of welfare substitution between these agents at zero transfers $\left(\frac{dV_2(T)}{dV_1(T)}\ |_{T=0}\right)$, conditional on the indicator pair $(\tilde{x}_1, \tilde{x}_2)$ and viewed from the perspective of the agent obliged to make a transfer there, for whom the incentive constraint binds. The RHS gives the same welfare substitution rate conditional on $(\hat{x}_1, \hat{x}_2)$. Rates of expected discounted welfare substitution [either side in (8) above] differ by the indicated multiplicative factors from per-period utility substitution rates $\left(\frac{v_2'(x_2,0)}{v_1'(x_1,0)}\text{ above}\right)$. This is because the incentive constrained giver [the least well endowed agent 1 at $(\tilde{x}_1, \tilde{x}_2)$ and the least well endowed agent 2 at $(\hat{x}_1, \hat{x}_2)$] incurs a cost with certainty in the current period, but only with probability $q_1(x_1)q_2(x_2)$ in each subsequent period. So, the term $\frac{(\frac{1-\delta}{\delta})\phi_i(x_i)}{q_1(x_1)q_2(x_2)v_i'(x_i,0)}$ represents the current marginal compliance cost per unit of expected discounted future cost, when agent i is the incentive constrained giver.[27] The larger is δ, the smaller is this term. So, the RHS in (8) is strictly decreasing in δ and the LHS is strictly increasing. Moreover, by the definition of $(\tilde{x}_1, \tilde{x}_2)$ and $(\hat{x}_1, \hat{x}_2)$, the RHS is less (greater) than the LHS for δ near one (for δ near zero.) So, equation (8) has a unique solution in the unit interval, as asserted.

Proof. To prove Theorem 1 notice first that, in view of Lemmas 1 and 2, positive surplus for both agents is feasible if and only if there is a feasible marginal risk sharing arrangement that moves resources from agent one to agent two only at $(\tilde{x}_1, \tilde{x}_2)$, and from agent two to agent one only at $(\hat{x}_1, \hat{x}_2)$, such that the outcome of these transfers strictly Pareto dominates autarky. In turn, it is clear, from (6) and (7), that *any* feasible, non-zero transfer arrangement must be a strict Pareto improvement over autarky. Therefore, a positive surplus is possible at the second stage if and only if there exists a pair of positive numbers (t_1, t_2) (representing the magnitudes of the transfers in a marginal arrangement) such that $\hat{T} \in \Im(C_1, C_2; \delta)$, for $\hat{T}$ satisfying:

$$\hat{T}(\tilde{x}_1, \tilde{x}_2) \equiv t_1 > 0, \hat{T}(\hat{x}_1, \hat{x}_2) \equiv -t_2 < 0, \text{ and } \hat{T}(x_1, x_2) = 0 \text{ otherwise.}$$

We show that such numbers exist if and only if $\delta > \bar{\delta}(C_1, C_2)$ as defined above.

Consider the inequalities (6) and (7). Let us take (t_1, t_2) to be a pair of positive numbers in the neighborhood of $(0, 0)$. For $\hat{T}$ the marginal transfer arrangement specified in terms of (t_1, t_2) above, define:

$$F_1(t_1, t_2) \equiv \delta V_1(\hat{T}) + \Delta u \left(\min\{C_1^{-1}(\tilde{x}_1)\}, -t_1\right),$$

and

$$F_2(t_1, t_2) \equiv \delta V_2(\hat{T}) + \Delta u \left(\min\{C_2^{-1}(\hat{x}_2)\}, t_2\right).$$

Obviously, there is a feasible marginal transfer arrangement if and only if there is a pair of positive numbers (t_1, t_2) near zero for which $F_i(t_1, t_2) > 0$, $i = 1, 2$. Totally differentiating the functions

[27] This interpretation may be verified by observing that, from the point of view of the least well off giving agent at $(\tilde{x}_1, \tilde{x}_2)$ (say), the expected marginal cost of a transfer is $\phi_1(\tilde{x}_1)$ in the current period, plus $q_1(\tilde{x}_1)q_2(\tilde{x}_2)v_1'(\tilde{x}_1, 0)$ starting in the next period and continuing in perpetuity. Moreover, the importance of a cost incurred in a single current period, relative to a cost incurred in perpetuity starting next period, is $(\frac{1-\delta}{\delta})$.

$F_i(\cdot,\cdot)$, while bearing in mind that $\partial F_i/\partial t_i < 0$ and $\partial F_i/\partial t_j > 0$, $i \neq j$, reveals that the inequalities:

$$[\partial F_1/\partial t_1]dt_1 + [\partial F_1/\partial t_2]dt_2 > 0 \text{ and } [\partial F_2/\partial t_1]dt_1 + \ [\partial F_2/\partial t_2]dt_2 > 0$$

can both hold simultaneously only if:

$$-\frac{\partial F_1/\partial t_1}{\partial F_1/\partial t_2} < \frac{dt_2}{dt_1} < -\frac{\partial F_2/\partial t_1}{\partial F_2/\partial t_2},$$

where the derivatives above are evaluated at $(t_1, t_2) = (0, 0)$. Carrying out the indicated differentiation, one can see that the inequality $-\frac{\partial F_1/\partial t_1}{\partial F_1/\partial t_2} < -\frac{\partial F_2/\partial t_1}{\partial F_2/\partial t_2}$ holds if and only if the LHS exceeds the RHS in equation (8). However, as noted, the LHS increases and the RHS decreases with δ. So, a feasible marginal transfer arrangement can be found which strictly Pareto dominates autarky if and only if $\delta > \bar{\delta}(C_1, C_2)$, as was to be shown. ■

3.3.2 Gains from Trade with Collective Identities

We now employ Theorem 1 to investigate the scope for risk sharing enjoyed by the agents when they embrace a common code (i.e., a collective identity.) Thus, suppose $C_1 = C_2 = C$, and denote by $\bar{\delta}_C \equiv \bar{\delta}(C, C)$ the minimal discount factor consistent with there being positive gains from trade when both agents embrace the same code. Notice that the functions on X defining the distributions of indicators, $q_i(\cdot)$, and the shadow prices, $v_i'(\cdot)$ and $\phi_i(\cdot)$, are now the same for both agents. So, we can drop the subscript i in what follows. Moreover the critical indicator pairs, $(\tilde{x}_1, \tilde{x}_2)$ and $(\hat{x}_1, \hat{x}_2)$ as defined above, will in this case be such that:

$$\begin{aligned} (\tilde{x}_1, \tilde{x}_2) &= (\hat{x}_2, \hat{x}_1) \equiv (x^H, x^L), \\ \text{where } x^H &\equiv \arg\min_{x \in X}\{v'(x, 0)\} \text{ and } x^L \equiv \arg\max_{x \in X}\{v'(x, 0)\}. \end{aligned}$$

Here, given a collective identity C, we think of x^H as the "high income" indicator (i.e., the one with the lowest conditional expected marginal utility), and x^L is the "low income" indicator (i.e., the one with highest conditional shadow price,) and these are the same for both agents. Define $\phi_C \equiv u'(\min\{C^{-1}(x^H)\})$. Then, straightforward manipulation of the formula in equation (8) reveals the following, which we state without proof:

Corollary 1 *If the agents have embraced a collective identity, so that $C_1 = C_2 = C$, then positive gains from trade can be achieved in equilibrium if and only if $\delta > \bar{\delta}_C$, where $\bar{\delta}_C$ is given by:*

$$\frac{1 - \bar{\delta}_C}{\bar{\delta}_C} = q(x^H)q(x^L)\left[\frac{v'(x^L, 0) - v'(x^H, 0)}{\phi_C}\right]. \tag{9}$$

Thus, under a collective identity the agents' scope for effective risk sharing depends on three factors: the gap $\left[v'(x^L,0) - v'(x^H,0)\right]$ in conditional shadow prices between their "worst" and "best" indicator states – this is the social benefit from a marginal transfer; the likelihood $[q(x^H)q(x^L)]$ of an encounter between them when one is in the "worst" and the other the "best" state; and the cost of a marginal transfer to the least well-off person announcing the "best" signal $[\phi_C]$. Equation (9) shows that these various factors combine in an intuitively appealing way to determine whether gains from trade can be attained with the collective identity, C: The term $\frac{v'(x^L,0)-v'(x^H,0)}{\phi_C}$ gives the ratio of benefits to costs from the best marginal trade, at (x^H, x^L), as viewed by the agent whose incentive constraint binds there. On the RHS of (9) this ratio is multiplied by the probability of that particular trading opportunity arising in a later period. Costs are incurred currently, while benefits accrue in perpetuity beginning in the next period. So, $\frac{\delta}{1-\delta}\cdot$[expected marginal benefit] $>$ [marginal cost] is a necessary condition for a marginal transfer to look profitable to the one who makes it. Equation (9) makes clear that this condition must fail, even for the best marginal transfer, if $\delta \leq \bar{\delta}_C$.

3.3.3 "Optimism" vs. "Pessimism" in the 3×2 Case

Recall now the 3×2 example introduced in section 2.2.1 above, where identity choice amounts to a decision on whether to code an intermediate endowment state as a "good" (optimist) or a "bad" (pessimist) event. Let $\bar{\delta}_P$ (resp. $\bar{\delta}_O$) denote the critical discount factors below which no surplus is possible, given that a pessimistic (optimistic) collective identity has been adopted by the agents. Then, using equation (9), we conclude that:

$$\begin{aligned}\frac{1-\bar{\delta}_P}{\bar{\delta}_P} &= p_h(1-p_h)\cdot\left[\frac{\alpha u'(l) + (1-\alpha)u'(m) - u'(h)}{u'(h)}\right] \text{ and} \\ \frac{1-\bar{\delta}_O}{\bar{\delta}_O} &= p_l(1-p_l)\cdot\left[\frac{u'(l) - \beta u'(m) - (1-\beta)u'(h)}{u'(m)}\right],\end{aligned}$$

where $\alpha \equiv \frac{p_l}{1-p_h}$ and $\beta \equiv \frac{p_m}{1-p_l}$. In light of the foregoing discussion, it is obvious that optimism affords a wider scope for risk sharing than does pessimism when $\frac{p_l(1-p_l)}{p_h(1-p_h)}$ is large, and/or when $\left|\frac{m-l}{h-m}\right|$ is large.

This result is intuitively satisfying. Pessimism conflates low and intermediate endowment states, while optimism conflates intermediate and high states. So, collective optimism will dominate collective pessimism when the information constraint of lumping together high and medium income states is less debilitating to the risk sharing enterprise than is the constraint of lumping together medium and low states. As an extreme example, as $h - m$ goes to zero, optimism will surely dominate because lumping the h and m endowments entails essentially zero information loss. Just the opposite is the case as $m - l$ goes to zero because lumping the m and l endowments then entails a trivial information loss. Likewise, as p_l goes to zero, pessimism will dominate optimism as a collective identity, and the opposite will be the case as p_h goes to zero. Moreover, as p_m goes to zero, the agents have full endowment information under both pessimism and optimism, so the two collective identities must be equivalent in that case.

Finally, to conclude our discussion of the 3 × 2 case, suppose the parameters of that example satisfy:

$$p_h = p_l \equiv p; \; p_m = 1 - 2p; \text{ and } h - m = m - l \equiv g.$$

We refer to this circumstance, in the context of the 3 × 2 example, as a *symmetric endowment distribution.* Since asymmetric distributions naturally favor optimism if left-skewed, or pessimism if right-skewed, examining this symmetric case provides a useful benchmark. Under symmetric endowment distributions high and low endowments are equally likely, and the intermediate endowment lies midway between the high and low realizations. Reasoning intuitively, if the endowment distribution is symmetric and if the demand for consumption insurance falls as the level of consumption rises, then a noisy signal at the lower range of endowments should be more of an impediment to welfare-enhancing risk sharing than a noisy signal at the higher range of endowments. So, optimism should dominate pessimism as a collective identity when the endowment distribution is symmetric, if the agent is less risk averse at higher levels of consumption. This is indeed the case, as the following result demonstrates.

Denote by $V_O(t)$ and $V_P(t)$ the agents' common level of welfare in the 3 × 2 example, under collective "optimism" and collective "pessimism" respectively, given that the transfer arrangement satisfies: $T(G, B) = t \geq 0$. (It is obvious that if both agents adopt the same identity code, then risk sharing transfers between them will take place only if they announce different indicators.) Then, using equations (4) and (5), a straightforward calculation reveals that:

$$\theta \cdot [V_O(t) - V_P(t)] = \begin{array}{c} \{[u(l+g) - u(l)] - [u(l+g+t) - u(l+t)]\} \\ - \{[u(l+2g-t) - u(l+g-t)] - [u(l+2g) - u(l+g)]\}, \end{array}$$

where $\theta \equiv \frac{1-\delta}{p(1-2p)}$. Using the Fundamental Theorem of Calculus, the RHS above can be written as follows:

$$
\begin{aligned}
RHS &= \int_0^g \left[u'\left(l+x\right) - u'\left(l+t+x\right)\right] dx - \int_0^g \left[u'\left(l+g-t+x\right) - u'\left(l+g+x\right)\right] dx \\
&= \int_0^g \int_o^t u''\left(l+g+x+w\right) dwdx - \int_0^g \int_0^t u''\left(l+x+w\right) dwdx \\
&= \int_0^g \int_0^g \int_0^t u'''\left(l+x+w+z\right) dwdxdz.
\end{aligned}
$$

Accordingly, collective "optimism" dominates collective "pessimism," holding fixed the level of transfer, if the third derivative of the utility function is positive (less risk aversion at higher levels of consumption), while the opposite is true if the third derivative is negative. For a quadratic utility function the two collective identities will be welfare equivalent. Thus, we have the following proposition:

Proposition 1 *Let there be a symmetric endowment distribution in the* 3×2 *example. Then for* δ *sufficiently large, we have that optimism welfare-dominates pessimism (*$V_O^* > V_P^*$*) if* $u''' > 0$*,while pessimism dominates optimism (*$V_O^* < V_P^*$*) if* $u''' < 0$*.*[28]

Proof. From the expression for RHS above, we know that $V_O\left(t\right) \gtreqless V_P\left(t\right)$ as $u''' \gtreqless 0$, for all t. Moreover, for δ sufficiently large we know that any level of transfer $t = T(G, B)$ that generates a positive utility surplus for the agents is feasible. Hence, by a "revealed preference" argument, the optimal transfer under optimism must generate higher (lower) welfare than the optimal transfer under pessimism when $u''' > 0$ $(u''' < 0)$. ■

3.3.4 Why Collective Identities Promote Risk Sharing

Equation (8) also provides further insight into the features of a code configuration that tend to be associated with a greater scope for risk sharing [i.e., a lower value of $\bar{\delta}(C_1, C_2)$.] Indeed, examining the equation gives us a hint as to why symmetric configurations (collective identities) may foster gainful trade among the agents in a wider range of environments than asymmetric configurations. Let us rewrite the equation as follows:

$$\frac{v_2'(\tilde{x}_2, 0)}{v_1'(\tilde{x}_1, 0)} \cdot \left[1 + Z(\delta)\right]^{-1} = \frac{v_2'(\hat{x}_2, 0)}{v_1'(\hat{x}_1, 0)} \cdot \left[1 + W(\delta)\right], \tag{10}$$

[28] It does *not* follow from this result that optimism is an equilibrium collective identity when $u''' > 0$. That would require $V_O^* > V_M^{P*}$. Indeed, as demonstrated in Figure 1 (see section 4 below), under constant relative risk aversion (so $u''' > 0$), if the agents are sufficiently risk averse then the collective identity of pessimism is a dominant strategy Nash Equilibrium of the reduced-form first stage game, even though the optimistic configuration welfare dominates the pessimistic one.

where

$$Z(\delta) \equiv \frac{(1-\delta)\phi_1(\tilde{x}_1)}{\delta q_1(\tilde{x}_1) q_2(\tilde{x}_2) v_1'(\tilde{x}_1, 0)} \text{ and } W(\delta) \equiv \frac{(1-\delta)\phi_2(\hat{x}_2)}{\delta v_2'(\hat{x}_2, 0) q_1(\hat{x}_1) q_2(\hat{x}_2)}.$$

As mentioned, the functions $Z(\delta)$ and $W(\delta)$ represents current marginal cost per unit of expected discounted future cost, for incentive constrained givers at $(\tilde{x}_1, \tilde{x}_2)$ and $(\hat{x}_1, \hat{x}_2)$, respectively. They are strictly decreasing functions, vanishing as $\delta \uparrow 1$ and growing without bound as $\delta \downarrow 0$. Reasoning informally about equation (10) above, we note that $LHS(\delta = 1) - RHS(\delta = 1) = \frac{v_2'(\tilde{x}_2,0)}{v_1'(\tilde{x}_1,0)} - \frac{v_2'(\hat{x}_2,0)}{v_1'(\hat{x}_1,0)} > 0$.] As δ falls from 1, $Z(\delta)$ and $W(\delta)$ both rise until $\delta = \overline{\delta}$, and the gap between LHS and RHS vanishes. So, the wider is the *endowment disparity* (as measured by the difference$\left|\frac{v_2'(\tilde{x}_2,0)}{v_1'(\tilde{x}_1,0)} - \frac{v_2'(\hat{x}_2,0)}{v_1'(\hat{x}_1,0)}\right|$), other things equal, the smaller is the value of $\overline{\delta}$ at which (8) is satisfied. Likewise, the larger are the *mismatch frequencies* [as measured by $q_1(\tilde{x}_1)q_2(\tilde{x}_2)$ and $q_1(\hat{x}_1)q_2(\hat{x}_2)$], and the smaller are the shadow prices $\phi_1(\tilde{x}_1)$ and $\phi_2(\hat{x}_2)$, the smaller will be $\overline{\delta}$.

We can see, therefore, that in general the factors determining the magnitude of $\overline{\delta}$ are the same as those mentioned in section 2.2.2 above for the special case, $|X| = 2$: (1) the endowment disparity (i.e., difference of conditional marginal utilities) at the two commonly known events most favorable for trading; (2) the mismatch frequencies (i.e., the probabilities of these events); and, (3) the marginal cost of transfers at these events. As mentioned, the endowment disparity is greater when the agents' codes specifically identify widely disparate endowment states (i.e., when very high endowments and/or very low endowments are provided with their own distinct signals so that trade between the agents conditional on these endowment realization becomes possible.) On the other hand, allocating separate indicators to very high and very low states uses up cognitive resources while lowering the mismatch frequency (i.e., the probabilities of high/low and low/high indicator realizations fall as the endowment disparity rises.) So, as in the example of section 2, a fundamental trade-off (*endowment disparity* versus *mismatch frequency)* is involved in the general case. *The basic reason why collective identity configurations afford the agents a greater scope to realize gains from trade is that symmetry between the agents promotes the efficient management of this trade-off.*[29]

[29] To see this, it may help to consider the following problem: Let $\tilde{y}$ be any random variable continuously distributed on some interval of real numbers, Y. Find sets of "high" and "low" realizations of $\tilde{y}$ for the two agents, respectively denoted Y_i^H and Y_i^L, $i = 1, 2$, so as to:

$$\max\left\{ \sum_{i \neq j} \left[E\left(u'(\tilde{y}) \middle| Y_i^L\right) - E\left(u'(\tilde{y}) \middle| Y_i^H\right)\right] \text{ s.t. } \sum_{i \neq j} \Pr\{Y_i^L\}\Pr\{Y_j^H\} \geq \theta \right\}.$$

So, the sets, (Y_i^H, Y_i^L), $i \in \{1, 2\}$, are to be chosen to maximize the average difference in conditional marginal utilities

3.4 Optimal Transfers and Dysfunctional Identities

3.4.1 The General Problem

At the start of the second stage the agents adopt a risk sharing arrangement in anticipation of their infinitely repeated interaction. As mentioned, we assume that they agree to adopt the feasible arrangement that maximizes the sum of their expected discounted utilities. Given a code configuration and a discount factor, denote this optimal arrangement by $T^*[C_1, C_2; \delta]$. Thus:

$$T^*\left[C_1, C_2; \delta\right] \equiv \arg\max\{V_1(T) + V_2(T) | \, T \in \Im(C_1, C_2; \delta)\}. \tag{11}$$

Now, observe that (since X is a finite set) a period-stationary transfer arrangement T is simply an array of $|X|^2$ elements that are real numbers. Furthermore, recall that at each indicator pair (x_1, x_2) the LHS of inequalities (6) and (7) (which define feasible transfer arrangements) are strictly concave functions of (the elements of) T. Therefore, since no transfer can exceed the giver's endowment, and since the endowment set Y is bounded, we can identify the set of feasible transfer arrangements $\Im(C_1, C_2; \delta)$ with the points of a compact, convex subset of a finite dimensional Euclidean space. It follows that for every code configuration and discount factor there exists a

of the agents across high/low indicator realizations, subject to keeping the probability of a high/low realization above some bound, θ. A little thought and a bit of algebra (which we leave to the reader) reveals that, for numbers $y^H > y^L \in Y$, the solution for this maximization problem entails:

$$Y_1^H = Y_2^H = \left\{y \in Y \mid y \geq y^H\right\} \text{ and } Y_1^L = Y_2^L = \left\{y \in Y \mid y \leq y^L\right\},$$

where y^H and y^L are such that:

$$u'(y^H) - E[u'(\tilde{y}) \mid \tilde{y} \geq y^H] = E[u'(\tilde{y}) \mid \tilde{y} \leq y^L] - u'(y^L), \text{ and}$$

$$2\left[\sum_{y \geq y^H} p(y)\right]\left[\sum_{y \leq y^L} p(y)\right] = \theta.$$

Thus, for a given sum of mismatch frequencies, the following symmetric configuration yields the widest endowment disparity (and thus the greatest potential gain from the marginal trade): both agents are assigned identical "high" and "low" events at the upper and lower ends (respectively) of the endowment distribution. The two boundaries defining these events must be such that a high/low realization occurs with the required frequency, and such that the spread between the average shadow price of consumption and the shadow price at the boundary is equated across events. (Otherwise, one could widen the disparity in average shadow prices while maintaining the frequency of low/high encounters by adjusting the boundaries of the high and the low events.) Note that this argument is general, and does not depend on the assumption that $|X| = 2$, nor (if random signalling is allowed) on the assumption that Y is an interval of real numbers.

unique [since $V_i(T)$ are strictly concave functions] socially optimal transfer arrangement, as defined by (11).

Intuitively, at any indicator pair (x_1, x_2) the optimal arrangement $T^*[C_1, C_2; \delta](x_1, x_2)$ shifts resources from the agent with the lower to the one with the higher conditional shadow price, either until the post transfer conditional marginal valuations have become equal, or until the associated incentive constraint binds. For $\delta \approx 1$, the incentive constraints are guaranteed to hold for any transfer arrangement generating positive surplus for both agents, so in that case the optimal transfer at each indicator pair (x_1, x_2) is simply the unconstrained maximizer of the sum of conditional expected utilities there. Let $\tilde{T}(x_1, x_2)$ denote this best *unconstrained* arrangement. Then:

$$\tilde{T}(x_1, x_2) \equiv \arg\max \{ v_1(x_1, -t) + v_2(x_2, t) | \, t \in \Re \}, \text{ for all } (x_1, x_2) \in X^2.$$

Obviously, we must have $v_1'(x_1, -t) = v_2'(x_2, t)$ at $t = \tilde{T}(x_1, x_2)$.

It follows that, with a sufficiently large discount factor, the receiver of an optimal transfer will enjoy higher net consumption than the giver with positive probability. Each pair of indicators (x_1, x_2) is associated with a "rectangle" of endowments, $C_1^{-1}(x_1) \times C_2^{-1}(x_2)$. The unconstrained optimal arrangement equalizes conditional expected marginal utilities at indicator pairs. But then, because utility is strictly concave, when a giving agent consumes least at a fixed indicator pair her marginal utility exceeds the conditional expectation there. And, when a receiving agent consumes most her marginal utility falls short of the corresponding conditional expectation. Thus, for $\delta \approx 1$ the consumption of low endowment givers is always less than that of high endowment receivers at a given indicator pair. This is a useful fact, recorded for future reference as:

Lemma 3 *Given a code configuration $\langle C_1, C_2 \rangle$ there is a δ' sufficiently large such that, for any $\delta > \delta'$ and any indicator pair (x_1, x_2): If the optimal transfer $t = T^*(x_1, x_2) > 0$ then $y_1 - t < y_2 + t$ for some $y_1 \in C_1^{-1}(x_1)$ and $y_2 \in C_2^{-1}(x_2)$, while the opposite inequality obtains if $t < 0$.*[30]

For $\delta << 1$ the incentive constraints (6) and (7) become relevant, and the best feasible risk sharing arrangement can no longer be described quite so simply. However, we can readily characterize the optimal transfer arrangement in the general case. (This characterization can be used to compute optimal transfer arrangements in parametric examples, as we do in section 4 below.) This characterization is derived by adapting to our context an observation familiar from the theory of

[30] Strictly speaking, this should be a weak inequality, since consumption levels must be identical when both of the sets $C_i^{-1}(x_i)$ are singletons. But, that occurence is not to be expected in equilibrium, given that $|X| << |Y|$.

discounted repeated games: If one knew in advance the overall payoff accruing to each agent from the optimal arrangement, then explicitly deriving the detailed features of that arrangement would be a trivial exercise.

Accordingly, for a given code configuration $\langle C_1, C_2 \rangle$ and discount factor δ, and for arbitrary $w = (w_1, w_2) \in \Re^2_+$, consider the the function $\Psi : \Re^2_+ \to \Re^2_+$ defined by:

$$\Psi_i(w) \equiv V_i\left(\hat{T}\left[w\right]\right), i = 1, 2, \tag{12}$$

where $\hat{T}[w]$ solves:

$$\begin{aligned}
&\max\left\{V_1(T) + V_2(T)\right\} \text{ subject to, for all } (x_1, x_2) \in X^2\text{:}\\
&\delta w_1 \geq -\Delta u\left(\min\{C_1^{-1}(x_1)\}, -T(x_1, x_2)\right), \text{ and}\\
&\delta w_2 \geq -\Delta u\left(\min\{C_2^{-1}(x_2)\}, T(x_1, x_2)\right).
\end{aligned}$$

Thus, $\Psi_i(w)$ is the payoff to agent i associated with the socially most desirable, "pseudo-feasible" transfer arrangement, where "pseudo-feasibility" refers to transfer arrangements that would be feasible if the "psuedo-payoff" w_j were what agent j's anticipated to lose upon reversion to autarky, $j = 1, 2$. Notice that, so long as at least one incentive constraint binds for agent j, then $\Psi_i(w)$ rises and $\Psi_j(w)$ falls as w_j rises, $i \neq j$. This is because raising agent j's pseudo-payoff, w_j, loosens the incentive constraints for transfers going from agent j to agent i, but does not affect the pseudo-feasibility of transfers going from agent i to agent j. So, raising w_j can only lead to an increase in Ψ_i and a decline in Ψ_j.

Now, it is obvious that if $w' = \Psi(w')$, then $\hat{T}[w'] \in \Im(C_1, C_2; \delta)$. [That is, all transfer arrangements $\hat{T}[w']$ associated with the fixed points of $\Psi(\cdot)$ are feasible.] Moreover, it is also obvious that for $w_i^* \equiv V_i(T^*\left[C_1, C_2; \delta\right])$, $i = 1, 2$, we must have $w^* = \Psi(w^*)$. [That is, the payoffs engendered by the optimal transfer arrangement constitute a fixed point of $\Psi(\cdot)$.] Of course, not any fixed point of Ψ will do the trick here, since $(w_1, w_2) = (0, 0)$ (with $\hat{T} \equiv 0$) is always going to be among the "self-generating" payoffs. Nevertheless, we have the following characterization:

Theorem 2 *Fixing the code configuration $\langle C_1, C_2 \rangle$ and the discount factor δ, let $\Psi : \Re^2_+ \to \Re^2_+$ be given by (12) above, and define $\Gamma \equiv \{w' \in \Re^2_+ \mid \Psi(w') = w'\}$. Then Γ is non-empty and there exists a vector-maximal element, $w^* \in \Gamma$.*[31] *Moreover, $\hat{T}\left[w^*\right] = T^*\left[C_1, C_2; \delta\right]$.*

[31] That is, w^* is such that $w^* >> w'$, for all $w' \in \Gamma$, $w' \neq w^*$.

Proof. To prove the Theorem notice (from the foregoing discussion) that, given a code configuration and discount factor, a unique optimal transfer arrangement T^* exists (maximizing a strictly concave function over a compact, convex set.) Moreover, by the definition of optimality, $\langle V_1(T^*), V_2(T^*)\rangle \equiv w^* \in \Gamma$. Thus, we only need to show that w^* is the vector-maximal element of Γ.

So, let $w' \in \Gamma$, with $w' \neq w^*$. Since $\hat{T}[w']$ is feasible and T^* is optimal, it is impossible that $w' >> w^*$. So, without loss of generality, we assume that $w_2^* > w_2'$ and then proceed to show that $w_1^* > w_1'$.

The result is a straightforward consequence of the facts that $\Psi_i(w)$ increases and $\Psi_j(w)$ decreases as w_j rises, $i \neq j$. To see this, observe that: $w_1^* = \Psi_1(w^*) > \Psi_1(w_1^*, w_2')$ (since Ψ_1 increases with w_2 and $w_2^* > w_2'$.) Thus,

$$w_1^* - w_1' > \Psi_1(w_1^*, w_2') - \Psi_1(w').$$

At the same time, because Ψ_1 decreases with w_1, we must have that

$$\left(w_1^* - w_1'\right)\left[\Psi_1(w_1^*, w_2') - \Psi_1(w')\right] < 0.$$

It follows from these two inequalities that $w_1^* - w_1' > 0$, as was to be shown. (If the product of two numbers is negative, the larger of these numbers must be positive.) ■

3.4.2 Equilibrium in the Case $|X| = 2$, with Quadratic Utility and $\delta \approx 1$

We conclude the analysis of this section by considering the full equilibrium of the two-stage model in the special case where just two indicators are available. To keep the computations tractable, we assume further that δ is sufficiently large so incentive constraints can be ignored, and that $u(\cdot)$ may be closely approximated by a quadratic function. These are strong assumptions, to be sure. But our goals here are merely illustrative: (i) to show how decentralized, self-interested identity choices by the agents can lead them to embrace a dysfunctional collective identity; and, (ii) to see how the bias associated with this inefficient identity choice depends on the fundamentals of the problem.

Thus, for the remainder of this section we study the first stage, reduced-form game under the presumption that the unconstrained optimal transfer arrangement $\tilde{T}$ is to be implemented in each period of the second stage. We assume that the utility function can be written as follows:

$$u(y) = \alpha y - \frac{\beta}{2}y^2 + \frac{\gamma}{3}y^3, \tag{13}$$

where α and β are positive constants, and γ is a real number, of either sign, near zero.[32] Endowments y are assumed to be continuously distributed over some bounded interval of $\Re_+$, and the parameters are taken to be such that $u'(y) > 0$ and $u''(y) < 0$ throughout this interval.

We denote the density function of the endowment distribution by $p(y)$ and the cumulative distribution function by $P(y) = \int_{\{v \leq y\}} p(v)dv$. The mean endowment is denoted by $\mu = \int yp(y)dy$. For any quantile of the endowment distribution, $z \in [0,1]$, we denote by $\mu(z) \equiv P^{-1}(z)$ the endowment level associated with that quantile, and we define

$$\mu_z^+ \equiv E\left[y \middle| y > \mu(z)\right] \text{ and } \mu_z^- \equiv E\left[y \middle| y \leq \mu(z)\right].$$

That is, μ_z^+ (μ_z^-) is the mean endowment conditional on the being above (below) quantile z in the endowment distribution. Obviously, $\mu_z^+ > \mu > \mu_z^-$, and $z\mu_z^- + (1-z)\mu_z^+ \equiv \mu$ for all $z \in (0,1)$. Finally, again for $z \in [0,1]$, we define

$$\rho_z^+ \equiv E\left[y^2 \middle| y > \mu(z)\right] \text{ and } \rho_z^- \equiv E\left[y^2 \middle| y \leq \mu(z)\right].$$

We will undertake a perturbation analysis of the following form: Staring with $\gamma = 0$, we solve for the (unconstrained) optimal second-stage transfer arrangement. Holding this arrangement fixed but allowing γ to vary in a neighborhood of zero, we then derive equilibrium and optimal first-stage identity configurations, as functions of γ. That is, we ignore the impact on the optimal transfer arrangement that arises due to a "small" perturbation of the utility function in the neighborhood of a quadratic. This is justified since, for γ near zero, this impact is a second-order effect. The motivation for proceeding in the way will become clear in what follows. For now, it suffices to observe that quadratic utility ($\gamma = 0$) is a "knife-edge" case, where equilibrium and optimal identity configurations coincide, but this is generally not true for the perturbed utility function ($\gamma \neq 0$). So, we can use the perturbation analysis to see how the divergence of equilibrium from optimal configurations depends on the endowment distribution and on (the third derivative of) the utility function.

As mentioned, in the case $|X| = 2$, monotonic codes are defined by thresholds y_i^* such that $C_i(y) = B$ for $y \leq y_i^*$. Equivalently, we can describe such codes by the quantiles of the endowment distribution $z_i \equiv P(y_i^*)$ below which agents report a "bad" outcome. Thus, for the remainder of this section we identify a code configuration with a pair of numbers $(z_1, z_2) \in [0,1]^2$, and we denote

[32] That is, the term $\frac{\gamma}{3}y^3$is a perturbation of the quadratic utility function. We will examine how equilibrium and socially optimal identities vary with γ in a neighborhood of zero.

the per-period payoff to agent one under such a configuration by $U(z_1, z_2) \equiv (1-\delta)V_1$. Let (z_1^e, z_2^e) and (z_1^o, z_2^o) denote, respectively, the equilibrium and socially optimal configurations. Then, as discussed in section 2.2.2 above, the relationship between the equilibrium and the socially optimal configurations is determined by the sign of $U_2(z_1^e, z_2^e)$.

Now, with a quadratic utility function, marginal utility is linear in consumption. So, equating conditional expected marginal utilities between the agents amounts to equating conditional expected consumption levels. Thus, in this case the unconstrained optimal arrangement must be such that the better-off agent transfers to the worse-off agent an amount equal to half the difference in their conditional mean endowments. Moreover, with two indicators, $x \in \{B, G\}$, there are four possible indicator pairs,

$$(x_1, x_2) \in X^2 \equiv \{(B,B),\ (B,G),\ (G,B),\ (G,G)\},$$

and these indicator pairs are realized, respectively, with probabilities

$$q_1(x_1)q_2(x_2) \in \{z_1 z_2,\ z_1(1-z_2),\ (1-z_1)z_2,\ (1-z_1)(1-z_2)\}. \tag{14}$$

In light of the discussion to this point, it is clear that the unconstrained optimal transfer arrangement in the quadratic, two-indicator case satisfies:

$$\begin{aligned} T(B,B) &= \frac{\mu_{z_1}^- - \mu_{z_2}^-}{2},\ T(B,G) = \frac{\mu_{z_1}^- - \mu_{z_2}^+}{2}, \text{ and} \\ T(G,B) &= \frac{\mu_{z_1}^+ - \mu_{z_2}^-}{2},\ T(G,G) = \frac{\mu_{z_1}^+ - \mu_{z_2}^+}{2}. \end{aligned} \tag{15}$$

We conclude that for $T(x_1, x_2)$ given in (15), and for $C_1^{-1}(B) = \{y \leq \mu(z_1)\}$ and $C_1^{-1}(G) = \{y > \mu(z_1)\}$, agent one's per period expected payoff is given by:

$$U(z_1, z_2) = \sum_{(x_1,x_2)\in X^2} q_1(x_1)q_2(x_2)E\left[\Delta u\left(y, -T(x_1,x_2)\right)\middle|\, y \in C_1^{-1}(x_1)\right] \tag{16}$$

A simple computation shows that if $u(y) = \alpha y - \frac{\beta}{2}y^2 + \frac{\gamma}{3}y^3$, then

$$\Delta u(y,t) \equiv u(y+t) - u(y) = u(t) - \beta t y + \gamma(t^2 y + t y^2).$$

Hence, taking conditional expectations above (for $A \subset Y$ an event, and with $\mu_A \equiv E[y \mid A]$ and $\rho_A \equiv E\left[y^2 \mid A\right]$), we have that:

$$E\left[\Delta u(y,t) \mid A\right] = \alpha t - \beta\left(t\mu_A + \frac{t^2}{2}\right) + \gamma\left(t^2\mu_A + t\rho_A + \frac{t^3}{3}\right). \tag{17}$$

Now, using the formula in (16) above, the expression in (17) [taking conditional expectations at each indicator pair], the transfer arrangement given in (15), and the probabilities as given in (14), we can derive agent one's payoff via a straightforward but tedious computation (which is omitted here.)[33] To state the result compactly, we require just a bit more notation. Thus, for any quantile $z \in [0,1]$ define:

$$\sigma^2(z) \equiv z\left(\mu - \mu_z^-\right)^2 + (1-z)\left(\mu - \mu_z^+\right)^2 \text{ and } \sigma^3(z) \equiv z\left(\mu - \mu_z^-\right)^3 + (1-z)\left(\mu - \mu_z^+\right)^3.$$

Then, we have the following result:

Proposition 2 *For the utility function $u(\cdot)$ given in (13) and the transfer arrangement $T(\cdot,\cdot)$ given in (15), the expected per period surplus for agent one is:*

$$\begin{aligned} U(z_1,z_2) &= \left(\frac{3\beta+2\gamma\mu}{8}\right)\sigma^2(z_1) + \left(\frac{2\gamma\mu-\beta}{8}\right)\sigma^2(z_2) + \left(\frac{7\gamma}{24}\right)\sigma^3(z_1) \\ &\quad -\left(\frac{\gamma}{24}\right)\sigma^3(z_2) - \left(\frac{\gamma}{2}\right)H(z_1) + K, \end{aligned} \tag{18}$$

where $H(z) \equiv z\mu_z^-\rho_z^- + (1-z)\mu_z^+\rho_z^+$ and K is a constant, independent of (z_1,z_2).

We are now in a position to compare the equilibrium and socially optimal collective identities in this "near quadratic utility-two indicator" case. Notice that $U(z_1,z_2)$ is additively separable, so agent one has a dominant strategy identity choice in the first stage. Therefore, the unique equilibrium identity choice is the same for both agents in this case, and satisfies:

$$z_1^* = z_2^* = z^e \equiv \arg\max_{z\in[0,1]} \left\{(3\beta+2\gamma\mu)\,\sigma^2(z) + \left(\frac{7\gamma}{3}\right)\sigma^3(z) - 4\gamma H(z)\right\}. \tag{19}$$

By contrast, the unique socially optimal collective identity [which maximizes $U(z,z)$] is given in this case by:

$$z^o \equiv \arg\max_{z\in[0,1]} \left\{(\beta+2\gamma\mu)\,\sigma^2(z) + \gamma\sigma^3(z) - 2\gamma H(z)\right\}. \tag{20}$$

For γ in a neighborhood of zero, denote by $z^e(\gamma)$ the solution of (19), and let $z^o(\gamma)$ be the solution of (20). Then, the following result is immediate:

[33] Details of this computation are available from the authors upon request. The "trick" is to rewrite the optimal transfer (at indicator pair (B,B), say) as

$$T(B,B) = \frac{\mu_{z_1}^- - \mu_{z_2}^-}{2} = \frac{1}{2}[(\mu - \mu_{z_2}^-) - (\mu - \mu_{z_1}^-)],$$

and to use the identity

$$z(\mu - \mu_z^-) + (1-z)(\mu - \mu_z^+) \equiv 0$$

when evaluating the expectation over indicator pairs of the various terms in (17).

Corollary 2 *In the exact quadratic utility case ($\gamma = 0$), with $\delta \approx 1$, the equilibrium collective identity is socially optimal: $z^e(0) = z^o(0) = \arg\max_{z\in[0,1]} \left\{\sigma^2(z)\right\}$.*

To carry forward our perturbation analysis, we need to derive the sign of $\frac{d}{d\gamma}[z^e(\gamma) - z^o(\gamma)]$ at $\gamma = 0$. If this derivative is positive, then a near quadratic utility function with $\gamma > 0$ yields a dysfunctional collective identity where the agents are too pessimistic (that is, their probability of announcing a "good" endowment is too low). While, if this derivative is negative then the agents are too optimistic in equilibrium when $\gamma > 0$. (The opposite inferences apply when $\gamma < 0$.) Using the first-order conditions in (19) and (20) above, and applying the Implicit Function Theorem, we conclude that:

$$\frac{d}{d\gamma}\left[z^e(\gamma) - z^o(\gamma)\right]\Big|_{\gamma=0} = \frac{\left(\frac{4}{3}\right)\frac{d\sigma^3(z^o)}{dz} + 2H'(z^o)}{\beta\frac{d2}{dz^2}\left[\sigma^2(z^o)\right]}. \tag{21}$$

Since the second-order condition for (20) requires the denominator above to be negative, we have the following result:

Corollary 3 *In the near quadratic utility case ($\gamma \approx 0$), with $\delta \approx 1$, the equilibrium identity is too pessimistic (resp., too optimistic) if $\gamma \cdot \left[\frac{2}{3}\frac{d\sigma^3(z^o)}{dz} + H'(z^o)\right] < 0$ (resp. > 0), where $z^o \equiv \arg\max_{z\in[0,1]} \left\{\sigma^2(z)\right\}$.*

The condition (21) above is difficult to interpret, We note, however, (as may be easily verified) that when the endowment distribution is uniform the RHS in (21) vanishes. While, if a linear density function is posited, the RHS in (21) is positive when the density is increasing with y (i.e., when the distribution of endowments has a relatively fat right tail), and the RHS is negative when the density is decreasing (i.e., when the distribution of endowments has a fat left tail.) From this it follows (assuming linear densities and nearly quadratic utility) that if $u''' > 0$, then the equilibrium identity will be too pessimistic (optimistic) if the endowment distribution has a fat left (right) tail. The opposite conclusion obtains if $u''' < 0$.

4 Numerical Analysis of the 3×2 Case

We return now to the 3×2 example. We will employ a parametric class of utility functions $u(\cdot)$ to explore the comparative statics of that case. Consider the constant relative risk aversion (CRRA) family of utility functions

$$u(y) = \frac{y^{1-\rho}}{1-\rho}, \text{ with } \rho \in (0,1]$$

when $\rho = 1, u(y) = \ln y$. Given this utility function, the outcome in our model is determined by the distribution of random incomes, the discount factor, δ, and the risk aversion parameter, ρ. Note that CRRA utilities satisfy $u''' > 0$. Thus, the set of economic environments within which we work is defined as follows:

$$\left\langle \left(\{k, p_k\}_{k \in \{l,m,h\}}, \delta, \rho \right) : l < m < h, p_k \in (0,1), \sum_{k \in \{l,m,h\}} p_k = 1, \delta \in (0,1), \rho \in (0,1] \right\rangle.$$

In what follows, we study how equilibrium identities chosen in the first stage depend on the discount factor and the degree of risk aversion. We do this by calculating numerically the second stage continuation values under autarky, under the "optimistic" and "pessimistic" collective codings, and under the mixed coding. We then examine how these continuation values vary with the pair of parameters, (δ, ρ). Our numerical results are summarized in Figures 1-4.

4.1 Risk Aversion and Collective Identities

We first consider how the equilibrium coding choices are affected by the relative risk aversion parameter ρ. As ρ gets larger, the agent becomes more risk averse, which means (of course) that risk sharing becomes more valuable to them, other things equal. Thus, one way of interpreting the comparative statics exercise below is to think of an increase in the risk aversion parameter as reflecting a raising of the stakes for the agents in their second stage interactions. The equilibrium coding depends on the relative value of V_M^{O*} as compared with V_O^* and V_P^*. Figure 1 shows the differences between $\left(V_P^*, V_O^*, V_M^{P*}, V_M^{O*}\right)$ and the autarky value V_A as ρ varies. Note the figure depicts a threshold ρ^* such that when for any $\rho \in (0, \rho^*)$, we have the following inequalities:

$$V_O^* - V_A > V_M^{P*} - V_A; \quad V_P^* - V_A > V_M^{O*} - V_A.$$

The first equality implies that if the other agent is choosing C_O, I will be better off by choosing C_O, which will secure myself a value of V_O^*, than choosing C_P - which will only yield a value of V_M^{P*} for *me* since we would be at a mixed code equilibrium. Therefore, the first equality implies that $\langle C_O, C_O \rangle$ is an equilibrium. Analogously, the second inequality implies that the other agent is choosing C_P, I am better off choosing C_P than C_O, because a choice of C_P yields value of V_P^* and a choice of C_O a value of V_M^{O*}.

Therefore when $\rho \in (0, \rho^*)$, we have *multiple* equilibria collective identities. Moreover, the equilibria are Pareto-ranked: the "optimistic" equilibrium $\langle C_O, C_O \rangle$ Pareto dominates the "pessimistic" equilibrium $\langle C_P, C_P \rangle$.

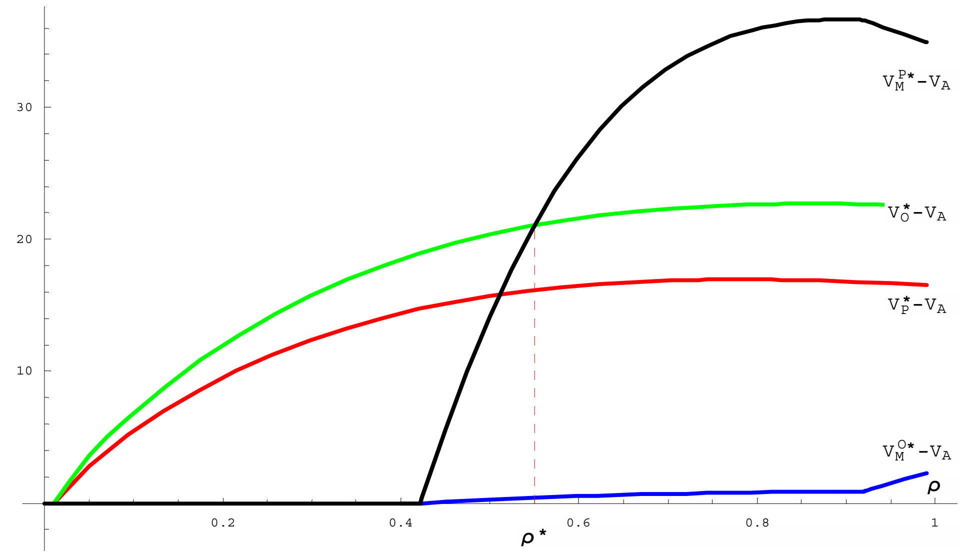

Figure 1: The Value Differences from the Autarky Value V_A as Functions of $\rho : p_l = 0.5, p_m = 0.2, l = 1, m = 6, h = 10, \delta = 0.99$.

When $\rho > \rho^*$, Figure 1 shows that

$$V_O^* - V_A < V_M^{P*} - V_A, \text{ but } V_P^* - V_A > V_M^{O*} - V_A.$$

Therefore, the unique equilibrium collective identity is the "pessimistic" identity $\langle C_P, C_P \rangle$. It is worth noting that if both agents can commit to choose the "optimistic" coding, both agents' value would be higher than the equilibrium value V_P^*. The "optimistic" coding does not constitute an equilibrium due to forces similar to the familiar "Prisoner's Dilemma."

4.2 Discount Factors and Collective Identities

The second comparative statics we do in this numerical exercise is with respect to the discount factor δ. A discount factor can capture many things in repeated game models, including expected stability of the relationship or the length of time elapsing between the repeated encounters. Here it is natural to think of the discount factor as capturing the density of the social network within which the agents interact follow their identity choices. That is, a large discount factor (near one) can be interpreted to mean that the repeated encounters are quite frequent, and thus the social network within which agents are embedded is dense.

Figure 2 shows three curves $V_P^* - V_M^{O*}$, $V_O^* - V_M^{P*}$ and $V_O^* - V_P^*$, all as functions of the discount

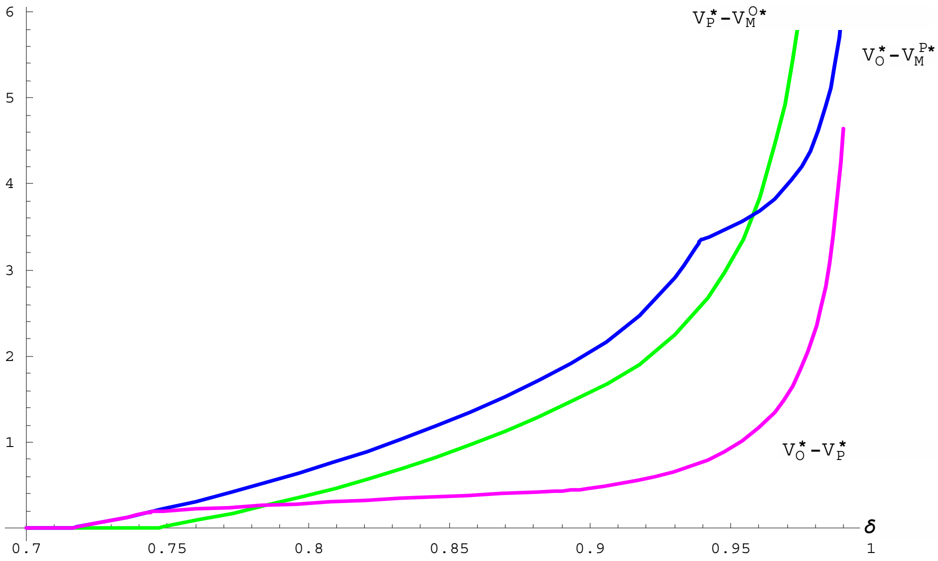

Figure 2: Relevant Value Differences as Function of δ : $p_l = 0.5, p_m = 0.2, l = 1, m = 6, h = 10, \rho = 0.5$.

factor δ, under a risk aversion parameter $\rho = 0.5$. Note that both $V_P^* - V_M^{O*}$ and $V_O^* - V_M^{P*}$ are strictly positive for all values of δ plotted. This indicates that there are two equilibrium coding for $\rho = 0.5$. The reason is simple. When $V_P^* - V_M^{O*}$ is positive, it means that, if the other agent is choosing a code C_P, I will be better off choosing C_P to secure a value of V_P^* than choosing C_O and obtain value V_M^{O*}. When $V_O^* - V_M^{P*}$ is positive, it means that, if the other agent is choosing a code C_O, then I will be better off choosing C_O to secure a value of V_O^* than choosing C_P and obtain value V_M^{P*}. Also note from Figure 2 that $V_O^* - V_P^*$ is also strictly positive for all values of δ plotted. This means that the "optimistic" coding equilibrium Pareto dominates the "pessimistic" coding equilibrium.

Figure 3 also shows three curves $V_P^* - V_M^{O*}$, $V_O^* - V_M^{P*}$ and $V_O^* - V_P^*$, all as functions of the discount factor δ, but under a risk aversion parameter $\rho = 0.8$. Note that, for this case, while $V_P^* - V_M^{O*}$ is strictly positive for all values of δ plotted, $V_O^* - V_M^{P*}$ is only positive when δ is less than a threshold δ^*. That is, communities with high degree of isolation, which implies a higher level of δ tend to have difficulty forming "optimistic" codes. Note that, a negative value of $V_O^* - V_M^{P*}$ when δ is high does not mean that V_O^* and V_M^{P*} are low, it just means that there is a profitable deviation when the other agent is choosing C_O. At all levels of δ, the simulation shows that $V_O^* > V_P^*$, that is, agents are better off under the "optimistic" code.

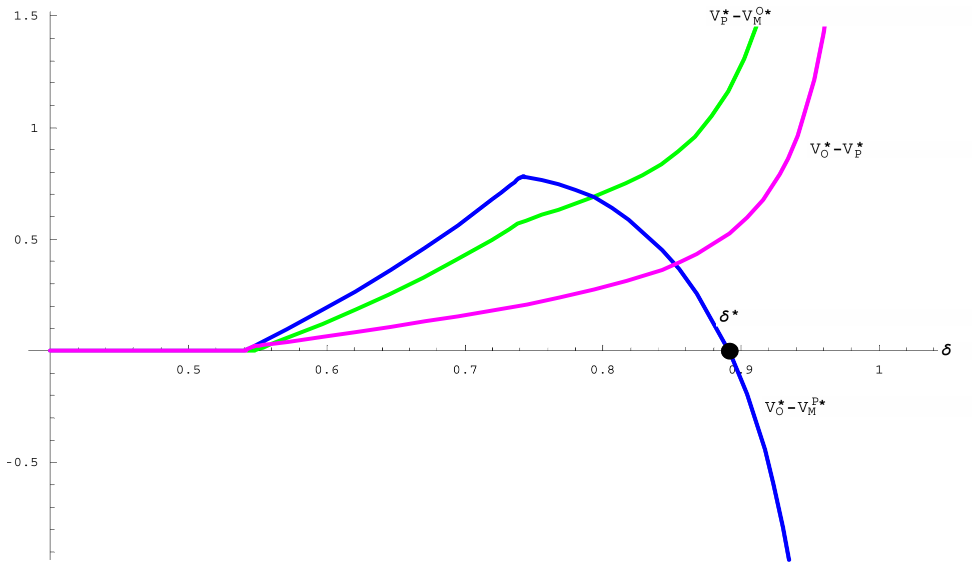

Figure 3: Relevant Value Differences as Function of δ : $p_l = 0.5, p_m = 0.2, l = 1, m = 6, h = 10, \rho = 0.8$.

4.3 Income Generating Processes and Collective Identities

Figures 1-3 are simulated from a particular income generating process characterized by (p_l, p_m, p_h, l, m, h). In general, the exact set of equilibrium collective identities will depend on the fine details of the income generating process. Here we show some simulation result for a symmetric environment to illustrate the results in Section 3.3.3. Consider a symmetric environment in which $p_l = p_h = 0.3$, $p_m = 0.4, l = 2, m = 6, h = 10$. In Figure 4, we let $\delta = 0.99$ and depict the difference between $V_P^*, V_O^*, V_M^{P*}, V_M^{O*}$ and the autarky value V_A. It turns out that under this particular income generating process, there is no scope for risk sharing under mixed codes. Thus $V_M^{O*} = V_M^{P*} = V_A$. Figure 4 shows that there are two equilibrium collective identities and the optimistic identity dominates the pessimistic identity, confirming our prediction in Proposition 1 because $u''' > 0$ for CRRA utility functions. Figure 5 shows the relevant payoff differences, as a function of δ, with ρ fixed at 0.8. It turns out the "prisoner dilemma" like situation does not arise under this symmetric income generating process.

5 Discussion and Next Steps

We now offer a few observations about our general approach and its limitations, as well as some suggestions for further work. The risk sharing problem studied here is simply a laboratory within

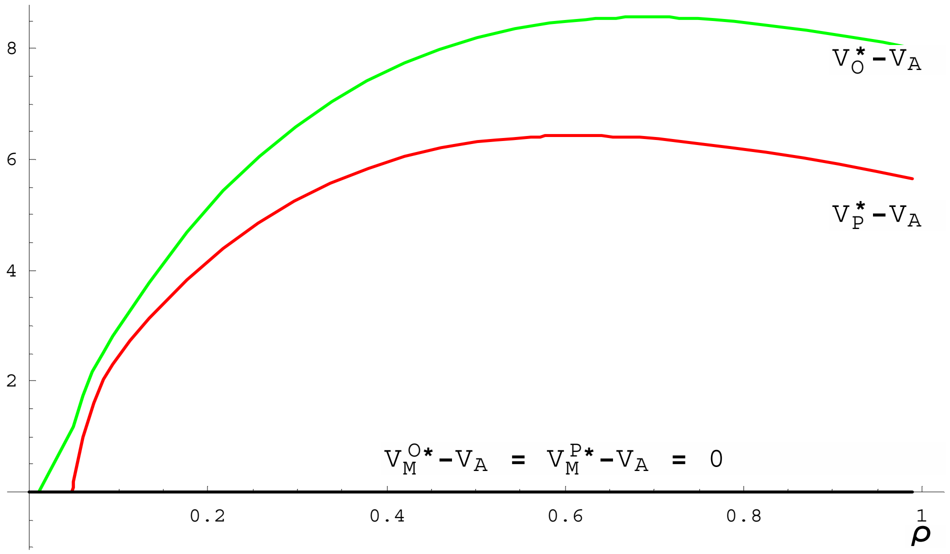

Figure 4: The Value Differences from the Autarky Value V_A as Functions of $\rho : p_l = 0.3, p_m = 0.4, l = 2, m = 6, h = 10, \delta = 0.99$.

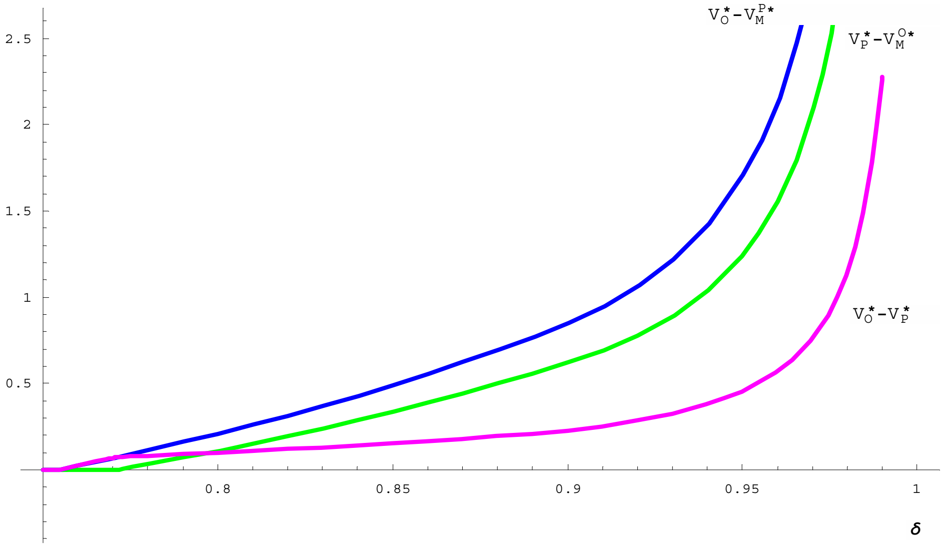

Figure 5: Relevant Value Differences as Function of $\delta : p_l = 0.3, p_m = 0.4, l = 2, m = 6, h = 10, \rho = 0.8$.

which to explore our main idea – which is modelling identity as the "coding" of personal history into a simplified form.[34] One nice thing about the risk sharing formulation, though, is that it easily accommodates an analysis of some classic identity oppositions – the "optimist" versus the "pessimist," for instance; or, one with a "wide" versus a "narrow" sense of what constitutes her "victimization." This framework also makes it easy to get an intuitive grasp of how identity choice interacts with economic behavior: In our model, choosing an identity is equivalent to deciding upon a way to limit the information that is publicly available about one's (income) experience. Studying the consequences of such endogenous restrictions on public information is a relatively easy thing to do in a risk sharing context, and the resulting model is rich with implications.

We have derived results in our simple, two-person setting about when the strategic interaction leading to collective identity choice is a coordination game, and when it is a prisoner's dilemma, based on the size of the discount factor and the degree of risk aversion. The numerical analysis (which assumed constant relative risk aversion) showed that when risk aversion is below a threshold, the simultaneous choice of identity in the first stage is a coordination game; when it is above a threshold, and when the discount factor is high enough, then we have a prisoner's dilemma with "pessimistic" coding being the only equilibrium, although it is Pareto dominated by the "optimistic" coding. This seems to us a very interesting finding. The degree of risk aversion may be taken as a proxy for the importance of the economic interactions that are being influenced by identity choices. The more risk averse are the agents, the more is at stake in their risk sharing interactions. This finding from our numerical analysis, therefore, can be interpreted as saying that when a great deal is at stake in their risk sharing interactions, "pessimism" (or, embracing a "wide," not "narrow," sense of what constitutes "victimization") is likely to emerge as a dysfunctional collective identity. We also find that, when the risk aversion is low, first stage identity choice is a coordination game for all discount factor. Thus, when the stakes are not very high for the economic interaction, multiple equilibria are likely. This means that two similarly situated but socially isolated populations could

[34] For instance, we might just as well have studied a repeated Prisoner's Dilemma – with pair-wise random matching of agents to play in each period, and public but noisy signals about each agent's history of play depending on her choice of code. Or, we could have pursued our agenda in a market setting – an exchange economy, say, with many heterogenous agents whose preferences depend on code-mediated narratives about their "types." Then, the aggregate demand function would vary with the distribution of adopted codes in the population, and market-clearing prices would both depend on but also help to detemine the equilibrium distribution of codes. One can surely think of other interactions where the coordinated choices of identities have economic consequence. All of this makes for a fit subject for further research.

end up making widely different (though equally “rational”) identity choices.

While we believe the approach to “identity” offered here is promising, we recognize that our analysis has some serious limitations. Our model endogenizes the choice of identity, but this is a once-and-for-all choice. We allow for no evolution of identities, no chance for agents to “invest” in remaking their identities (through education, relocation, sex change operations(!), etc.) In reality, of course, identities can evolve. One vexing question is why this malleability of identity is greater in some “cultures” than in others.[35] Relatedly, suppose agents cannot perfectly observe each other’s identities, but they can learn about each other over time. To revisit some examples mentioned in the introduction: Is he gay or not? Is she a single-minded career woman, or not? Is that white guy over there really angry? In reality, the first stage identity choice game would not be in normal form. And, this being the case, it is far from clear that the multiple equilibria we find to exist in a static identity choice setting would survive the dynamics of equilibrium selection if agents’ identities were imperfectly observed, but agents could learn about each other over time.

Our theory (but also, common sense) emphasizes that “identity” is endogenous, and is shaped by social contacts. So, the question arises: What kind of social networks in which people might be embedded lead to what kinds of choices about identity? This is a particularly interesting question for someone studying race, culture and social inequality in the U.S. One implication of our theory, in a slightly expanded model allowing for the assortative matching of agents from distinct groups before playing the second stage repeated game, is that distinctive patterns of identity choices by individuals in distinct groups is more likely if patterns of social interaction are more group-segregated. This leads us to speculate that anyone who believes “culture” is important in sustaining racial inequality in a society like the US should look seriously at the linkages between identity and social integration. Casual empiricists make much of the observable differences in “values” between distinct groups. But, our analysis points toward a recognition of the fact that such cultural difference may be parasitic upon a pre-existing disparity in the structures of social interaction.[36] If group inequality is partly due to cultural differences, if cultural variation is partly a matter of distinct identity choices, and if identity choices diverge in part because of segregated social networks, then social *integration* of some sort might be an antidote for inequality.

But, what kind of integration would be most important, and which egalitarian interventions

[35] This question is taken-up by Chris Barrett in his contribution to this volume.

[36] After all, one hears very little about collective identities based on hair length, eye color or shoe size. This may be due to the fact that segregated patterns of social interaction along such lines as these are virtually non-existent!

might be most effective? Consider, for instance, the distinction emphasized by political scientist Robert Putnam between "bridging" and "bonding" social connection.[37] "Bridges" are connections between people belonging to different (racial/ethnic) groups; "bonds" are connections between people in the same group. In general, a social network is characterized by its "nodes" (people) and its "links" (connections), where the links might be thought of as coming in these two flavors – bridges and bonds.[38] All of this suggests what might be a useful way of thinking about the connection between culture and inequality. Whether or not a given (possibly dysfunctional) pattern of behavior becomes normative within an economically backward, socially isolated group (so that conformity pressures favoring that behavior can develop) could depend in interesting ways on what might be called the *architecture* of the social network in which the group is embedded – that is, on the density and relative frequency of these two types of bonds. Our approach could be extended in this direction, perhaps even to include the study of endogenously generated racial identities.[39] We plan to pursue this possibility in future work.

In another vein that we intend to pursue, our conceptualization of identity seems to be similar to *language* in the following sense: Saying that different communities can embrace different collective identities (in the sense that states of the world are understood differently due to the limited cognitive capacity) is isomorphic to saying that different communities can adopt different languages.[40] That is, while one community may have a word to describe a particular state of the world, in another community no word may exist to specifically describe that state of world. So, our framework (extended to incorporate moral hazard in the endowment generating process) might be useful for asking why we do not observe complete languages – that is, languages in which a different word exists to describe each possible state of nature. One reason could be that ambiguity is sometimes useful in a world where there is a trade-off between effort incentives and risk sharing. More specifically, it may be true that, given any fixed income generating process, a complete vocabulary could achieve better risk sharing; but, if incomes are endogenous, then this completeness might undermine incentives to

[37] See the article describing Putnam's recent work in *The Economist*, Feb. 26, 2004.

[38] An interesting and accessible discussion of the mathematics of such networks, emphasizing the importance of this kind of qualitative distinction between different types of links, is Barabási (2002).

[39] Sociologist Mary Waters (2001) provides an interesting illustration of the complexity of racial identity choice among black Americans. Based on her extensive interviews with first and second-generation West Indian immigrants, she draws a rich and enlightening contrast between the self-definitions adopted by these subjects versus those embraced by the more indigenous black American population.

[40] This interpretation and possible extension of our work has been suggested to us by Antonio Merlo.

take income-enhancing effort.

Finally, we wish to discuss informally one more possible extension of this line of inquiry. It is an implication of Lemma 3, as applied to the 3×2 case, that when the discount factor is sufficiently large the post-transfer consumption of an agent with the intermediate income will be greater than that of an agent with the high income under "pessimism," but less than that of an agent with the low income under "optimism." This finding could have interesting implications when the model is extended to allow for endogenous efforts. It suggests that an effort disincentive could exist under either type of collective identity, and that the nature of this disincentive might depend on the precise way that effort shifts the distribution of random incomes.

Thus, consider the following speculative argument: Let the parameters be such that the identity choice game is a coordination game. Modify the game by inserting an (unobservable) effort choice prior to the realization of incomes in each period of the second stage. Effort is costly, but it makes higher income realizations more likely to occur. If the result of more effort is to raise the probability of a high income and to lower the probability of an intermediate income, leaving the probability of a low income unchanged, then the "pessimistic" coding (which makes post-transfer consumption lower for the high than for the intermediate income agent) may lead to an overall equilibrium with low effort, compared to the "optimistic" coding. That is, *"pessimism" could be a dysfunctional collective identity when it is difficult to reduce the chance of poverty but possible to increase the chance of becoming rich through high effort.* Similarly, if the result of effort is mainly to raise the probability of an intermediate income and to lower the probability of a low income, leaving the probability of a high income unchanged, then the "optimistic" coding (which makes post-transfer consumption lower for the middle than the low income agent) may lead to equilibrium with low effort, compared to the "pessimistic" coding. In this case, *"optimism" could be a dysfunctional collective identity when it is difficult to increase the chance of becoming rich but possible to reduce the chance of being poor through high effort.*

Of course, this is all just conjecture at this point. But these conjectures, based on the interaction between collective identity and the technology of income improvement, seem quite intriguing to us. Note that in both of these speculative instances, a certain monotonicity property fails: higher effort does not increase the likelihood ratio of every income level relative to all lower income levels. What the discussion suggests is that incentive problems may cause a collective identity choice to be inefficient when this non-monotonicity overlaps with the clustering of income states under a code. That is, a dysfunctional collective identity may come about when effort causes the higher income state within a code to become relatively less likely than the lower income state. We will be looking at this possibility in the next phase of this research program.

6 Conclusion

Managing collective action problems is itself a collective action problem. In our model, two agents need to share resources in order to smooth consumption over time. This is their collective action problem, and how the agents manage it depends on how they interpret their personal experiences to one another. We imagine that agents decide on a way to publicly render their private experiences, mindful of the fact that any subsequent transactions between them must be framed in terms of those renderings. This framework implies a non-cooperative game of identity choice, where "identity" is understood to be a way of rendering "the self" to others. We have shown that, under a wide range of conditions, the strategic forces of this game favor the agents adopting a common, collective identity in equilibrium. Moreover, we have also shown that when the density of their interactions and their potential gains from trade are sufficiently great, the equilibrium of this implied identity game has a "tragedy of the commons" character, and a universally superior way exists for agents to render their experiences to one another. So, under these conditions their collective identity can be said to be *dysfunctional.* This classical economic insight is a principle benefit of the approach to "identity" that we are proposing here.

References

[1] Akerlof, George. 1997. "Social Distance and Social Decisions," *Econometrica,* 65 (5): 1005-1027.

[2] Akerlof, George and Rachel Kranton. 2002. "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature, 40* (4): 1167-1201.

[3] Akerlof, George and Rachel Kranton. 2000. "Economics and Identity," *The Quarterly Journal of Economics.* 115 (3): 715-753.

[4] Aronson, Joshua, Claude M. Steele, M.F. Salinas, and M.J. Lustina. 2003. "The Effects of Stereotype Threat on the Standardized Test Performance of College Students." in *Readings About the Social Animal*, 8th edition, E. Aronson (ed.) New York: Freeman.

[5] Banfield, Edward. 1970. *The Unheavenly City: The Nature and Future of Our Urban Crisis.* Little Brown, Boston.

[6] Baumeister, Roy. 1998. "The Self," in Gilbert, Daniel T., S. Fiske and G. Lindzey. *Handbook of Social Psychology*, (2 volumes). Oxford University Press. Vol. 1, 680-740.

[7] Barabási, Albert-Lászlo. 2002 *Linked: How Everything Is Connected to Everything Else and What It Means.* Perseus Books Group.

[8] Barrett, Christopher. 2005. "Smallholder Identities and Social Networks: The Challenge of Improving Productivity and Welfare," (this volume, Chapter 9).

[9] Benoit, Jean-Pierre and Vijay Krishna. 1985. "Finitely Repeated Games," *Econometrica* (July): 905-922.

[10] Bernheim, Douglas. 1994. "A Theory of Conformity," *Journal of Political Economy*, 102 (5): 841-877.

[11] Burns, Justine. 2004. "Race and Trust in Post Apartheid South Africa," unpublished manuscript. Santa Fe Institute.

[12] Denizet-Lewis, Benoit. 2003. "Double Lives on the Down Low," *The New York Times Sunday Magazine.* (August 3, 2003).

[13] Douglas, Mary. 2004. "Traditional Culture – Let's Hear No More About It." Chapter 4 in *Culture and Public Action,* editted by V. Rao and M. Walton, Stanford University Press, 85-109.

[14] *The Economist.* 2004. "The kindness of strangers?" Feb. 26, 2004.

[15] Ferguson, Ronald. (in press). "Why America's Black-White School Achievement Gap Persists," in *Ethnicity, Social Mobility and Public Policy in the US and the UK,* editted by G. Loury, T. Modood and S. Teles. Cambridge University Press.

[16] Fryer, Roland. 2003. "An Economic Approach to Cultural Capital," unpublished manuscript, Harvard University.

[17] Fryer, Roland and Matt Jackson. 2002. "Categorical Cognition: A Psychological Model of Categories and Identification in Decision Making," unpublished manuscript, Harvard University.

[18] Goffman, Erving. 1963. *Stigma: Notes on the Management of Spoiled Identities.* Simon & Schuster.

[19] Grief, Avner. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *The Journal of Political Economy*, 102 (5): 912-950.

[20] Hoff, Karla and P. Pandey. 2003. "Why are Social Inequalities So Durable?" Discussion Paper. World Bank and Pennsylvania State University.

[21] Loury, Glenn C. 2002. *The Anatomy of Racial Inequality.* Harvard University Press.

[22] Loury, Glenn C., Steven Teles and Tariq Modood, eds. (in press). *Ethnicity, Social Mobility and Public Policy in the US and the UK.* Cambridge University Press.

[23] McWhorter, John. 2000. *Losing the Race: Self-Sabatoge in Black America.* The Free Press.

[24] North, Douglas. 1981. *Structure and Change in Economic History.* W.W. Norton & Company.

[25] Ogbu, John. 2003. *Black Students in Affluent Suburbs: A Study in Academic Disengagement.* Mahwah, NJ: Lawrence Erlbaum.

[26] Sidanius, James and Felicia Pratto. 2001. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression.* Cambridge University Press.

[27] Waters, Mary. 2001. *Black Identities: West Indian Immigrant Dreams and American Realities.* Harvard University Press.