The cost of social division

Manuel Muñoz-Herrera[*] [†]

Saturday 16th October, 2021

For the most recent version click here

or go to: https://manumunoz.github.io/files/JMP_ManuMunoz.pdf

Abstract

At times groups collectively follow inefficient behaviors motivated by their group identity. Changes in social behavior are hard because they are a collective action problem that requires coordination. So, groups may end up socially divided into a fraction that abandons the group's identity and behavior for a more efficient one and the rest that persist on it. In this paper, I study both theoretically and experimentally the conditions under which social division arises and the consequences it has on coordination and welfare. In the model, efficiency requires all individuals to abandon the inefficient group identity. Thus, social division is inefficient but stable. In the experiment, I find social division to be a prominent outcome, which has negative consequences in multiple dimensions: (i) individuals that adopt the efficient group identity fewer chances to integrate with others if they come from a socially divided group, (ii) individuals that persist on the inefficient group identity lose social support as part of their group abandons them, and (iii) divided groups display less altruistic behavior towards others, which has distributive consequences. This highlights that partially abandoning an inefficient group identity can have negative consequences beyond those individuals that persist on it.

Keywords: social norms, networks, coordination, inter-group contact, identity change
JEL Classification: C92, D91, D85

---

[*]Center for Behavioral Institutional Design, New York University Abu Dhabi (e-mail: manumunoz@nyu.edu, website: manumunoz.me).

[†]I am grateful to Kanchan Chandra, Aleksei Chernulich, Sanjeev Goyal, Frederic Moisan, Nikos Nikiforakis, Ernesto Reuben, Pedro Rey-Biel, Marc Witte, Yves Zenou and participants at a number of seminars for helpful comments. I am grateful for financial support from Tamkeen under the NYUAD Research Institute award for Project CG005.

1 Introduction

Group identities prescribe rules of behavior, and individuals act in conformity to those rules when they expect others they care about to do the same (Bicchieri 2006, 2016). For example, if a person is asked to split a sum of money and she divides it into equal parts, this may be because she believes others in her reference network would expect her to do so, and do the same in a similar situation. At times the prescriptions from an identity may be beneficial, such as those sustaining fairness or reciprocity (Gachter et al. 2017). Other times, people may follow behaviors that are inefficient or even detrimental because their choices are conditional on what they believe others in their group will (expect them to) do. Examples such as child marriage (Bicchieri 2016) or corruptive behavior (Xiao and Bicchieri 2012; Xiao 2013) showcase behaviors that are sustained because of social expectations, even when some people would rather change them and adopt different ones.

The challenge is that to change a group behavior a collective action problem must be overcome, because the benefits from conforming to a group's prescription stem from coordinating one's choices with many others (Baronchelli 2018). That is, an individual may end up worse off if she abandons her group's prescription and adopts a different one on her own than if she sticks to the inefficient behavior together with her peers. The reason is that by conforming to a different group identity that person may experience social sanctions from those she cares about, such as exclusion, discrimination or penalties (Battu and Zenou 2010; Bisin et al. 2016). Only when there is enough support from others to adopt a new group's prescription, the sanctions can be minimized to the point that it is beneficial to abandon the current one (Riedl et al. 2016; Goyal et al. 2020). Another notable barrier to identity (and behavior) change arises at times when it is not possible to publicly signal whether someone supports a change in behavior. When group identities are difficult to detect, entire groups may hold false views of the prevailing prescriptions among their peers, and end up in a belief trap where they think they are the only ones willing to abandon the inefficient behavior. This barrier to identity change is known as pluralistic ignorance (Schank 1932; Prentice and Miller 1993).

Studying how barriers to group-identity change such as social sanctions or pluralistic ignorance impact collective outcomes is central to our understanding of social change and economic mobility (Boyd and Richerson 1992, 2002; Henrich and Boyd 2001). However, looking at the collective behavior of a social group, beyond the aggregate behavior of a set of individuals, is an empirical challenge that tends to be unfeasible in the field. For instance, surveys may be able to reach a subset but not the entire social group (Bicchieri 2016). Or even when all group members are reached, the self-report of their choices is made in isolation, which impedes to capture the complex interdependence embedded in collective outcomes (Charness et al. 2007). With this in mind, in this paper I study both theoretically and through a laboratory experiment *how barriers to group identity change shape the collective choices of a social group that initially conforms to an inefficient behavior but has the option to abandon it and adopt a more efficient one*.

Naturally, identity change can occur in multiple settings. A commonly studied environment is one where the entire population belongs to the same social group and a subset of group members wants to change the prescribed behavior they conform to. In these settings the choice for group-identity changers lies between conforming to the group's behavior or leaving the group to form an independent community (Bicchieri 2006, 2016). An example is the case of the British settlers that initially migrated to North America. In this paper, I will focus on a different environment that is widely pervasive in our complex world: the case in which the population is divided into various social groups, where each group conforms to a different form of behavior (Eckel and Grossman 2005; Goyal et al. 2020). In one of the social groups (e.g., a minority) an inefficient behavior is prescribed and followed, while members of the other group (e.g., a mainstream majority) conform to a more efficient one. In this setting, the choice of group-identity changers from the inefficient group lies between persisting on the in-group's identity and behavior or adopting the out-group's prescription and act accordingly.[1]

There are key considerations in the process of group identity change when different social groups interact, because if someone chooses to abandon her current group's identity and behavior there is already a group of people conforming to the behavior she wants to adopt. This has the potential to facilitate identity change, compared to the case where everyone belongs to one same group. However, because those others belong to the out-group, identity concerns may arise and limit inter-group contact (McPherson et al. 2001; Lowe 2021; Mousa 2020). Specifically, to benefit from the new group identity, adopters require ties with members of the out-group. However, the out-group may be unwilling to connect with identity changers, especially if group membership is associated with external markers (e.g., skin color, accent, height, tattoos) that can trigger doubts that the "new comers" will actually conform to a behavior that opposes their original group's identity (Farrell and Saloner 1985; Lane 2016; Barr et al. 2018). On top of this challenges to inter-group contact, there may be explicit barriers to identity change, such as social sanctions or pluralistic ignorance (as mentioned above), creating a complex mix from which it is not clear how social groups will behave and which collective outcomes will emerge.

[1] There are numerous examples of how minority groups are trapped in (or escape from) disadvantageous positions given their differences with the mainstream majority. See Hoff and Pandey (2006) for an example of ethnic minorities, Saleh and Tirole (2021) for one of religious minorities, Candelo et al. (2017) and Choi et al. (2019, 2021) for linguistic minorities, and Chandra (2012) for political minorities.

I first tackle this problem by formalizing a model of social coordination where two social groups interact but one is guided by a more efficient normative prescription (group identity) than the other. The model is based on three main elements that encompass identity change: (i) groups have opposing identities prescribing how to behave, (ii) individuals choose with whom to relate, i.e. their reference network, and the benefit of adhering to a given identity is interdependent on the behavior of their connections, and (iii) individuals have agency to abandon their group's identity and adopt a different. My model extends previous work on social coordination by introducing identity change as an endogenous choice (see e.g., Goyal et al. 2020).[2]

The main findings of the model indicate that, independently of which barrier to identity change a group is exposed to, three types of collective outcomes can emerge in equilibrium. First, a *no change outcome* where everyone persists on conforming to their group's identity, regardless of how efficient it is. This implies segregation between social groups because the behaviors they follow are incompatible. On the opposite end, a *global change outcome* where everyone in the group following an inefficient prescription abandon their group identity and adopts the other one. This leads to inter-group contact, given that everyone chooses the same behavior, making relations across groups beneficial. Finally, a *partial change outcome* where the inefficient group is divided into those who adopt and those who oppose the out-group's identity. This means that in equilibrium, members of the social group following the inefficient identity can act either as a united collective where all choose to persist or abandon the group's prescription on how to behave, or can be fractured into subgroups following opposing group identities.

Efficiency-wise, the *global change outcome* is the first-best because everyone benefits from the choices of both in-group and out-group members. The second-best is the *no change outcome* where the social group conforming to the inefficient identity sticks together. So, although they do not improve by choosing the efficient identity, acting in unity keeps the groups size intact and everyone benefits from the choices of all others in their in-group. Finally, the worse case arises in the *partial change outcome*, where the social group is fractured. Because of this, those who persist on their group's identity are conforming to an inefficient group identity and only coordinating with a fraction of their in-group and no one from the out-group. Given the theory predicts multiple equilibria, identifying which of these outcomes is most likely to emerge is an empirical question. For this, I design a laboratory experiment that allows me to exogenously vary the barriers identity changers face and to evaluate individual and collective outcomes when different social groups interact. There are three main treatments in the experiment: BASELINE, SANCTIONS, and

[2] See Andreoni et al. (2021) for a study of collective outcomes where changes in group's identities are exogenous, so that individuals have no choice over them.

IGNORANCE.

In the BASELINE treatment there are no barriers to identity change. If an individual abandons her group's identity, there are no social sanctions imposed on her and, everyone sees the group identity she has chosen. In the SANCTIONS treatment, I exogenously impose social sanctions to identity changers through a monetary cost. The cost decreases on the number of in-group members that also abandon the group's identity, to capture the essence of social sanctions: identity change can be costly but with enough support from others in one's group, the intensity of the sanction is reduced (Alesina and La Ferrara 2002; Fryer and Torelli 2010). Finally, in the IGNORANCE treatment, I eliminate the possibility to observe chosen group identities. Specifically, group membership is visible by external markers but identity choice is not. It is evident whether individuals come from one group or the other but not whether they oppose or adhere to the efficient group identity, which allows me to capture the essence of pluralistic ignorance: individuals cannot identify when others have abandoned their group's identity and the prescription associated with it.

I do not conduct a treatment combining social sanctions and pluralistic ignorance because, as argued by Fryer and Torelli (2010) and Bisin et al. (2011), these two barriers are mutually exclusive. An individual can either signal her adoption of a new group identity or conceal it. Making the group identity visible guarantees integrating with the out-group but exposes the adopter to social sanctions from her in-group. Keeping it hidden impedes sanctions but puts the adopter in a setting of pluralistic ignorance, where her choice is not identified by either the out-group or other adopters in her in-group. I impose each barrier exogenously in separate treatments to better understand the way social groups confront them. To complement this choice in design, I conduct a fourth treatment where I make barriers endogenous. In the REVEAL treatment, individuals choose whether to adopt the other group's identity without sanctions to those that change. However, as in the IGNORANCE treatment, no one can see the identity choice of others, only the group markers. Subsequently, every individual is able to make her chosen identity public, which exposes them to social sanctions (as in the SANCTIONS treatment).

The main result of the experiment indicates that groups following an inefficient identity are fractured and fail to achieve a *global change outcome* significantly more often when exposed to social sanctions as opposed to pluralistic ignorance. In the SANCTIONS treatments, social groups end up in *partial change* half of the times, which is twice more likely than in the IGNORANCE treatment and four times more than in the BASELINE. That is, social sanctions are barriers that erode the unity of the group while pluralistic ignorance is a barrier that drives the group to remain united (even if it prevents identity change). My work in general, and this result in particular, contributes to the research on cultural assimilation, which aims to identify the conditions under which social groups display oppositional identities (see e.g., Bisin et al. 2016).[3] I show that both barriers impede the abandonment of inefficient group identities, but highlight that social sanctions are most likely to prevent social change. This finding also provides some implications for public policy, by suggesting that the potential economic benefits of identity change may not be enough to guarantee individuals from abandoning inefficient group prescriptions, reinforcing the need for political leaders to look at group processes and at the types of barriers faced.

The result showing that groups act in unity in the case of pluralistic ignorance is quite remarkable and echoes previously observed phenomena, in which similar collective action problems are solved by tacit coordination (see e.g., Opp and Gern 1993; Bicchieri 2016). To further this finding, the REVEAL treatment provides an ideal setting where individuals facing pluralistic ignorance are given the option to publicly broadcast their identity choice, at the expense of potential social sanctions. The findings indicate that when individuals keep their group identities concealed, groups are more likely to choose one same identity in unison, while revealing chosen identities leads more often to a division within the social group. My work raises a warning for policies aiming at solving issues of pluralistic ignorance via public announcements of chosen identities (see e.g., Bicchieri 2016): providing communication channels that broadcast a change in group identities can backfire when the fear of sanctions remains, for it may divide the social group and limit economic mobility.

Finally, at the end of the experiment every individual participates in an other-other allocation decision, by distributing a pool of resources between a receiver from the in-group and one from the out-group (see e.g., Chen and Li 2009). This allows me to test the extent to which sharing a common group identity in the coordination setting impacts the distributive preferences of the individuals in each group in an independent environment. The literature on group identity, intergroup contact and social preferences (see e.g., Tajfel and Turner 1979; Akerlof and Kranton 2000), has found that adopting a common identity across groups can reduce in-group bias, leading individuals to treat those in the out-group as they treat their in-group members (Chen and Li 2009; Gaertner and Dovidio 2000; Dasgupta and Goyal 2019; Lees and Cikara 2020). However, I find an asymmetric effect between social groups. Individuals in social groups where everyone changes to the efficient group identity make more equitable distributions of resources between the two receivers. However, this has no differential effect on the distributive preferences of the passive group. This result also provides insights for policies promoting changes in group's

[3] Oppositional identities have been shown to affect employment (Battu and Zenou 2010) and academic performance of those individuals who oppose the mainstream group (Akerlof and Kranton 2002; Austen-Smith and Fryer 2005).

prescribed behavior, suggesting that if the prescriptions targeted are those a group already possesses, inter-group attitudes towards those that assimilate may remain unchanged. As such, efforts to promote integration and to reduce out-group tensions should not be placed only on the identity changers but also on the mainstream group.

The rest of the paper is organized as follows. In Section 2, I formalize the model and characterize equilibrium outcomes. Section 3 contains the experimental design and hypotheses. I report the results of the experiment in Section 4. In Section 5, I explore the effect of endogenously choosing between barriers. Finally, in Section 6, I conclude with a discussion.[4]

2 Theory

In this section, I model group identity change to evaluate collective outcomes through a coordination game where two social groups interact: a mainstream majority and an opposing minority. Minority and majority members are guided by different group identities and, as such, have preferences to behave in opposing ways. Benefits from conforming to a group identity come from the number of others a player coordinates with. As such, the size of the group serves to model economic differences and inequality. That is, the identity of the majority group is more efficient because more people would want to conform to it (for related models see Hernandez et al. 2013; Ellwardt et al. 2016; Goyal et al. 2020). Minority players can, however, abandon their group's identity and adopt that of the mainstream majority. Identity change has the potential to foster network connections across groups, resulting in economic mobility for the minority. A particular interest is to evaluate if in equilibrium the minority adopts one same group identity collectively or if it divides into a fraction that becomes mainstream and a fraction that persists in opposing the majority.

2.1 The model

Suppose the population N is fixed, of size n, and composed of a mainstream majority group and an opposing minority group. I denote by subscript m the majority group N_{m}, of size n_{m} and by subscript x the opposing minority group N_{x}, of size n_{x} (with $n_{\mathsf{m}} > n_{\mathsf{x}}$ and $n_{\mathsf{m}} + n_{\mathsf{x}} = n$). The two groups can be *ex-ante* differentiated by group markers and

[4] In addition, in Appendix A I include proofs and additional formalizations of the model. Appendix B contains all regression tables. Appendix C provides a detailed description of the sample and additional analysis from the experiment. Finally, in Appendix D I include the experimental instructions.

group identities. Group markers are external and fixed, such as skin color, hair, size, etc., which are exogenous to the individual (see e.g., Efferson et al. 2008).[5] On the other hand, group identities are expressed as normative prescriptions, e.g. preferences, views and values of the group, which are initially exogenous but can be affected by individual decisions. That is, before the game begins, a player i is exogenously assigned an identity $\theta_i = \mathsf{m}$ if she belongs to the majority and $\theta_i = \mathsf{x}$ if she belongs to the minority.[6]

The game has three stages. In Stage 1, every minority player $i \in N_{\mathsf{x}}$ chooses which group identity to adhere to $\bar{\theta}_i \in \{\mathsf{m},\mathsf{x}\}$. This results in two types of minority players: a player i who changes her group identity by adopting the mainstream views of the majority, $\bar{\theta}_i = \mathsf{m} \neq \theta_i$, and a player i who maintains her identity, $\bar{\theta}_i = \theta_i = \mathsf{x}$. Majority players are mainstream and do not change their group identity, so that $\bar{\theta}_i = \theta_i$ for all $i \in N_{\mathsf{m}}$.[7] Thus, I define $\bar{N}_{\mathsf{m}} = \{i \in N : \bar{\theta}_i = \mathsf{m}\}$ of size $\bar{n}_{\mathsf{m}}$ and $\bar{N}_{\mathsf{x}} = \{i \in N : \bar{\theta}_i = \mathsf{x}\}$ of size $\bar{n}_{\mathsf{x}}$ as the sets of players choosing group identity m and x, respectively. The resulting vector of identities $\bar{\theta} = \{\bar{\theta}_1, \ldots, \bar{\theta}_n\}$ is common knowledge.

In Stage 2, every player $i \in N$ chooses a set of link proposals g_i with others, $g_i = (g_{i1}, \ldots, g_{ii-1}, g_{ii+1}, \ldots, g_{in})$, where $g_{ij} \in \{0, 1\}$ for any $j \in N\backslash\{i\}$. Let $G_i = \{0, 1\}^{n-1}$ define i's set of link proposals. The cost of proposing a link is $c > 0$, which is independent of whether a link was reciprocated or not. The induced network $g = (g_1, \ldots, g_n)$ is a directed graph. The closure of g is an undirected network denoted by $\bar{g}$, where $\bar{g}_{ij} = g_{ij}g_{ji}$ for every $i, j \in N$. The set of i's neighbors in the network $\bar{g}$ is $N_i(\bar{g}) = \{j \in N : \bar{g}_{ij} = 1\}$ with cardinality $|N_i(\bar{g})| = n_i$. Define the set of all undirected networks $\bar{g}$ as $\bar{G}$.

In Stage 3, every player i decides between two possible actions $a_i = \{\mathsf{m}, \mathsf{x}\}$, the same for all of her neighbors. Denote the vector of chosen actions as a. Building on Akerlof and Kranton (2000) and Bicchieri (2016), a group identity specifies prescriptions and preferences on how to act. Accordingly, I model normative prescriptions through differences in the gains from choosing alike with others. A mainstream individual, i.e. either a majority player or a minority player who adheres to the group identity of the majority, prefers

[5] Arguably, even attributes such as skin color could change over time or across generations. Therefore, conceptualizing markers as fixed includes features that are permanent as well as those that cannot be altered in the short term. For a detailed discussion on the endogenous construction of identities and the mutability of traits see Chandra (2012).

[6] Group identities and, consequently, prescriptive norms can be exogenously imposed by a process of intergenerational transmission of traits (see e.g., Bisin and Verdier 2001). Also, there are exogenous alterations of social and group identities through publicity and marketing (Akerlof and Kranton 2000), or via organizations, such as universities or the army, which commonly induce identification within its members (see also Andreoni et al. 2021).

[7] See Section 2.4 for a discussion on the implications of assuming that only the minority can change their group identity and consequently their prescriptive norms. Also, see Section 4.1 in the Results for a test of this assumption.

the mainstream action m to the opposing action x, while a minority player who does not abandon her group's identity prefers action x to m. Specifically, the return for i from coordinating on her prescribed action with a neighbor j is v. On the other hand, coordinating on the non-prescribed action returns $v-r$, where r represents a loss in utility i experiences for acting in misalignment with her adopted group identity. There are no returns from a connection if neighbors do not coordinate their actions. Given an outcome profile $(\bar{\theta}, g, a)$, the utility of player i is defined as:

$$u_i(\bar{\theta}_i, \bar{g}, a) = \lambda(1 + \sum_{j \in N_i(\bar{g})} \mathbb{1}_{\{a_j = a_i\}}) - cn_i \tag{1}$$

where $\mathbb{1}_{\{a_j = a_i\}}$ is the indicator function that yields 1 if player j chooses the same action as player i, $\lambda = v$ if the action i chooses is aligned with her group identity, $a_i = \bar{\theta}_i$, and $\lambda = v - r$ if i acts against her group identity. The number of proposals made is n_i, which is multiplied by the cost of making a link proposal c, where $0 < c < v - r$. [8]

This is a multi-stage game with simultaneous decisions in each stage. A player's strategy is a contingent plan, specifying a choice at each stage conditional on each possible previous decision profile. In the following section, I characterize equilibrium outcomes.

2.2 Equilibrium analysis

The aim of the equilibrium characterization is to identify if outcomes where the minority is divided are stable, in comparison to outcomes where the minority group collectively shares one same group identity. For this, I look at two settings with different assumptions of what it implies to adopt a group identity (i.e., normative prescription). In the first, it is assumed that adherence to a group identity implies conformity on the action prescribed by such an identity (Schelling 1960; Bernheim 1994; Akerlof 1997). This means that when a player chooses the mainstream identity m in Stage 1, she will follow through by playing the mainstream action m in Stage 3. The same holds for a player choosing the oppositional prescription x. In the alternative setting, I evaluate the assumption that conformity is not enforced and group identities are, thus, non-enforceable (Bicchieri 2006; Schram and Charness 2015; Fehr and Schurtenberger 2018). In a nutshell, I find that outcomes predicted under enforceable identities remain stable when conformity is not imposed. I report the results for the enforceable case in the main text, and relegate the case of non-enforceability of group prescriptions to Appendix A, for it is more extensive. I

[8] Observe that if $v - r < c$, then no player will benefit from choosing an action in misalignment with their group's prescription. Moreover, if $v < c$, then no player benefits from proposing any link.

discuss differences between the two settings in Section 2.4.

The assumption that the prescriptions of behavior from adopting a group identity are enforceable binding transforms the 3-stage game into a 2-stage game. That is, players choose a group identity and action in Stage 1 and then form links in Stage 2. I now solve the two-stage game with identity (and action) choice and link formation. For this, I adapt the pairwise stability notion from Jackson and Wolinsky (1996) to my setting.[9] In the spirit of their definition, I say that a network is stable if no individual can profitably deviate either unilaterally or with one other individual.

Definition 1 *An identity-network pair $(\bar{\theta}, \bar{g})$ is pairwise stable if:*

- *for every $\bar{g}_{ij} = 1, u_i(\bar{\theta}, \bar{g}) \geq u_i(\bar{\theta}, \bar{g} - \bar{g}_{ij})$ and $u_j(\bar{\theta}, \bar{g}) \geq u_j(\bar{\theta}, \bar{g} - \bar{g}_{ij})$.*
- *for every $\bar{g}_{ij} = 0, u_i(\bar{\theta}, \bar{g}) \geq u_i(\bar{\theta}, \bar{g} + \bar{g}_{ij})$ or $u_j(\bar{\theta}, \bar{g}) \geq u_j(\bar{\theta}, \bar{g} + \bar{g}_{ij})$*

In this definition, part (1) says that no player can delete an existing link and profit, while part (2) says that no pair of players can form an additional link and increase their payoffs. This leads to the following proposition:

Proposition 1 *The pair $(\bar{\theta}^*, \bar{g}^*)$ is pairwise stable if one of the following obtains:*

(i) All minority players adopt the mainstream group identity, so that $\bar{\theta}_i^* = \mathsf{m}$ for all $i \in N$ and $\bar{g}^*$ is a complete network.

(ii) Some or no minority players adopt the mainstream group identity, so that the network $\bar{g}^*$ is divided into two complete components, $Q_\mathsf{m} = \bar{N}_\mathsf{m}$ and $Q_\mathsf{x} = \bar{N}_\mathsf{x}$, where every player in component Q_m follows identity m, while every player in Q_x follows identity x.

Proof. Player i's payoff from linking to a player j is $v - c$ if both choose the same group identity, and consequently the same action, or $-c$ if i chooses different from j. So, if $\bar{\theta}_i = \bar{\theta}_j$, both players have incentives to connect, and given deviations are bilateral they will. Otherwise, if their chosen identities are not the same, neither of them has incentives to connect. □

The intuition from Proposition 1 is that in a pairwise-stable outcome, players in the majority and the minority are willing to establish connections across groups when the latter

[9] I use pairwise stability as a solution concept instead of Nash equilibrium because of the two-sided linking in the model, for which Nash equilibrium is too permissive and, thus, uninformative. For example, it is stable under Nash equilibrium for two players choosing the same group identity/action to not be connected with each other, despite both having incentives to connect.

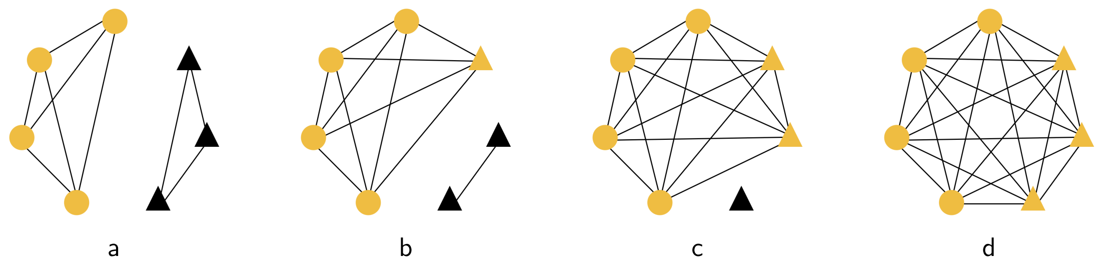


Figure 1 Examples of pairwise-stable outcomes.

A player from the majority is illustrated as a circle and a player from the minority as a triangle. Nodes of players choosing the mainstream group identity and action are filled with the lighter color, while those rejecting the mainstream group identity are filled with the darker color. A line connecting two nodes represents an undirected link. Network 1a portrays the *no change outcome* where no minority player abandons her group's identity. Networks 1b and 1c each illustrates a case of the *partial change outcome*, where the minority is divided into a fraction that becomes mainstream and the rest that oppose the majority. Network 1d illustrates the *global change outcome*.

adopt the mainstream group identity, for it implies they will also choose the mainstream action. This points to three types of equilibrium outcomes. Proposition 1(i) describes the *global change outcome*, where all minority players abandon their group's prescription to adopt the identity of the mainstream majority. Therefore, any pair of unconnected players i and j are better off forming a link and increasing their utility by v, given that both follow the same normative prescription. If all players adopt the mainstream group identity the level of inter-group contact is maximal, as the complete network is the pairwise-stable outcome.

Proposition 1(ii) describes two outcomes, one when some minority players adopt the mainstream identity and the other when no one adopts it. The first is a *partial change outcome*, which implies that the minority is socially divided into a fraction that abandons the group's prescription to become mainstream and the rest that persist on it. When the minority is divided, adopters integrate with the majority players in Component Q_{m}, while non-adopters remain isolated from the majority in the network component Q_{x}. The second case is the *no change outcome*, where all players maintain their original group identity in unity. As such, there is no integration between groups because the minority collectively chooses the same identity. All minority players are in Q_{x} and there are only majority players in Q_{m}.

Figure 1 illustrates examples of these outcomes ($n = 7$, $n_{\mathsf{m}} = 4$, $n_{\mathsf{x}} = 3$). Majority players are displayed using circles and minority players using triangles. Irrespective of their group markers, the nodes of players choosing the mainstream prescription (and action) are filled with the lighter color, while the nodes of players who do not change prescriptions are filled with the darker color. The sequence of outcomes in Figure 1 illustrates the progression from a pairwise-stable outcome where no one in the minority adopts the mainstream views to one where all minority players become mainstream. In Figure 1a all players maintain their original group identity. In Figure 1b, one of the minority players adopts the majority's identity while the other two oppose it. In Figure 1c, there is only one minority player opposing the prescription of the majority, while the other two are adopters. Finally, in Figure 1d, all minority players abandon their group identity to become mainstream.

Now I turn to social welfare, which I define as the sum of payoffs of all players. An outcome is said to be socially efficient if it maximizes aggregate welfare. I show that the *global change outcome* maximizes social welfare.

Proposition 2 *A complete network where all minority players change their group identity for the mainstream prescription is socially efficient.*

The proof is presented in Appendix A. Intuitively, this result states that a fraction of minority players opposing the mainstream views is never socially desirable. This is the case, because opposers divide into a separate community from those who become mainstream. As such, opposers are connected with fewer others and earn less than those who adopted the mainstream group identity. However if everyone in the minority becomes mainstream, all players in both the majority and the minority are better off. Figure 2 illustrates the utility of a majority player (solid line) and of a minority adopter (long-dash line) as a function of the number of adopters. Notably, without any identity change, minority players are worse-off than majority players. Such inequality is due to differences in group size (see the difference between lines at the intercept, when there are zero adopters). However, once a player in the minority becomes mainstream, her utility in the pairwise-stable outcome is the same as that of a majority player. Naturally, although not depicted in Figure 2, the utility of opposers decreases linearly on the number of adopters, given their network shrinks. Therefore, in a *partial change outcome* inequality between adopters and majority players disappears, while inequality between adopters and non-adopters within the minority group emerges. However, when all minority players adopt the mainstream group views, every player $i \in N$ gets $u_i = vn - c(n-1)$, which maximizes efficiency and eliminates inequality within and between groups.

2.3 Costs and social sanctions

In multiple settings identity change is not free. There are material and social costs for minority individuals who choose to become mainstream. Let $\delta > 0$ be a positive cost that any minority player pays for changing her group identity.[10] The effect of the identity-

[10] This transforms the payoff function in Equation 1 into $u_i(\bar{\theta}_i, \bar{g}, a) = \lambda(1 + \sum_{j \in N_i(\bar{g})} \mathbb{1}_{\{a_j = a_i\}}) - (cn_i - \delta)$

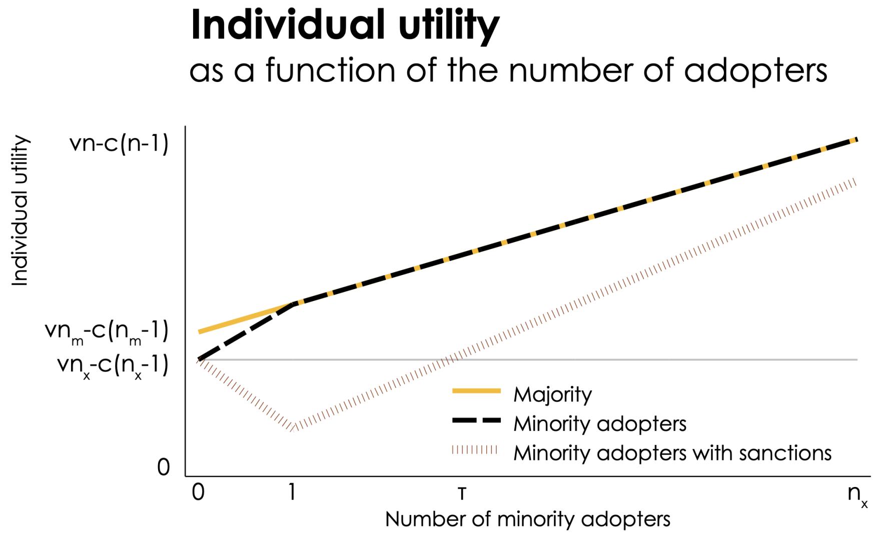


Figure 2 Individual utility with respect to the number of minority adopters.
The graph displays the utility of individual players in pairwise-stable outcomes (vertical axis) as a function of the number of adopters of the mainstream group identity within the minority (horizontal axis). The solid black line displays the utility for a majority player. The long-dashed line displays the utility for a minority player who adopts the mainstream views. The dotted line displays the utility for a minority adopter who is exposed to social sanctions, i.e., the cost of identity change increases with the number of opposers. The horizontal gray line references the utility of a minority player in the *no change outcome*. Values below this line display cases where group-identity change is detrimental and values above display cases where identity change leads to economic mobility for the minority player.

change cost is presented in the following proposition:

Proposition 3 *If the identity-change cost δ is:*

(i) $\delta > vn - c(n-1)$*, a minority player can never adopt the mainstream prescription and be in a pairwise-stable outcome.*

(ii) $0 < \delta < v(n_{\mathrm{m}} + 1) - cn_{\mathrm{m}}$*, a minority player can adopt the mainstream group identity and be in a pairwise-stable outcome, even when no one else in the minority becomes mainstream.*

(iii) $v(n_{\mathrm{m}} + 1) - cn_{\mathrm{m}} < \delta < (v - c)(n_{\mathrm{m}} + n_{\mathrm{x}} - \tau) + v$*, a minority player can adopt the mainstream group identity and be in a pairwise-stable outcome, as long as τ others in the minority adopt the mainstream prescription as well.*

The proof is presented in Appendix A. Intuitively, Proposition 3(i) states that if the cost of identity change is higher than the payoff a minority player earns in the *global change* *outcome*, her utility from adopting the majority's views would be always negative. As such, only the *no change outcome* is pairwise stable. Proposition 3(ii) gives a lower bound on the identity-change cost, such that any fraction of minority players adopting the majority's views can be sustained in a pairwise-stable outcome. For this, the cost must be lower than the earnings that a minority player gets in the *partial change outcome* where she is the only adopter. Finally, 3(iii) states that if the identity-change cost is within the two bounds mentioned above, a minority player only adopts the mainstream views when there are τ other minority players adopting, as well. Thus, the higher the value of δ, the larger the fraction of the minority that must become mainstream to sustain a *partial change outcome*.

A particular case of Proposition 3(iii) takes place when the cost of identity change is not fixed but rather decreases with the number of adopters in the minority group. This type of social cost has been evidenced in settings where non-adopters exert peer pressure on those in-group members who become mainstream (see e.g., Austen-Smith and Fryer 2005; Battu and Zenou 2010; Fryer and Torelli 2010). The aim of such social sanctions on adopters is to police the group's boundaries and prevent peers from leaving, i.e., becoming mainstream. Consequently, instead of only having a fixed element, identity-change costs may increase as a function of the number of minority players opposing the mainstream views. For example, the cost of identity change can be expressed by $\delta + \omega(n_{\mathsf{x}} - \tau)$, where ω is the rate at which the cost increases on the number of non-adopters, where $n_{\mathsf{x}} - \tau$ is the number of players in the minority minus those who adopt the mainstream group identity.[11] This leads to the following corollary to Proposition 3:

Corollary 1 *In a setting of social sanctions, where the identity-change cost $\delta + \omega(n_{\mathsf{x}} - \tau)$ decreases with the number of adopters, a minority player becomes mainstream if $\tau \geq \frac{\delta+\omega n-v}{(v+\omega-c)} - n_{\mathsf{m}}$ of her peers adopts the mainstream identity as well.*

The intuition from Corollary 1 is that if identity-change costs encompass social sanctions, utility for adopters does not increase monotonically, as in the case without cost. Instead, it is a convex function in which partial identity-change leads to lower utility than no change at all, when the number of adopters is below the thresholds τ, and it leads to higher utility when it is above it. An example of this is illustrated by the dotted line in Figure 2.[12]

[11] Naturally, there are many more functional forms one could use to express identity-change costs under social sanctions. However, the linear function is useful because of its simplicity and because the main goal of the example is to show how the cost can vary with the fraction of adopters in the group.

[12] For this illustration I use the following parameters: $\omega = c$, $\delta = v$, and $N_{\mathsf{m}} = N_{\mathsf{x}} + 1$. Parameters are discussed in more detail in Section 3.

2.4 Discussion of the assumptions

I conclude the theory section with a discussion of the main assumptions underlying the analysis. I do so by addressing variations to the characterization presented above, which highlights the role played by some of its key features. All proofs are present in Appendix A.

A1. *Group-identity choice for the minority alone*

In the model, I restrict group-identity choice to the minority. Assuming that majority members are mainstream and do not change their group identity is in line with other models in the literature, see for example Bisin et al. (2011). However, to further explore the implications of this assumption, I characterize pairwise-stable outcomes for the case in which majority players can also change their normative prescriptions and oppose the mainstream views (see Appendix A). I show that relaxing this assumption does not exclude any of the pairwise-stable outcomes characterized in Proposition 1. It does, however, allow for other outcomes to be pairwise stable. Among these is, for example, the *partial change outcome* where the entire majority gives up on being mainstream and adopts the views of the minority, while the entire minority becomes mainstream. Although possible, outcomes like these are unlikely and not of central interest to my study on minority identity-change. Nonetheless, in the experiment I test the robustness of this assumption by allowing majority players to choose whether to keep or change their group identity. Previewing the findings, even when the assumption is relaxed and all players can change prescriptions, majority players persist on keeping their group identity, irrespective of what minority players do (see Section 4.1).

A2. *Enforceable normative prescriptions*

A key assumption in the equilibrium characterization above is that the group identity a player chooses in Stage 1 works as an enforceable commitment on the action she plays in Stage 3. Relaxing this assumption, by making the normative prescriptions associated with a group identity non-enforceable, turns the setting back into a 3-stage sequential game. Players are still potentially better-off aligning their action with the identity they choose, so that each gets v instead of $v-r$ for every connection in a pairwise-stable outcome. That is, all outcomes characterized with enforceable group identities remain pairwise stable if the prescription of a group identity is non-enforceable. However, with non-enforceability the action is chosen after the linking stage. As such, there are pairwise-stable outcomes in which a player best responds by conforming to her reference network rather than by conforming to her group identity. This is the case of a player who is mostly connected to neighbors choosing her disliked action, where she is better off acting against her group identity and benefiting from the large share of neighbors she can coordinate with (see Appendix A). As such, in addition to the pairwise-stable outcomes described in Proposition 1, there are also outcomes where players choose an action in misalignment with their chosen group identity. In all cases, the same type of properties for the equilibrium outcomes hold: players are connected to all others choosing the same action, and to no one else.

I test the robustness of the enforceability assumption in the experiment, by allowing all players to choose actions independently of their chosen group identity. Previewing the findings, behavior in the experiment shows strong correlations between the *chosen* group identity in Stage 1 and the action chosen in Stage 3, for both majority and minority players.

A3. *Visible group identities*

In the model there is complete information about the vector of chosen group identities. There are however, many settings of incomplete information where minority players who adopt the mainstream identity and minority players who oppose it are, a priori, indistinguishable (see e.g., Efferson et al. 2008). Such cases are the basis for pluralistic ignorance. I do not provide a formal characterization of this setting, but instead discuss some of its implications: If identity choice is unobservable, there is strategic uncertainty for everyone in the game. Even for those in the minority respect to others in the minority. Because of this, even if beliefs about the number of adopters are accurate and the same for all players, there are high risks when only a fraction of the minority becomes mainstream. For example, in a *partial change outcome*, if a single minority player adopts the mainstream identity, she is unable to show she is the adopter. This can lead to losses for the adopter if she proposes links to players in the majority. Also, for those in the majority if they mistakenly propose links to a minority player that did not become mainstream, given they know for sure someone adopted but cannot tell who among the minority players is the adopter. It can also lead to losses for non-adopters in the minority because they may propose to the adopter, instead of a non-adopter, for the same reasons mentioned above.

Naturally, with incomplete information about group-identity choices also includes the case in which players propose links correctly. As such, the outcomes described in Proposition 1 are also pairwise stable when identity choices are not observable. Note also, that if beliefs are correct and symmetric (everyone has information about the number of adopters but not their identities), the case where no one becomes mainstream or the case where everyone does are equivalent to those of complete information, and so the risk is minimized compared to outcomes of partial change.

To conclude this section, I summarize and comment on the theoretical analysis: there are three specific outcomes of interest. A *no change outcome* where every player maintains her original group identity, such that the minority acts in unity and forms an isolated community separate from the majority. A *global change outcome* where all minority players adopt the mainstream identity and society fully integrates. The latter is the socially efficient outcome and also the case where the minority attains economic mobility, because inequality between groups is minimized. Finally, a *partial change outcome* where unity of the minority is eroded, so that a fraction of the minority becomes mainstream while the remaining fraction opposes the majority's normative prescriptions. Minority players can be worse-off in *partial change* compared *no change* or *global change*, if there are barriers such as pluralistic ignorance or social sanctions. Thus, the multiplicity of equilibria leaves open the question of which outcome is more likely to emerge in the different settings. I conduct a laboratory experiment to examine this and test the predictions of the model.

3 Experimental design

3.1 The experimental coordination game

For the experiment, I consider groups of 7 participants who interact together during the entire session. Prior to the start of play, four participants are randomly assigned to the majority and three to the minority.[13] The coordination game follows the setup of the model, in that participants from both the majority and the minority make decisions in all stages. By allowing participants in the majority to choose group identities, I avoid biasing outcomes towards minority identity change and, instead, allow these outcomes to emerge naturally. Also, by allowing all players to choose actions, I can experimentally test the conjecture that identity choice works as a commitment device, even when the chosen normative prescriptions are non-enforceable.

In Stage 1 of the game, each individual chooses whether to maintain her original group identity or to change it. Everyone can be differentiated on the screen through group and identity symbols. Specifically, there is an internal and an external symbol put together for each participant (see Figure 3a). The external symbol is a fixed marker that signals the group a participant is assigned to: *empty circle*, $\bigcirc$, for the majority and *empty triangle*, $\triangle$,

[13] In Goyal et al. (2020) it was shown that having a small difference in size between groups enhances tensions, while having a very large majority is always conducive of conformity by the minority (see also Ellwardt et al. 2016). Based on this, I choose a group composition of 4 majority and 3 minority members, which represents an ideal ratio to test the effects of identity change.

for the minority. The internal symbol signals the group identity a participant has chosen: *filled circle*, ●, for the mainstream identity of the majority and *filled triangle*, ▲, for the identity of the minority.

While the external symbol is kept fixed, the internal symbol may change according to the decisions individuals make. This results in four combinations of symbols that visually display group assignment and identity choice. Figure 3b shows an example where participant 7 is assigned to the minority and adopts the group identity of the majority (△, ●), while 1, also from the minority, maintains her group identity (△, ▲). Similarly, 6 is a majority participant who keeps her group identity (○, ●), while 5 changes it (○, ▲).

A participant who chooses the group identity of the majority, ●, prefers the mainstream action m to the opposing action x, and one who chooses the identity of the minority, ▲, prefers action x to m; in both cases irrespective of her group markers.[14] Participants earn $v = 6$ points for each coordination if they choose their prescribed action and $v - r = 4$ points for coordinating on their non-prescribed action, where $r = 2$ is the utility loss for acting in misalignment with their chosen group identity.[15]

In Stage 2, individuals simultaneously make costly link proposals to any of the other 6 participants. Each proposal, whether reciprocated or not, costs $c = 2$. Figure 3b presents the screen participants see in the linking stage. Every participant is assigned a numeric label to facilitate identification in the current period, and also a node displaying her corresponding pair of symbols (larger for the decision-maker).

In Stage 3, participants observe the emerging network and choose an action. As illustrated in Figure 3c, participants are informed of the proposals made and received, displayed as incoming and outgoing arrows respectively (thicker for the decision-maker). For example, participant 7 has links with $2, 5$ and 6. She does not reciprocate the proposal from 1, while she makes an unreciprocated proposal to 4. Reciprocated proposals result in the creation of links, which lead to the relation of being *neighbors*. Then, each chooses one of two actions, m or x, to coordinate with her neighbors.

Everyone sees the outcome of the game and their net payoffs on the screen, as in Figure 3d. The figure shows that participant 7 coordinates successfully on her *prescribed* action with neighbors 2 and 6, and fails to coordinate with 5. Thus, her net payoff is $18 - 8 = 10$ points. At the beginning of any subsequent period, every participant is assigned a

[14] In the experiment, the action prescribed to the majority, m, was labeled mustard, and that of the minority, x, was labeled black, to facilitate identification with the colors displayed on the screen (see Figure 3c). For consistency, I will not use the labels mustard and black in the paper, instead I will continue referring to them as actions m and x.

[15] For consistency and comparability with related models, I use the same parameters as in Goyal et al. (2020). See also Ellwardt et al. (2016).

new numeric label and a new position in the network, to preclude identification across periods.[16]

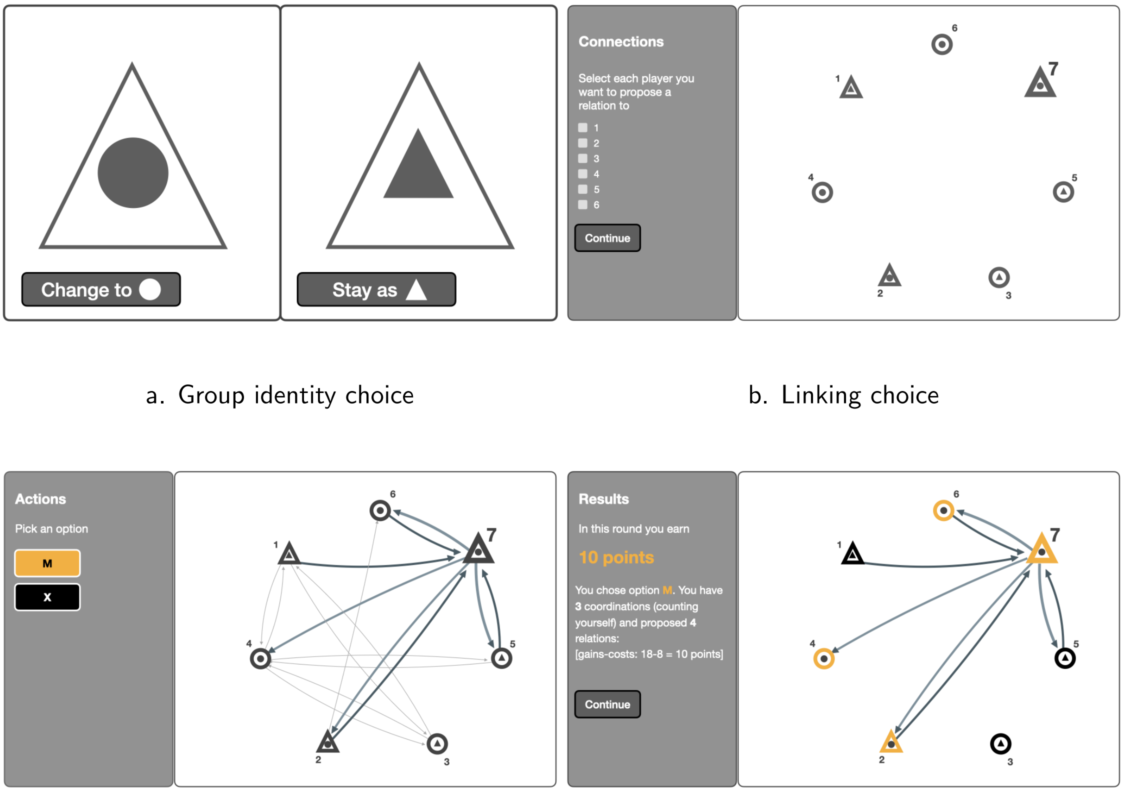
There is a benchmark version of the game with fixed identities, as in Goyal et al. (2020). I will use this game as a within-subject control in the experiment, where identities cannot be changed (see Section 3.2). If identities are fixed, the first stage in the game is the linking choice (see Figure 3b). Subsequently, after making their linking proposals and observing the resulting network, participants choose one of the two actions: m or x (see Figure 3c). Payoff parameters are the same as in the game with flexible identities: $v = 6$ is the payoff for coordinating on the same action with a connection, $r = 2$ is the loss for playing the action that opposes one's group identities, and $c = 2$ is the cost of a link proposal (see Figure 3d). However, unlike the case with identity choice, identifying markers do not change with fixed identities, so that all participants are displayed with matching internal and external symbols: (○, ●) for the majority and (△, ▲) for the minority.

3.2 Treatments

An experimental session consists of a sequence of four parts: (1) Group assignment, (2) Coordination game with *fixed* group identities, (3) Coordination game with *flexible* group identities, and (4) Other-Other allocation. Individuals know the number of parts in the experiment, but only read the instructions for a specific part once the previous one has finished. I introduce experimental variations in the game with flexible identities (PART 3), while PART 1, PART 2 and PART 4 are identical across treatments. I describe each part in the sequence below, and explain treatments in the description for PART 3.

PART 1. Group assignment: At the beginning of the experiment, individuals are randomly assigned to the majority or the minority subgroups. Members from each subgroup separately participate in a task where they are asked to vote for a name for their subgroup. To do so, members of each subgroup communicate through a chat and together must choose the same group name from a list of 5 options (see the Instructions in Appendix D). Solving this task together through free-form communication is expected to prime group identity, without having to highlight differences in group size (Eckel and Grossman 2005;

[16] The theoretical analysis focuses on the stage game. To approximate the design closer to a one-shot interaction, I randomly assign new numeric labels and new positions on the screen to everyone in each period. This prevents identification of individual players. So, individual choices are anonymous despite interactions being fixed within each group of 7. Importantly, the repetition of a stage-game equilibrium outcome is an equilibrium in the repeated game. Naturally, there may also be other non-standard equilibria. But I do not investigate these for they are of less interest to the study.



a. Group identity choice

b. Linking choice

c. Action choice

d. Results

Figure 3 Screens in the experiment.

3a: Participants choose whether to maintain their group identity or adopt the other group's identity by clicking on the corresponding button. The image above each button illustrates how they will be displayed on the screen given their choice. The internal symbol, ● or ▲, represents the chosen identity, while the external symbol, ○ or △, keeps track of the original group a participant was assigned to. 3b: Participants see their own and others' numeric labels and symbols, and choose which links to propose by ticking on the boxes on the left. 3c: Then, they observe the proposals made (outgoing arrows) and the proposals received (incoming arrows), by everyone. Own link proposals are displayed with thicker lines. Participants choose an action, mustard (m) or black (x), by clicking on the corresponding button on the left. 3d: Participants see the action chosen by everyone else, illustrated as the border-color of each node, and a summary of their payoffs in the current period. For simplicity, only the decision-maker's reference network is displayed.

Chen and Li 2009).[17] During the experiment I refrain from using *majority* and *minority* as labels, and instead use the chosen names to refer to each subgroup.

PART 2. Game with *fixed* group identities: Once the task on group assignment concludes, the 7 participants play together the coordination game without identity choice for 10 periods. This is a 2-stage game where participants choose links and then actions. The aim of PART 2 is to allow participants to gain experience in the coordination setting and to further enhance in-group identification and out-group differentiation, given identity change is not an option.

PART 3. Game with *flexible* group identities: After the 10 periods of PART 2, participants play together the coordination game with identity choice for 10 periods. This is a 3-stage game where participants choose identities, links and actions. In PART 3, it is possible for minority players to abandon their group's normative prescription and adopt the mainstream one. Identity change can maximize efficiency and reduce inequality when compared to the case in which identities are retained.[18] I assign participants into one of three main treatments: BASELINE, SANCTIONS, or IGNORANCE, which I use to asses how different barriers impact the collective behavior of the minority (see Table 1 for details). I explain each treatment as follows:

BASELINE: In this treatment there are no barriers, so that identity change is free and visible. That is, any participant can change her assigned identity at no cost, independently of the group identity other in-group members choose. Also, adopters can be clearly identified by everyone else, before the linking proposals are made, because the internal symbol is adjusted to display the chosen identity (see Figure 3b).

SANCTIONS: This treatment explores the effect of social sanctions. Anyone who changes her group identity can be distinguished from those who maintain it, as in BASELINE, but the change of identities is costly. The cost of identity change for the minority is $\delta = 6+2\cdot(2-\tau)$, where τ refers to the number of in-group peers adopting the mainstream identity (see Section 2.3).[19] This means that anyone who changes her group identity pays a fixed

[17] The group chat takes place at the beginning of the experiment but before the participants receive any information about the coordination game they will play. Because of this, there is no room for participants to chat about potential strategies for the game. The sole purpose is for participants to choose a name for their subgroup out of a 5-item list. Participants in the majority chose a name between: *Cats, Tigers, Lions, Leopards, Jaguars*, and participants in the minority chose between: *Dogs, Jackals, Coyotes, Foxes, Wolves*. The lists of names are different to further induce group identification/differentiation.

[18] Note that the aim of a sequence going from fixed to flexible identities is to evaluate the benefits of identity change in solving tensions that arise in an environment where identities cannot be abandoned. Because the interest of this sequence is contingent on its direction, I do not test for order effects by running treatments where the game with flexible identities is played before that with fixed identities.

[19] The cost for a majority player is $\delta = 6 + 2 \cdot (3 - \tau)$, given there are 4 players in total in the majority group, and so up to 3 peers can adopt the minority's group identity.

cost of 6 points, plus a variable cost that increases by 2 points for each participant in her group who maintains the original group identity. In this way, I exogenously impose social sanctions from non-adopters upon those who change their group identity.[20]

IGNORANCE: This treatment explores the effect of pluralistic ignorance. Identity change is free, as in BASELINE, but the choice is not visible to others. As such, adopters cannot be differentiated from those who persist on their group's identity. Participants can see each others' assigned group through their external symbol. However, no one can identify each others' chosen identity, because the internal symbol is eliminated from all nodes on the screen, irrespective of their identity choice.

Table 1 Experimental treatments

Description of Parts and Treatments, as well as summary of payoffs (in points) in the *no change outcome* and *global change outcome*, separate for the majority and the minority.

	Fixed identities	**Flexible identities**		
		BASELINE	SANCTIONS	IGNORANCE
Identity change				
Costly	N/A	×	✓	×
Visible identity	✓	✓	✓	×
Payoffs: No Change				
Minority	14	14	14	14
Majority	18	18	18	18
Payoffs: Global Change				
Minority	16	30	24	30
Majority	30	30	30	30

PART 4, Other-Other allocation: After the end of PART 3, every participant is asked to allocate 10 points between two anonymous receivers. There is one receiver from the minority and another from the majority, and both are drawn from the same group of 7 as the allocator. No one was allowed to allocate points to herself. The novelty of including this task after the game with flexible identities is to explore if identity change leads to positive inter-group attitudes beyond the setting for which the adoption of an identity was relevant. Specifically, I explore if by sharing a common identity participants from the majority and the minority make more equitable allocations between the two receivers.

[20] By choosing $\delta = 6$ and $\omega = 2$ as parameters, it is possible to capture the convex relation between the number of adopters and the earnings of the minority. This is made clear when comparing payoffs for identity adopters in the sequence of outcomes in Figure 1, where an individual earns 14 points in the *no change outcome*, 12 points if she becomes mainstream alone, 18 points if someone else adopts the mainstream identity as well, and 24 points in the *global change outcome* where all three adopt the mainstream group identity.

3.3 Hypotheses

I now present the hypotheses I test in the experiment for PART 2, PART 3, and PART 4.

PART 2 replicates the coordination game in Goyal et al. (2020) with fixed group identities. The main results of their work suggest that if linking is costly, participants only form links with others in their group: minority with minority and majority with majority. Consequently, everyone choose actions in line with their group identity. This result is robust to different levels of linking costs and it is also consistent with findings in Ellwardt et al. (2016). Thus, I postulate the following hypothesis for PART 1:

Hypothesis 1 *In a setting with fixed identities, minority and majority participants avoid interactions across groups and choose actions in conformity with their assigned group identity.*[21]

In PART 3 participants play the coordination game with flexible group identities. As mentioned above, groups are assigned into one of three treatments: BASELINE, SANCTIONS and IGNORANCE. In all three treatments, minority participants can make use of the identity-change opportunity to integrate and coordinate with a larger number of neighbors than in an outcome with no change of identities. Among all outcomes, the *global change outcome* is socially efficient, leads to economic mobility and reduces inequality, across treatments. Thus, I postulate the following *null* hypotheses for PART 3.

Hypothesis 2 *In a setting with flexible group identities, all participants adopt the mainstream identity and form a complete network where everyone chooses the mainstream action, resulting in the global change outcome.*

Alternatively, it may be that not all individuals in the minority become mainstream, when they are exposed to social SANCTIONS or pluralistic IGNORANCE. Exposure to such barriers makes identity change riskier than in the BASELINE. If only one minority participant were to become mainstream, the entire minority group could end up worse-off than if there was no adoption of mainstream identities at all. Specifically, for the BASELINE treatment, as shown in Figure 2, any number of minority players can become mainstream and

[21] The outcome is equivalent to the *no change outcome* in the model, where the social network is composed by two disconnected but complete components; one formed only by majority players and the other by minority players, where all players choose an action in conformity with their original group identity (see Figure 1a).

benefit from this choice.[22] For the SANCTIONS treatment, a minority participant is worse-off becoming mainstream if she is the only adopter in her group. Only when two minority participants change their group identities, each can earn more than in the *no change outcome*.[23] Notably, in the IGNORANCE treatment, if a single minority participant becomes mainstream, it is not possible for her to be individually identified as the one changing her group identity, even if everyone has correct beliefs on the number of minority adopters. This is also true in the case of a partial change with two adopters, even though the probability of being identified increases. Note that, if all three minority participants become mainstream, and beliefs are correct, it is as if there was complete information, and all players can maximize their gains. As such, the *global change outcome* is the first-best outcome and the *no change outcome* is the second best.

In summary, identity change is riskier in SANCTIONS and in IGNORANCE than in BASELINE, because both adopters and non-adopters can be worse-off in a *partial change outcome* than in a *no change outcome*. I postulate the following *alternative* hypothesis for treatments SANCTIONS and IGNORANCE in PART 3:

Hypothesis 3 *If the minority is exposed to social sanctions or pluralistic ignorance, not all players adopt the mainstream group identity. Networks segregate into two components, one choosing the mainstream identity and another opposing it, resulting in either the no change outcome or a partial change outcome.*

Finally, in PART 4 I explore how changes in group identities impact social preferences, specifically distribution preferences between the minority and the majority. This is a way to explore if the benefits from identity change extend beyond the context in which it occurs. The existing experimental evidence on similar types of allocation choices consistently shows that individuals are likely to bias their allocation in favor of the receiver from their same group (Chen and Li 2009; Cikara and Van Bavel 2014). However, in my experiment, after being exogenously assigned to a group and identity in PART 2, participants have the option to endogenously change their group identity (and consequently change their reference group) in the game with flexible identities. I use this choice to evaluate if identity

[22] If a single minority participant changes her identity, she earns 22 points per period, compared to 14 points she earns in the *no change outcome*. Naturally, if all three adopt the mainstream group identity, each minority player earns 30 points, which is more than twice the earnings of not changing identities.

[23] If only one minority player adopts, she earns 12 points for integrating with the majority, because she has to pay $\delta = 10$. Thus, it is only when two become mainstream, that each would earn 18 points and improve compared to not changing their identity, given $\delta = 8$. If all three minority participants become mainstream, each earns 24 points and achieve economic mobility, because the *global change outcome* is socially efficiency and Pareto dominant. In SANCTIONS, unless two out of three adopt, earnings from a *partial change outcome* are not higher than earnings in the *no change outcome*.

change leads majority and minority individuals to see each other as sharing a common identity, despite being originally assigned to opposing out-groups. If this were the case, the more the minority collectively adopts the mainstream identity the more equitable the allocations between the two receivers. Consequently, I postulate the following hypothesis:

Hypothesis 4 *The more the minority collective adopts the mainstream group identity the more likely majority and minority participants are to make equitable allocations between the two receivers.*

3.4 Experimental procedures

The experiment was conducted at the experimental laboratory of the University of Rosario (REBEL). A total of 336 individuals participated in the study.[24] Participants interacted through computer terminals and the experiment was programmed using oTree (Chen et al. 2016). Each session lasted on average 120 minutes, including the time used to read the instructions and to anonymously pay.

From the beginning, participants were informed that the study consisted of a sequence of four parts played within the same group of 7. Instructions were administered on the screen at the beginning of each part, so participants did not know anything specific about a subsequent part before it started. After reading the instructions, participants were presented with a set of comprehension questions and could not advance until all answers were correct. At all times, participants could click on a button on the screen and a summary of the instructions for the corresponding part would be displayed (see Appendix D).

With respect to earnings, one of the 10 periods of play was randomly selected for payment from PART 2 and one from PART 3. For the allocation choice, the decision from one of the 7 players in a group was randomly selected for payment. In the instructions for each part, participants received information about the way payment for the current part would be calculated. However, they were only informed of their actual earnings at the end of the experiment. Participants were paid using the exchange rate of 2 points = 800 Colombian Pesos (COP). On average participants earned 46,000 COP (Approx. 15.5 USD), including a 3.5 USD show-up fee. The standard conditions of anonymity and non-deception were implemented in the experiment, and no one participated in more than one session.

[24] I conducted 3 sessions with 4 groups of 7 participants for each treatments, including participants in the treatment REVEAL discussed in Section 5. All participants were undergraduate students, out of which 58% where female. See Appendix C for more details on the sample.

4 Results

In this section, I summarize the main findings of the experiment. The data from the coordination game consist of decisions made in two Parts of 10 periods. There are 12 groups of 7 participants in each of the main treatments: BASELINE, SANCTIONS and IGNORANCE (I explore the results from treatment REVEAL in section 5), resulting in 5040 observations at the individual level and 720 observations at the group level. To control for any interdependence across periods, I run random effects GLS regressions, clustering standard errors on groups. I report two-sided p-values in the text and provide all regression estimates in Appendix B.[25]

Table 2 Summary statistics in the coordination game

Average fractions (percentages) for each of the main variables. Standard deviations in parenthesis. Observations are disaggregated by majority and minority; pooled for the game with *fixed* identities and separated by treatment for the game with *flexible* identities.

	Fixed identities		**Flexible identities**					
			BASELINE		**IGNORANCE**		**SANCTIONS**	
	Major	*Minor*	*Major*	*Minor*	*Major*	Minor	*Major*	*Minor*
Identity change	n/a	n/a	0.01	0.95	0.01	0.65	0.01	0.59
			(0.09)	(0.21)	(0.08)	(0.48)	(0.08)	(0.49)
Inter-group contact	0.04		0.77		0.42		0.51	
	(0.13)		(0.33)		(0.44)		(0.43)	
Efficiency (mobility)	0.88	0.75	1.42	1.68	1.18	1.15	1.28	0.91
	(0.25)	(0.36)	(0.33)	(0.57)	(0.39)	(0.71)	(0.35)	(0.67)
In-group bias	n/a	n/a	0.36	0.22	0.39	0.37	0.31	0.48
			(0.42)	(0.37)	(0.43)	(0.38)	(0.40)	(0.45)

Table 2 reports summary statistics of the main variables, separate for the majority and the minority. *Identity change* is the fraction of participants in a subgroup changing their group identity (not applicable when identities are fixed). *Inter-group contact* reports the fraction of links formed across the majority and the minority.[26] *Efficiency* is the fraction of earnings an individual attains per period, normalized by the expected earnings in the *no change outcome*.[27] This measures economic mobility; where anything above 1 means

[25] To test the extent to which the results are robust to the dynamics within each series of repetitions, I conducted the same analysis as in the main text applied to the last 5 periods. All findings are consistent when looking at either all periods or at the second half.

[26] See Table C-3 in Appendix C.1 for a detailed summary of other connectivity variables, such as within-group connectivity and link proposals.

[27] The benefit of normalizing efficiency by the earnings predicted for the *no change outcome* is that the value is the same for the benchmark game with fixed identities and across treatments in the game with flexible identities (see Table 1). Note that the *global change outcome* is the first-best outcome, for which efficiency is 2.1 for the minority in BASELINE and IGNORANCE and 1.7 in SANCTIONS, while it is 1.7 for the majority across treatments.

earnings are higher than predicted for the second-best outcome.[28] Finally, *in-group bias* is a measure from the Other-Other allocation choice. It refers to the difference between the allocation to the in-group receiver and the out-group receiver, normalized by the total endowment.[29]

4.1 Test of theoretical assumptions

I begin by reporting how behavior in the experiment aligns with the two following assumptions discussed in the theory: (A1) that majority participants do not change their group identity, and (A2) that participants play actions in conformity with the identity they choose.

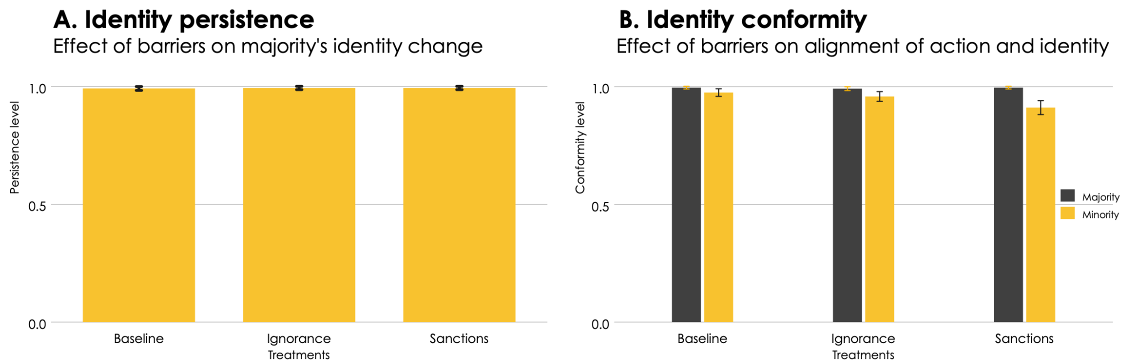


Figure 4 Test of theoretical assumptions

Panel A illustrates the average level of identity persistence of the majority (lines represent 95% confidence intervals). Panel B shows the fraction of participants who choose an action in Stage 3 that conforms to the identity they choose in Stage 1, in the minority (lighter bars) and in the majority (darker bars). In both panels choices are disaggregated by treatment.

With respect to A1, Figure 4A illustrates that when given the option to change identities, majority members chose to keep theirs 99% of the times and there are no differences between treatments (see also Table B-1 in Appendix B).

With respect to A2, Figure 4B illustrates that both majority (darker bars) and minority (lighter bars) participants are very consistent in playing the action that conforms to their chosen group identities. Participants in the majority choose their prescribed action 99% of the times and there are no differences across treatments (see also Table B-1 in Appendix B). For the minority, conformity is 98% in BASELINE, which is not different from 96% in

[28] This is illustrated as the area above the gray horizontal line in Figure 2.

[29] For ease in the exposition, I report the allocation bias under the treatment columns. That is why there is n/a for the fixed identity columns.

IGNORANCE (p = 0.168), or 91% in SANCTIONS (p = 0.170).[30] This leads to the first result:

Result 1 *When group identities are flexible the majority persist on the mainstream identity, and players from both groups consistently play actions that conform to their chosen group identity.*

4.2 Identity change, inter-group contact and economic mobility

In this section, I show how being exposed to either social SANCTIONS or pluralistic IGNORANCE impacts choice at the individual- and group-level, compared with the BASELINE without barriers. I also look at how identity choices affect inter-group contact and economic mobility.

With respect to the minority, Figure 5A shows that in the BASELINE the individual level of identity change is 95%, which drops to 65% in IGNORANCE (p = 0.007) and to 60% in SANCTIONS (p < 0.001), where the effect of SANCTIONS and IGNORANCE is not statistically distinguishable (p = 0.684). Thus, at the individual level, in line with existing findings (see e.g., Bisin et al. 2016), both social sanctions and pluralistic ignorance similarly prevent minority members from becoming mainstream.

The result, however, takes a different turn when looking at how the minority as a group chooses identities. For this, I evaluate the frequency of the *partial change outcome*, in which the minority is divided into those that become mainstream and those that oppose the majority's group identity. Figure 5B shows that minority groups exposed to social sanctions are less likely to avert the *partial change outcome* when compared to those exposed to pluralistic ignorance: about 50% of the groups in SANCTIONS are divided, which is twice more than the 25% in IGNORANCE (p = 0.051) and four times more than the 12% in BASELINE (p = 0.001). This leads to the following result:

Result 2 *Social sanctions and pluralistic ignorance reduce identity change. But sanctions divide the minority most.*

[30] In addition, I test whether alignment of the action played and the group identity differs when identities cannot be changed. I found that misalignment is more likely when group identities are fixed than when they are chosen, and the difference is mildly significant ($89\% < 95\%, p = 0.029$). I pool the data across treatments, given the conjecture is on alignment conditional on identity choice, not on the barriers faced. This speaks to the two characterizations of equilibria discussed in the theory section: identities are enforceable on actions or non-enforceable. The result suggests that a model assuming group identities as enforceable may be more suitable to capture settings where identities can be chosen, while the assumption of non-enforceability may be better to capture settings where identities are fixed and exogenously imposed, as in the case of socialization (see e.g., Bisin and Verdier 2001).

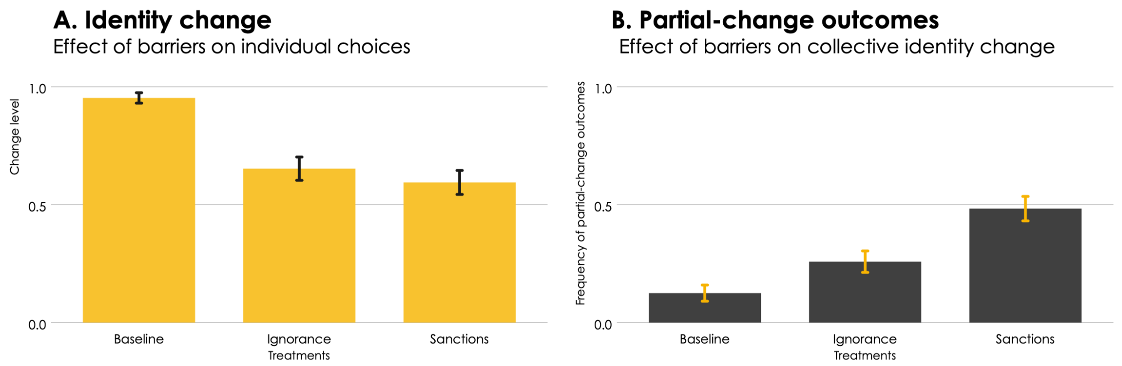


Figure 5 Individual and collective identity change

Panel A illustrates the average level of identity change of the minority disaggregated by treatment. Lines represent 95% confidence intervals. Panel B illustrates the frequency of the *partial change outcome*, in which the minority is divided. Choices in both panels are disaggregated by treatment.

The natural question to follow is how the difference in identity-change patterns and collective choices impact contact between the minority and the majority. Focusing on how being in a divided minority impacts identity adopters' chances to integrate with the majority. Figure 6A illustrates inter-group connectivity. For the fixed-identity stage, as predicted, connectivity is close to null, such that participants create incomplete networks where they segregate by majority and minority: 96% of the links across groups are missing.[31] Compared to fixed identities, inter-group contact significantly increases to 77% ($p < 0.001$) in the BASELINE. The increment is evident but not as large in SANCTIONS ($44\%, p < 0.001$) or in IGNORANCE ($51\%, p = 0.004$).

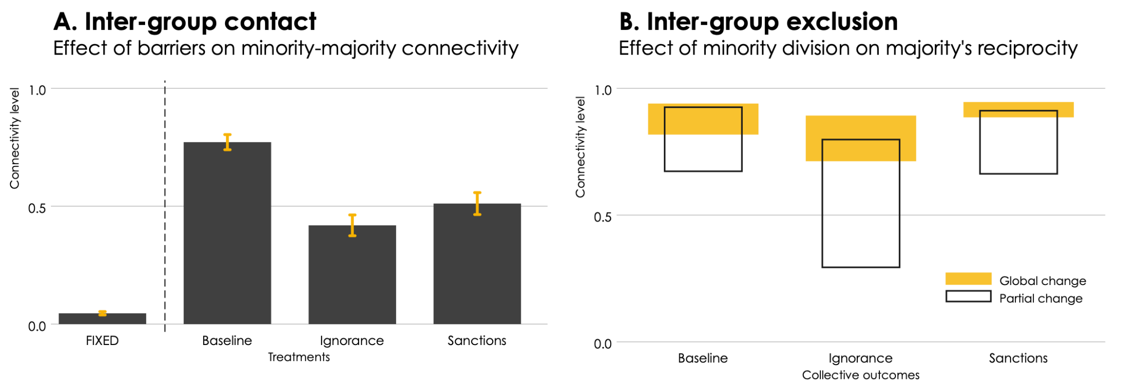


Figure 6 Inter-group contact and exclusion

Panel A illustrates the average level of connectivity across groups (minority with majority). Panel B illustrates the level of inter-group contact of minority individuals that adopt the mainstream identities, compares outcomes of global or partial identity change.

Next, I focus exclusively on the minority members that abandon their group identity and

[31] Participants can form up to 12 links across groups. So, 4% of inter-group connectivity indicates that on average 0.48 links were formed. This is practically a segregated network.

become mainstream. In equilibrium, any identity adopter should propose a connection to each of the four majority members and all proposals should be reciprocated. This, independently of whether the other minority members also become mainstream or oppose the majority's group identity. However, I find that reciprocity is significantly lower for a minority adopter when her group is divided, even if her chosen identity is visible. Figure 6B shows the level of inter-group contact of minority adopters, separate for partial and global change outcomes. The top of each range bar shows the level of proposals made to the majority (1 means she made 4 proposals), the bottom is the level of formed links, and a bar's height is the fraction of unreciprocated proposals. When the minority acts as a collective unit and all members change their identities, individuals form on average 82% of links in BASELINE, 71% in IGNORANCE ($p = 0.194$), and 88% in SANCTIONS ($p = 0.608$), indicating that unity facilitates individual connectivity and treatment differences disappear. However, when the minority is fractured, average inter-group connectivity is 67% in BASELINE, 29% in IGNORANCE ($p < 0.001$) and 66% in SANCTIONS ($p = 0.590$), and connectivity is significantly lower with partial than global identity change ($p < 0.001$). This suggests that integration is harder for adopters when others in the minority oppose the majority. This is summarized in the following result:

Result 3 *Identity change promotes inter-group contact. But in-group division reduces link reciprocity from the majority towards identity changers in the minority.*

To conclude this section I explore efficiency and economic mobility. I test if the minority attains higher earnings when identities are flexible rather than fixed. Figure7A shows the level of attained efficiency with fixed and flexible identities. In fixed, the minority gravitates towards the segregated (second best) outcome, attaining on average 76% of the predicted earnings. Identity change leads to an increase in efficiency for the minority in most but not all treatments. In BASELINE (line with circles) the minority more than doubles its gains due to identity change, from 0.76 to 1.68 ($p < 0.001$). In IGNORANCE (line with triangles), gains increased by 53%, from 0.75 to 1.15 ($p < 0.001$), resulting in economic mobility for the minority despite individuals being unable to identify the group identity chosen by others. However, in SANCTIONS (line with X) efficiency increases by 21%, to 0.91 and does not lead to economic mobility ($p = 0.073$).

I further explore why minority members reach economic mobility when exposed to pluralistic ignorance but not to social sanctions, despite average identity change and average inter-group connectivity being indistinguishable. Figure 7B shows the distribution of earnings of minority separate for adopters (filled bars) and opposers (empty bars). Notably, minority members who do not change their identities are significantly more likely to reach

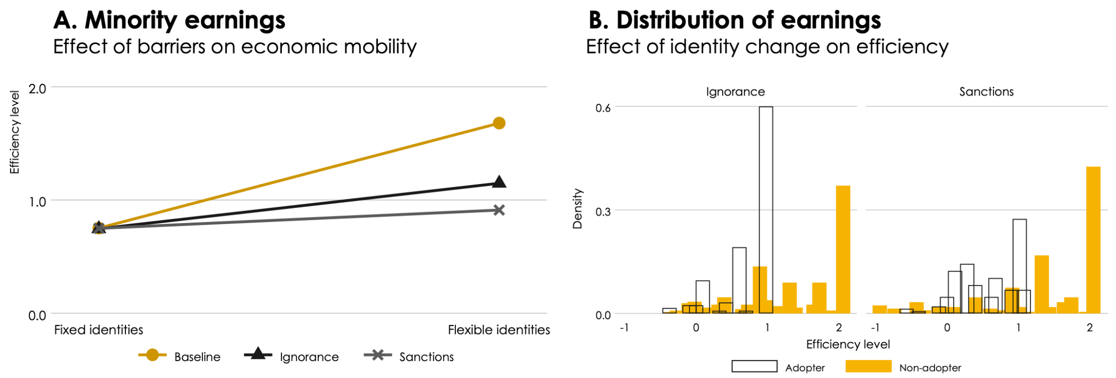


Figure 7 Efficiency, economic mobility and partial change outcomes.

Panel A illustrates the level of efficiency comparing the game with fixed identities and with flexible identities, for the minority. Panel B illustrates the fraction of groups in which the minority displays partial identity change, where not all three members choose the same group identity, disaggregated by adopters and non-adopters. In both panels, choices are disaggregated by treatment.

efficiency of 1, which is the highest level of earnings attainable, in IGNORANCE than in SANCTIONS. As such, the salience of partial change outcomes hurts the minority, especially those minority individuals who are deterred from identity change because of the potential sanctions, but end up in a social network that shrinks and prevents them from increasing their gains. This is summarized in the following result:

Result 4 *Identity change can lead to economic mobility (compared to no change). But division within the minority prevents it.*

In the next section, I look at whether changing identities as a collective unit in the coordination game has positive spillovers on social (distributive) preferences towards the out-group in an independent setting.

4.3 Identity change and social preferences

To conclude, I measure differences in the allocation participants make to a receiver from their in-group and to a receiver from their out-group, at the end of the experiment. The difference between the allocations represents the magnitude of the *in-group bias*. Following Hypothesis 4, equitable allocations (no bias) are expected to occur the more a participant perceives the out-group receiver as someone she identifies with. Figure 8A reports the magnitude of in-group bias for the minority.

The findings support the predictions in Hypothesis 4, that allocations to the out-group are expected to be positively related to the minority collectively adopting the mainstream group identity as a unit. Allocations in BASELINE are biased by 22%, the bias in IGNORANCE is higher but not significantly different to BASELINE (37%, p = 0.101), while the bias in SANCTIONS is significantly larger than in BASELINE (48%, p = 0.012). That is, the more minority participants change identities collectively as a unit, the more equitable are their preferences towards the out-group. Figure 8B illustrates the magnitude of in-group bias for the majority, and shows that there are no differences in bias in BASELINE (36%) with respect to IGNORANCE (39%, p = 0.728) or to SANCTIONS (31%, p = 0.528). So that even when the minority integrates and plays the coordination game the way the majority prefers, the majority seems unaffected by it. This leads to the following finding:

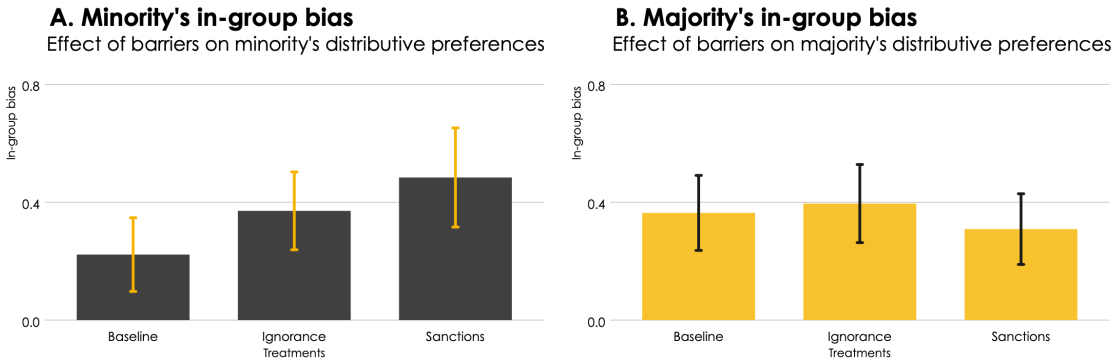


Figure 8 In-group bias for the minority and the majority, by treatment.
The graph shows the magnitude of allocation bias towards in-group members for the minority (Panel A) and the majority (Panel B). In both panels, choices are disaggregated by treatment.

Result 5 *Collective identity change can transform social (distributive) preferences for the minority, but not for the passive majority.*

In conclusion, the findings are not fully supportive of the predictions in the *null* Hypothesis 2 but more of the alternative Hypothesis 3. The findings highlight that social SANCTIONS divide the minority into predominantly displaying partial change outcomes, preventing inter-group contact and economic mobility in the coordination game. However, when the minority acts as a collective unit, as they do despite being exposed to pluralistic IGNORANCE, identity change and inter-group contact are instrumental in benefiting minority members. Naturally, the benefits of identity change are even more prominent when there are no barriers, as in the BASELINE, where the emerging outcome is almost exclusively the efficient *global change outcome*. The consequences of these outcomes transcend the coordination setting and give partial support to Hypothesis 4, in that identity change transforms the behavior of the minority towards the majority. However, majority members, who belong to the group others aim to integrate into, appear to be unaffected by the identity choices of the minority.

5 Revealing identity choices

In this section, I report results on treatment REVEAL, where I allow individual participants to endogenously select which barriers to face. Participants can freely change identities but their choice is not visible to others (as in IGNORANCE). Then, they can *reveal* their chosen identity, which exposes them to sanctions that increase on the number of opposers in their in-group (as in SANCTIONS). This is implemented for the minority through a *revelation cost* that has the same form as that in SANCTIONS: $\delta_{\text{REVEAL}} = 6 + 2(2 - \tau)$, where τ is the number of in-group members who adopt the other group's identity.[32]

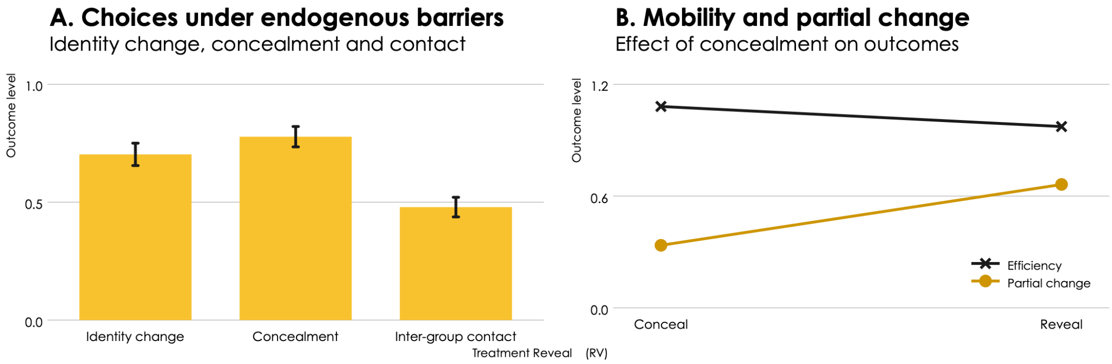


Figure 9 Main choices in treatment Reveal (RV).

Panel A illustrates the average level of identity change, identity concealment (choosing not to reveal), and inter-group contact. Lines represent 95% confidence intervals. Panel B illustrates the level of efficiency (line with X) and the frequency of partial-change outcomes players are in, conditional on revealing or concealing their chosen identities.

The main question with endogenous barriers is whether the choice individuals make of which barrier to expose themselves to leads to outcomes that reflect similar patterns to those observed with exogenous barriers.[33] Figure 9A shows that minority participants change identities 70% of the times. This is not distinguishable from either IGNORANCE ($p = 0.726$) or SANCTIONS ($p = 0.393$). Once players choose their identities less than a quarter (22%) revealed it. Instead, the larger majority (78%) kept their identities concealed. The consequent level of inter-group contact is 48%, which is significantly lower for participants concealing than revealing their group identity ($39\% < 77\%, p < 0.001$).

The minority attains a significantly higher level of efficiency in the game with flexible identities than in the game with fixed identities: $1.06 > 0.80$ ($p = 0.047$). Specifically, Figure 9B shows that efficiency is slightly higher, although not significantly different when

[32] For the majority the revelation cost is $\delta_{\text{REVEAL}} = 6 + 2(3 - \tau)$ given there are 4 participants in the majority group.

[33] Consistent with the main treatments, behavior in REVEAL supports assumptions A1 and A2 (see regressions in Appendix B).

identities are concealed (1.1 versus $0.97, p = 0.196$). But, the most notable outcome is that individuals revealing their identities end up in a*partial change outcome* significantly more often than those who conceal it.[34] In fact, when all three minority participants change group identities and become mainstream, there is significantly less revelation of chosen identities than when there is partial change ($p < 0.001$), which is indicative that being exposed to social sanctions is likely to be related to group cohesion division. Thus suggesting that by endogenously avoiding social sanctions, the minority is able to act cohesively and attain economic mobility. This leads to the final result:

Result 6 *When there is a choice between barriers, minority participants are less likely to reveal their chosen identities if this exposes them to social sanctions. But, when they do, identity revelation divides the minority the most.*

6 Conclusions

In this paper, I have studied how changes in group identities impact the collective behavior of groups, focusing on the interaction between a mainstream majority and an oppositional minority. I summarize the main results of the study below and discuss their implications.

An initial finding shows that having a dominant group creates a focal point. That is, it would be equally efficient (at least in the BASELINE) for the majority to adopt the group identity of the minority or for the minority to become mainstream. However, majority players do not even attempt to take the group identity of the minority. This shapes the outcomes in the social coordination study, as well as shows that by being focal (and possibly passive) majority members fail to change how the perceive the minority, despite the efforts of minority members to integrate. So that even in such a minimal-group setting, sharing a common identity does not impact social groups in the same way.

Unlike the majority, minority members use identity change extensively when given the chance. However, success from adopting the mainstream group identity is contingent on the group acting collectively as a unit. Because, when a minority group is divided into those who become mainstream and those who persist in opposing the majority's group identity, everyone is worse-off. The experimental findings show that when the minority group is exposed to social sanctions, the minority is divided the most. Half of the times groups end up fractured. On the other hand, when the minority is exposed to plural-

[34] The frequency of partial change outcomes is 40%, which is lower but neither significantly different from SANCTIONS ($p = 0.606$) nor from IGNORANCE ($p = 0.209$).

istic ignorance, where there is no objective information about the identities participants choose, the minority acts in unity and everyone in the group adopts one same identity three-quarters of the time.

This is a central finding that complements our understanding of barriers to identity change. Both social sanctions and pluralistic ignorance reduce identity change, but the impact they have on the collective as a whole is more detrimental with social sanctions, because it erodes group unity. I further explore this finding in the treatment REVEAL, where individuals have the option to keep their chosen identity concealed (and avoid sanctions) or to reveal it. The results complement the previous findings and show that individuals are much more prone to face uncertainty (by concealing their chosen identities) than to risk being sanctioned. On top of that, groups that keep their identities concealed remain united. That is, are more likely to choose the same identity as a collective, while making the identity visible is strongly correlated with a divided minority.

A notable effect of a divided minority is how it reduces the chances of better integration for those who actually change their group identities and become mainstream. This suggests that majority players not only look at those who share their group identity, but are also *distracted* but those in the minority who oppose them. Even when the chosen identity are visible, identity adopters' chances to integrate are hurt by the presence of minority members opposing the mainstream. In addition, division within the minority biases distributive preferences towards only in-group others. However, being in a united minority that integrates well with the majority changes social preferences, so that individuals act more equitable towards the in-group and the out-group. Unfortunately, this effect is one-sided, as mentioned above, and the passive majority appears unaffected by the efforts the minority puts to integrate.

In essence, changes in group identity are challenging, because minority groups tend to be exposed to barriers that not only reduce but outright prevent identity change. When groups manage to change identities as a unit, they integrate well. This leads to higher efficiency and to a change in social preferences, that is beneficial to the out-group. However, when a group is divided, which is most often if exposed to social sanctions, even those who adopt the mainstream identity are hurt. Integration is harder and social preferences appear unaffected. This provides useful insights for policy makers, who aim to help social groups integrate, by highlighting the trade-offs that some barriers bring. My work suggests that if groups have an internal process of sanctions to those who change identities, such sanctions will be most detrimental not because they prevent identity change but because they divide the social group. A possible avenue to explore, is how to minimize the effects of sanctions, by providing subsidies that help cover the trade-offs identity adopters face. For instance, by making the mainstream majority aware that the better the minority integrates not only benefits them directly in the context for which the normative prescription associated with a group identity is relevant, but also may change the views the minority has towards them, which will have positive spillovers into other contexts.

A limitation in my study is that I explore group identities that are induced by differences in payoffs. It is worth pursuing in further research, how the findings of this study inform settings with real identities. Despite being able to capture the strong interdependence that ties identity persistence to the choices on one's peers in the network, the individual attachment to a specific identity may be stronger when that identity is a marker of one's social group. As such, it is possible that the levels of identity change are more modest in the field.

References

Akerlof, G. (1997). Social distance and social decisions. Econometrica, 65:1005–1027.

Akerlof, G. and Kranton, R. (2000). Identity and economics. Quarterly Journal of Economics, 115:1–39.

Akerlof, G. and Kranton, R. (2002). Identity and schooling: some lessons for the economics of education. Journal of Economics Literature, 40:1167–1201.

Alesina, A. and La Ferrara, E. (2002). Who trusts others? Journal of Public Economics, 85:207–234.

Andreoni, J., Nikiforakis, N., and Siegenthaler, S. (2021). Predicting social tipping and norm change in controlled experiments. Proceedings of the National Academy of Sciences, x:1–1.

Austen-Smith, D. and Fryer, R. G. (2005). An economic analysis of "acting white". Quarterly Journal of Economics, 120:551–583.

Baronchelli, A. (2018). The emergence of consensus: A primer. Royal Society Open Science, 5:172–189.

Barr, A., Lane, T., and Nosenzo, D. (2018). On the social inappropriateness of discrimination. Journal of Public Economics, 164:153–164.

Battu, H. and Zenou, Y. (2010). Oppositional identities and employment for ethnic minorities: Evidence from england. The Economic Journal, 120:F52–F71.

Bernheim, B. (1994). A theory of conformity. Journal of Political Economy, 102:841–877.

Bicchieri, C. (2006). The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press.

Bicchieri, C. (2016). Norms in the wild. How to diagnose, measure and change social norms. Oxford University Press.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Formation and persistence of oppositional identities. European Economic Review, 55:1046–1071.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2016). Bend it like beckham: Ethnic identity and integration. European Economic Review, 90:146–164.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. Journal of Economic Theory, 97:298–319.

Boyd, R. and Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology, 13:171–195.

Boyd, R. and Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. Journal of Theoretical Biology, 215:287–296.

Candelo, N., Croson, R. T., and Li, S. X. (2017). Identity and social exclusion: an experiment with hispanic immigrants in the u.s. Experimental Economics, 20:460–480.

Chandra, K. (2012). Constructivist theories of ethnic politics. Oxford University Press.

Charness, G., Rigotti, L., and Rustichini, A. (2007). Individual behavior and group membership. American Economic Review, 97:1340–1352.

Chen, D., Schonger, M., and Wickens, C. (2016). otree-an open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9:88–97.

Chen, Y. and Li, S. X. (2009). Group identity and social preferences. American Economic Review, 99:431–457.

Choi, D. D., Poertner, M., and Sambanis, N. (2019). Parochialism, social norms, and discrimination against immigrants. Proceedings of the National Academy of Science, 116:16274–16279.

Choi, D. D., Poertner, M., and Sambanis, N. (2021). ”linguistic assimilation does not reduce discrimination against immigrants: Evidence from a field experiment in germany. Journal of Experimental Political Science, pages 1–14.

Cikara, M. and Van Bavel, J. J. (2014). The neuroscience of intergroup relations: An integrative review. Perspectives on Psychological Science, 9:245–274.

Dasgupta, P. and Goyal, S. (2019). Narrow identities. Journal of Institutional and Theoretical Economics, 175:395–419.

Eckel, C. C. and Grossman, P. J. (2005). Managing diversity by creating team identity. Journal of Economic Behavior & Organization, 58:371–392.

Efferson, C., Lalive, R., and Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. Science, 321:1844–1849.

Ellwardt, L., Hernández, P., Martínez-Canovas, G., and Muñoz-Herrera, M. (2016). Conflict and segregation in networks: An experiment on the interplay between individual preferences and social influence. Dynamic and Games, 3:191–216.

Farrell, J. and Saloner, G. (1985). Standardization, compatibility, and innovation. The RAND Journal of Economics, 16:70–83.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. Nature Human Behaviour, 2:458–468.

Fryer, R. G. and Torelli, P. (2010). An empirical analysis of “acting white”. Journal of Public Economics, 94:380–396.

Gachter, S., Gerhands, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. European Economic Review, 97:72–86.

Gaertner, S. and Dovidio, J. (2000). Reducing intergroup bias: The common ingroup identity model. Psychology Press.

Goyal, S., Hernandez, P., Moisan, F., Martinez, G., Muñoz-Herrera, M., and Sanchez, A. (2020). Integration and diversity. Experimental Economics, XX:1–60.

Henrich, J. and Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. Journal of Theoretical Biology, 208:79–89.

Hernandez, P., Muñoz-Herrera, M., and Sanchez, A. (2013). Heterogeneous network games: Conflicting preferences. Games and Economic Behavior, 79:56–66.

Hoff, K. and Pandey, P. (2006). Discrimination, social identity, and durable inequalities. American Economic Association: Papers and Proceedings, 96:206–211.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. Journal of economic theory, 71:44–74.

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. European Economic Review, 90:375–402.

Lees, J. and Cikara, M. (2020). Inaccurate group meta-perceptions drive negative outgroup attributions in competitive contexts. Nature Human Behavior, 4:279–286.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. American Economic Review, x:1–32.

McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27:415–444.

Mousa, S. (2020). Building social cohesion between christians and muslims through soccer in post-isis iraq. Science, 369:866–870.

Opp, K.-D. and Gern, C. (1993). Dissident groups, personal networks, and spontaneous cooperation: The east german revolution of 1989. American Sociological Review, 58:659–680.

Prentice, D. and Miller, D. (1993). Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. Journal of Personality and Social Psychology, 64:243–256.

Riedl, A., Rohde, I., and Strobel, M. (2016). Efficient coordination in weakest-link games. The Review of Economic Studies, 83:737–767.

Saleh, M. and Tirole, J. (2021). Taxing identity: Theory and evidence from early islam. Econometrica, 13705:1–80.

Schank, R. L. (1932). A study of community and its group institutions conceives of as behavior of individuals. Psychological monographs, 43:1–133.

Schelling, T. (1960). The Strategy of Conflict. Harvard University Press.

Schram, A. and Charness, G. (2015). Inducing social norms in laboratory allocation choices. Management Science, 61:1531–1546.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. In Austin, W. G. and Worchel, S., editors, The Social Psychology of Intergroup Relations, pages 33–47, Monterey, CA: Brooks/Cole.

Xiao, E. (2013). Profit seeking punishment corrupts norm obedience. Games and Economic Behavior, 77:321–344.

Xiao, E. and Bicchieri, C. (2012). Words or deeds: Choose what to know about others. Synthese, 187:49–63.

Online Appendix: The cost of social division

Manuel Muñoz-Herrera[35]

Saturday 16th October, 2021

[35] Social Science Division, New York University Abu Dhabi (e-mail: manumunoz@nyu.edu, website: manumunoz.me).

A Proofs and Propositions

A.1 Proof of Proposition 2

Proof. Let y and z be the number of players choosing action x in n_{m} and n_{x}, respectively. The sum of individual payoffs is

$$W(y,z) = (n-y-z)(v(n_{\mathsf{m}}-y)+(v-r)(n_{\mathsf{x}}-z)) + (y+z)((v-r)y+vz) \tag{A-1}$$

For fixed z, social welfare is decreasing in y if $y < y^*$ and increasing in y for $y > y^*$, where

$$y^* = \frac{(v-r)(n_{\mathsf{x}}-2z)+v(n_{\mathsf{m}}-2z)+v(n)}{2(2v-r)}. \tag{A-2}$$

Similarly, for any y, social welfare is decreasing in z if $z < z^*$, and increasing in z for $z > z^*$, where

$$z^* = \frac{v(n_{\mathsf{m}}-2y)+(v-r)(n_{\mathsf{x}}-2y)+(v-r)n}{2(2v-r)} \tag{A-3}$$

Since $0 \le y \le n_{\mathsf{m}}$ and $0 \le z \le n_{\mathsf{x}}$, it follows that $W(y,z)$ is maximized for some $y \in \{0, n_{\mathsf{m}}\}$ and some $z \in \{0, n_{\mathsf{x}}\}$. Note that $W(0,n_{\mathsf{x}}) = v(n_{\mathsf{m}}^2 + n_{\mathsf{x}}^2)$, and $W(n_{\mathsf{m}},0) = \beta(n_{\mathsf{m}}^2 + n_{\mathsf{x}}^2)$, which directly implies that $W(0,n_{\mathsf{x}}) > W(n_{\mathsf{m}},0)$ (because $r > 0$). Furthermore, since $W(0,0) = n(vn_{\mathsf{m}} + (v-r)n_{\mathsf{x}})$, then $W(0,0) > W(0,n_{\mathsf{x}})$ if and only if

$$\frac{n_{\mathsf{m}}}{n_{\mathsf{x}}} > \frac{r}{2v-r} \tag{A-4}$$

This inequality holds whenever $n_{\mathsf{m}} > n_{\mathsf{x}}$.

Similarly, since $W(n_{\mathsf{m}},n_{\mathsf{x}}) = n((v-r)n_{\mathsf{m}} + vn_{\mathsf{x}})$, then $W(n_{\mathsf{m}},n_{\mathsf{x}}) > W(0,n_{\mathsf{x}})$ if and only if

$$\frac{n_{\mathsf{x}}}{n_{\mathsf{m}}} > \frac{r}{2v-r} \tag{A-5}$$

This inequality holds whenever $n_{\mathsf{x}} > n_{\mathsf{m}}$. Furthermore, note that equations A-2 and A-3 hold for $n_{\mathsf{m}} = n_{\mathsf{x}}$ as long as $v > r$. To summarize, there is always either $W(0,0) > W(0,n_{\mathsf{x}})$ or $W(n_{\mathsf{m}},n_{\mathsf{x}}) > W(0,n_{\mathsf{x}})$ as long as $n_{\mathsf{m}} \neq n_{\mathsf{x}}$ or $v > r$.

Now, consider the case where $y = n_{\mathsf{m}}$ and $z = n_{\mathsf{x}}$: this implies that $y + z = n$. Since

$r > 0$, it can be shown that $W(0,0) > W(n_{\mathsf{m}}, n_{\mathsf{x}})$ so long as $n_{\mathsf{m}} > n_{\mathsf{x}}$. Moreover, $W(0,0) < W(n_{\mathsf{m}}, n_{\mathsf{x}})$ holds as long as $n_{\mathsf{m}} < n_{\mathsf{x}}$. Finally, $W(0,0) = W(n_{\mathsf{m}}, n_{\mathsf{x}})$ if $n_{\mathsf{m}} = n_{\mathsf{x}}$.

This shows that in a setting with fixed norms social welfare is maximized when the network is complete and all players choose the mainstream action: m.

Consider now the setting with flexible norms. There are two different symmetric scenarios: (i) a case in which all minority players maintain their initial norm, $\bar{\theta}_i = \theta_i$, and (ii) a case where all players choose the same norm, because all players in the minority change their initial norm to adopt the mainstream one.

The results so far have shown the efficient outcome in case (i). Compare this to the setting in which all players in the minority change their norm. Thus, $\bar{\theta}_i = \bar{\theta}_j = \mathsf{m}, \forall i, j \in N$. Thus, if the cost of norm change is $\delta = 0$ then: $W(0,0) = n(vn_{\mathsf{m}} + (v-r)n_{\mathsf{x}}) < W(n_{\theta_i=\theta_j=\mathsf{m}}) = vn^2$ which is always true given that $r > 0$. This shows that if norms are flexible, it is efficient for all players to choose the norm of the majority, as well as consistently choose the action prescribed to that norm. This outcome is better than a combination of norms, and thus it is more efficient than the case where only a subset of the minority changes its norm while the rest maintains it.

Finally, consider the case in which the assimilation cost is positive: $W(0,0) = n(vn_{\mathsf{m}} + (v-r)n_{\mathsf{x}}) < W(n_{\theta_i=\theta_j=\mathsf{m}}) = vn^2 - \delta n_{\mathsf{m}}$ which is true as long as:

$$\delta < r \cdot n \tag{A-6}$$

This simply states that as long as the cost of assimilation is not too high, the efficient outcome is that in which all players choose the same norm and the action that is prescribed with the chosen norm, the *global change outcome*. If norm change is costly the chosen norm/action is that of the majority, given that fewer players pay the norm change cost. This completes the proof. □

A.2 Proof of Proposition 3

Proof. Proposition 3 looks at the role of costs to norm change. If a minority player abandons her group norm to become mainstream her utility is:

$$u_i = v(n_{\mathsf{m}} + \tau + 1) - c(n_{\mathsf{m}} + \tau) - \delta$$

where τ is the number of minority players who also change their social norm and become mainstream, while $n_{\mathsf{x}} - \tau - 1$ is the number of minority players persisting on their group's social norm. I look at three cases. First, the *global change outcome*, where all minority players become mainstream. This leads to $\tau = n_{\mathsf{x}} - 1$ and $n_{\mathsf{m}} + \tau + 1 = n$. In this case, it is a best response for i to change her social norm and become mainstream, if:

$$\delta < vn - c(n-1) \tag{A-7}$$

The second case is a *partial change outcome* with full opposition, where only player i becomes mainstream, while everyone else in the minority persist on their group norm. This leads to $\tau = 0$. Thus, i only links to the set of majority players n_{m}. In this case, it is a best response for i to change her social norm and become mainstream, if:

$$\delta < v(n_{\mathsf{m}} + 1) - c(n_{\mathsf{m}}) \tag{A-8}$$

The third case is a *partial change outcome* with some opposition, where a subset of minority players become mainstream. This leads to $0 < \tau < n_{\mathsf{x}}$. Naturally, if δ is as in Equation A.2, player i can change her norm independently of the size of τ. However, if δ is higher, then it is a best response for i to change her social norm and become mainstream, if:

$$v(n_{\mathsf{m}} + 1) - c(n_{\mathsf{m}}) < \delta < (v - c)(n_{\mathsf{m}} - \tau) + v \tag{A-9}$$

This completes the proof. □

A.3 Proof of Corollary 1

Proof. If the cost to change norms is $\delta + \omega(n_{\mathsf{x}} - \tau)$, the utility of player i for changing her social norm is:

$$u_i = v(n_{\mathsf{m}} + \tau + 1) - c(n_{\mathsf{m}} + \tau) - (\delta + \omega(n_{\mathsf{x}} - \tau))$$

where $n_{\mathsf{x}} = n - n_{\mathsf{m}}$. It is a best response for i to change her social norm and become mainstream, if:

$$\tau > \frac{\delta + \omega n - v}{v + \omega - c} - n_{\mathsf{m}} \tag{A-10}$$

so that the threshold necessary for i to become mainstream is smaller the higher the size of the majority (n_{m}), also smaller the larger the gains from choosing the action in conformity with the chosen norm (v), and larger the size of the fixed cost δ and of the rate at which the variable cost increases (ω). □

A.4 Proof of Assumption A1

Proof. If the majority players can also change social norms, the intuition of the pairwise stable equilibria from Proposition 1 still holds, although other outcomes arise. This is described in the following corollary:

Corollary 2 *The pair $(\bar{\theta}^*, \bar{g}^*)$ is pairwise stable if one of the following obtains:*

(i) All players adopt the same social norm, so that $\bar{\theta}_i^* = \{\mathsf{m}, \mathsf{x}\}$ for all $i \in N$ and $\bar{g}^*$ is a complete network.

(ii) The network $\bar{g}^*$ is divided into two complete components, $Q_{\mathsf{m}} = \bar{N}_{\mathsf{m}}$ and $Q_{\mathsf{x}} = \bar{N}_{\mathsf{x}}$, where every player in component Q_{m} follows norm m, while every player in Q_{x} follows norm x.

The proof for Corollary 2 is the same as that for 1 and can be found in the main text. Importantly, Corollary 2 implies that the *global change outcome* can take two forms. Either all players become mainstream, so that the minority changes norms but the majority does not, or all players oppose the mainstream norm, so that the majority changes norms and the minority does not. The second item, states that the *no change outcome* where no one changes norms remains a pairwise stable. It also states that a *partial change outcome* can take multiple forms. Not only those in which a fraction of the minority changes norms while no one in the majority does, but also some in which a fraction in each subgroup changes norms, and also one where all players change norms. □

A.5 Proof of Assumption A2

Proof. If social norms are not binding, the game has three stages. I solve backwards starting with the choice of actions in a given network where players have also decided on their social norm. Let τ refer to the number of neighbors choosing the action that conforms to i's social norm. That is, if i's social norm is $\overline{\theta}_i = m$ then the conforming action is $a_i = m$, and the opposite when the chosen norm is x. This means that player i's payoff from choosing in conformity with her chosen norm is $v(\tau + 1)$ and from choosing the non-conforming action is $(v - r)(n - \tau + 1)$. So i is strictly better off choosing in conformity with her chosen social norm if and only if

$$v(\tau + 1) > (v - r)(n - \tau + 1) \tag{A-11}$$

This inequality can be rewritten as

$$\tau > \frac{v - r}{2v - r}n - \frac{r}{2v - r} \tag{A-12}$$

so that a player is better off selecting in conformity with her chosen norm if and only if the proportion of her neighbors selecting the same action is sufficiently large.

From this, it follows that the pairwise stable networks are those in which the action profile is an equilibrium, and where players are linking as in Proposition 1. That is, no player connects to someone choosing a different action. And, only networks where all players choosing the same action are connected are pairwise stable.

Finally, given norm choice is non-binding and there are no direct benefits from choosing one norm or the other in Stage 1 (naturally there are at the end of the game), the set of equilbria is such that: (i) there is a complete network where all players choose the same action, similarly to the *global change outcome*, but where players may or may not choose in conformity with their social norm, (ii) there is a segregated network where all players maintain their social norm and choose in conformity, as in the *no change outcome*, and (iii) there are segregated networks where a subset of players choose action m and the other subset chooses action x. In both cases, players may also be choosing the corresponding social norm or not. Naturally, the case where the action conforms to the social norm Pareto dominates the same action profile and network but where the vector of social norms is such that a player is not matching her social norm to the action in stage 3, give that she is earning v in the former and $v - r$ in the latter.

For more details on the equilibria for the 3-stage game see Goyal et al. (2020). □

B Regression tables

The data in the coordination game consists of the decisions made by 336 individuals who interact over two blocks of 10 periods in groups of 7 players in four treatments. The tables below report random effects GLS regressions with standard errors clustered on groups for (i) the setting with fixed norms, (ii) the setting with flexible norms, and (iii) differences between the two settings.

Table B-1 reports regressions testing differences in the level of norm change of the minority and the majority across treatments. In all regressions the independent variables are dummies for each treatment, for which the BASELINE treatment is the omitted category. Moreover, I use group random effects in all regressions. The first dependent variable is the level of norm change, representing the case in which a player changes her social norm, within the minority in column **I** and within the majority in column **II**. The subsequent dependent variable is the level of conformity, representing the case in which a player chooses the action that conforms to her chosen norm, for the minority in column **III** and the majority in column **IV**. Finally, the last dependent variable is the level of economic mobility, expressed as the fraction of earnings normalized by the prediction for the *no change outcome*, for the minority in column **IV** and for the majority in column **V**.

Table B-1 The effect of barriers on norm change and conformity

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the level of norm change of the minority in column I and of the majority in column II, and the level of conformity of the minority in column III and of the majority in column IV. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	Norm change		**Conformity**	
	Minor	*Major*	*Minor*	*Major*
	I	**II**	**III**	**IV**
IGNORANCE	−0.300***	−0.002	−0.017	−0.004
	(0.112)	(0.006)	(0.012)	(0.007)
SANCTIONS	−0.358***	−0.002	−0.064	0.000
	(0.091)	(0.006)	(0.047)	(0.004)
REVEAL	−0.250***	0.010	−0.031	−0.006
	(0.090)	(0.009)	(0.024)	(0.007)
Constant	0.953***	0.008*	0.975***	0.996***
	(0.013)	(0.005)	(0.007)	(0.003)
χ^2	29.39***	2.16	4.54	1.22
# Obs.	1440	1920	1440	1920
# Groups	48	48	48	48

Table B-2 reports regressions testing differences in group variables across treatments. In all regressions the independent variables are dummies for each treatment, for which the

BASELINE treatment is the omitted category. Moreover, I use group random effects in all regressions. The first dependent variable in column **I** is the level of partial norm change. This represents the case in which not all three minority players in a group chose the same social norm. The subsequent dependent variable is the level of inter-group contact, measured by the fraction of links formed across groups. The next dependent variables is the level of inter-group contact contingent on having changed norms, when outcomes display partial norm change, in column **III**, or global norm change in column **IV**.

Table B-2 The effect of barriers on group outcomes

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the level of partial norm change outcomes in column I. The subsequent variable is the level of inter-group contact in column II. Then, inter-group contact of norm changers i norm change is partial in column III or global in column IV. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	Partial norm change *Minor*	**Inter-group contact** *Minor-Major*	**Contact (Partial)** *Minor*	**Contact (Global)** *Minor*
	I	**II**	**III**	**IV**
IGNORANCE	1.333**	−0.353***	−0.156	−0.406***
	(0.063)	(0.113)	(0.120)	(0.099)
SANCTIONS	0.358***	−0.260**	0.046	−0.062
	(0.104)	(0.105)	(0.089)	(0.106)
REVEAL	0.283***	−0.292***	−0.178*	−0.085
	(0.109)	(0.093)	(0.097)	(0.109)
Constant	0.125***	0.772***	0.822**	0.661***
	(0.027)	(0.052)	(0.056)	(0.082)
χ^2	19.87***	17.29***	6.40*	26.05***
# Obs.	480	480	789	256
# Groups	48	48	34	99

Table B-3 reports regressions testing differences in inter-group proposals by players from the minority who adopt the mainstream norm of the majority. In all regressions the independent variable is a dummy for the *partial change outcome*, for which the outcome where all players choose the mainstream norm is the omitted category. Moreover, I use group random effects in all regressions. The table reports regressions for the BASELINE in column **I**, IGNORANCE in column **II**, SANCTIONS column **III**, and REVEAL in column **IV**.

Table B-4 reports regressions testing differences in inter-group ties of players from the minority who adopt the mainstream norm of the majority. In all regressions the independent variable is a dummy for the partial norm change outcome, for which the outcome where all players choose the mainstream norm is the omitted category. Moreover, I use group random effects in all regressions. The table reports regressions for the BASELINE in column **I**, IGNORANCE in column **II**, SANCTIONS column **III**, and REVEAL in column **IV**.

Table B-5 reports regressions testing differences in the randomization of groups by treat-

Table B-3 The effect of partial change outcomes on inter-group proposals

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the level of inter-group proposals from the minority to the majority, by treatment. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	BASELINE *Minor-Major*	IGNORANCE *Minor-Major*	SANCTIONS *Minor-Major*	REVEAL *Minor-Major*
	I	**II**	**III**	**IV**
Partial change	−0.041	0.018	−0.048	−0.022
	(0.056)	(0.067)	(0.029)	(0.050)
Constant	0.936***	0.813***	0.914***	0.849***
	(0.025)	(0.067)	(0.037)	(0.039)
χ^2	0.54	0.07	2.68	0.20
# Obs.	343	235	214	253
# Groups	12	12	11	12

Table B-4 The effect of partial change outcomes on inter-group ties

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the level of inter-group ties between the minority and the majority, by treatment. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	BASELINE *Minor-Major*	IGNORANCE *Minor-Major*	SANCTIONS *Minor-Major*	REVEAL *Minor-Major*
	I	**II**	**III**	**IV**
Partial change	−0.242***	−0.327***	−0.282***	−0.119**
	(0.069)	(0.126)	(0.059)	(0.059)
Constant	0.823***	0.663***	0.870***	0.648***
	(0.058)	(0.110)	(0.070)	(0.070)
χ^2	12.36***	6.70***	22.51***	4.10**
# Obs.	343	235	214	253
# Groups	12	12	11	12

ment, on the effect that fixed social norms (i.e., no norm change) have on choices and earnings. In all regressions the independent variables are dummies for each treatment, for which the BASELINE treatment is the omitted category. Moreover, I use group random effects in all regressions. The first dependent variable in column **I** is the level of inter-group contact, measured by the fraction of links formed across groups. The subsequent dependent variable is the level of conformity. For the fixed-norms game, this represents the case in which a player chooses the action that aligns with her assigned social norm, for the minority in column **II** and the majority in column **III**. Finally, the last dependent variable is the level of economic mobility, expressed as the fraction of earnings normalized by the prediction for the *no change outcome*, for the minority in column **IV** and for the majority in column **V**.

Table B-6 reports regressions testing differences in economic mobility between the setting with flexible social norms and the setting with fixed social norms for the *minority*, to

Table B-5 The effect of fixed social norms on choices and earnings

GLS regressions with standard errors clustered on groups (in parenthesis). The first dependent variable is the level of inter-group contact in column I. The second dependent variable is the level of conformity, for the minority in column II, and for the majority in column III. Finally, the level of economic mobility for the minority in column IV and for the majority in column V. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	Inter-group contact	**Conformity**		**Economic mobility**	
	Min-Maj	*Minor*	*Major*	*Minor*	*Major*
	I	**II**	**III**	**IV**	**V**
IGNORANCE	0.004	0.022	−0.010	−0.006	−0.035
	(0.018)	(0.611)	(0.019)	(0.099)	(0.074)
SANCTIONS	0.012	−0.006	0.002	−0.002	0.063
	(0.033)	(0.087)	(0.007)	(0.103)	(0.039)
REVEAL	0.034	−0.009	−0.027	0.047	−0.068
	(0.033)	(0.089)	(0.021)	(0.100)	(0.055)
Constant	0.033***	0.894***	0.988***	0.752***	0.869***
	(0.011)	(0.055)	(0.006)	(0.078)	(0.030)
χ^2	1.11	0.46	2.31	0.46	7.54
# Obs.	480	1440	1920	1440	1920
# Groups	36	36	36	36	36

evaluate the effect of norm change on earnings. In all regressions the independent variable is a dummy for the case with flexible social norms, so that having fixed norms is the omitted category. I use group random effects in all regressions. The magnitude of the economic mobility, i.e., earnings as a fraction of the predicted gains in the *no change outcome*, is the dependent variable for treatment BASELINE in column **I**, for treatment IGNORANCE in column **II**, for SANCTIONS in column **III** and for REVEAL in column **IV**.

Table B-6 Effect of flexible norms on economic mobility for the minority

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the magnitude of the economic mobility of the minority in BASELINE in column I, in IGNORANCE in column II, in SANCTIONS in column III, and in REVEAL in column IV. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	BASELINE	**IGNORANCE**	**SANCTIONS**	**REVEAL**
	Minor	*Minor*	*Minor*	*Minor*
	I	**II**	**III**	**IV**
Flexible norms	0.928***	0.402***	0.162*	0.258**
	(0.064)	(0.109)	(0.090)	(0.130)
Constant	0.752***	0.746***	0.750***	0.799***
	(0.081)	(0.064)	(0.069)	(0.065)
χ^2	207.20***	13.61***	3.22*	3.96**
# Obs.	720	720	720	720
# Groups	12	12	12	12

Table B-7 reports regressions testing differences in economic mobility between the setting with flexible social norms and the setting with fixed social norms for the *majority*, to evaluate the effect of norm change on earnings. In all regressions the independent variable is a dummy for the case with flexible social norms, so that having fixed norms is the omitted category. I use group random effects in all regressions. The magnitude of the economic mobility, i.e., earnings as a fraction of the predicted gains in the *no change outcome*, is the dependent variable for treatment BASELINE in column **I**, for treatment IGNORANCE in column **II**, for SANCTIONS in column **III** and for REVEAL in column **IV**.

Table B-7 Effect of flexible norms on economic mobility for the majority

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the magnitude of the economic mobility of the majority in BASELINE in column I, in IGNORANCE in column II, in SANCTIONS in column III, and in REVEAL in column IV. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	BASELINE *Major*	**IGNORANCE** *Major*	**SANCTIONS** *Major*	**REVEAL** *Major*
	I	**II**	**III**	**IV**
Flexible norms	0.554***	0.344***	0.345***	0.383***
	(0.051)	(0.072)	(0.066)	(0.079)
Constant	0.869***	0.835***	0.932***	0.801***
	(0.031)	(0.070)	(0.026)	(0.048)
χ^2	117.55***	22.58***	27.47***	23.24***
# Obs.	960	960	960	960
# Groups	12	12	12	12

At the end of the experiment each participant makes an allocation choice once, by dividing 10-points between an in-group and an out-group receiver. Therefore, there are 108 observations for the minority and 144 for the majority, across treatments. I run OLS regressions with robust standard errors to test differences in the allocation bias towards the in-group, across treatments. Table B-8 reports these regressions for the minority in column **I** and the majority in column **II**. In all regressions the independent variables are dummies for each treatment, for which the BASELINE treatment is the omitted category.[36]

Table B-9 reports regressions testing differences in choices in the treatment REVEAL, contingent on whether a player revealed her chosen norm for the *minority*. In all regressions the independent variable is a dummy for the case where a player revealed her chosen norm, so that keeping the norm concealed is the omitted category. I use group random effects in all regressions. The dependent variable is the level of norm change in column **I**, the level of inter-group contact in column **II**, the level of economic mobility in column **III** and the frequency of partial norm change outcomes in column **IV**.

Table B-10 reports regressions testing what drives participants in the minority to reveal

[36] Note there are 7 observations missing for the minority (101 remaining) and 11 for the majority (133 remaining). These were dropped because the allocation was biased towards the out-group, which is most likely due to a failure in understanding the instructions.

Table B-8 The effect of assimilation on in-group allocation bias

OLS regressions with robust standard errors (in parenthesis). The dependent variable is the magnitude of the bias towards the in-group for the minority in column I and for the majority in column II. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	In-group bias	
	Minor	*Major*
	I	**II**
IGNORANCE	0.148	0.032
	(0.090)	(0.091)
SANCTIONS	0.262^{**}	-0.055
	(0.103)	(0.087)
REVEAL	0.105	-0.099
	(0.097)	(0.084)
Constant	0.222^{***}	0.364^{***}
	(0.062)	(0.063)
Adjusted R^2	0.051	0.016
# Obs.	134	180

their chosen norm in treatment REVEAL. In both regressions, the dependent variable is the revelation choice for the minority, without controls in column **I** and controlling for the number of adopters in the in-group in column **II.** I use group random effects in all regressions.

Table B-9 Effect of revelation of norms on choices and outcomes for the minority

GLS regressions with standard errors clustered on groups (in parenthesis). The dependent variable is the level of norm change in column I, the level of inter-group contact in column II, the level of economic mobility in column III and the frequency of partial norm change outcomes in column IV. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	Norm change	**Inter-group contact**	**Economic mobility**	**Partial norm change**
	Minor	*Minor*	*Minor-Major*	*Minor*
	I	**II**	**III**	**IV**
Revealed norms	0.164^{***}	0.232^{***}	-0.250^{*}	0.223^{***}
	(0.055)	(0.078)	(0.149)	(0.039)
Constant	0.666^{***}	0.428^{***}	1.113^{***}	0.359^{***}
	(0.1000)	(0.084)	(0.131)	(0.108)
χ^2	8.84^{***}	8.80^{***}	2.82^{*}	33.05^{***}
# Obs.	360	360	360	360
# Groups	12	12	12	12

Table B-10 The effect of endogenous barriers on norm revelation

GLS regressions with standard errors clustered on groups (in parenthesis). In all regressions the dependent variable is the level of revelation for the minority, without controls in column I and controlling for the number of adopters in the in-group in column II. ***, ** and * indicate statistical significance at the 0.01, 0.05 and 0.10 levels.

	Norm revelation	
	Minor	*Minor*
	I	**II**
Norm change	0.227^{***}	0.457^{***}
	(0.066)	(0.101)
1 adopter in subgroup		0.043
		(0.055)
2 adopters in subgroup		−0.013
		(0.037)
3 adopters in subgroup		-0.396^{***}
		(0.097)
Constant	0.063	0.085^{**}
	(0.049)	(0.038)
χ^2	11.86^{***}	57.81^{***}
# Obs.	360	360
# Groups	12	12

C Additional analysis

C.1 Description of the sample

Table C-1 reports the count and percentage of females participants and males participants in the experiment. Totals are reported separately by subgroup: majority or minority. I run non-parametric Wilcoxon-Mann-Whitney tests to evaluate the underlying distributions of females between treatments as well as between subgroups across treatments. The pairwise comparison shows no significant difference in the distribution of females between any pair of treatments, except for BASELINE with respect to IGNORANCE ($p = 0.013$) and with respect to REVEAL ($p = 0.089$). By taking a deeper look into the subgroups, majority and minority, I find that the effect is driven by lower levels of females in the majority subgroup in BASELINE, while there are no differences in the distribution of females in the minority subgroup across treatments.

Table C-1 Descriptive statistics of the sample

Frequency of female and male participants by treatment. Counts are reported separately by subgroup: majority or minority. Percentages in parentheses.

	Treatments								**Totals**
	BASELINE		IGNORANCE		SANCTIONS		REVEAL		
	Major	*Minor*	*Major*	*Minor*	*Major*	Minor	*Major*	*Minor*	
Female	21	19	34	22	30	19	31	20	196
	(0.44)	(0.53)	(0.71)	(0.61)	(0.63)	(0.53)	(0.65)	(0.53)	(0.58)
Male	27	17	14	14	18	17	17	16	140
	(0.56)	(0.47)	(0.29)	(0.39)	(0.37)	(0.43)	(0.35)	(0.47)	(0.42)

Table C-2 reports the average earnings per period in the game with fixed social norms, the game with flexible social norms, and the other-other allocation choice. Means are reported separately by gender (female and male) and by subgroup (majority and minority). As reported in the main text, there are no significant differences in earnings in the game with fixed social norms across treatments. On the contrary, in the game with flexible social norms, the minority earns significantly more in BASELINE than in both IGNORANCE and SANCTIONS and there are no differences between the last two.

Table C-3 reports summary statistics of the main variables in the experiment for all four treatments (BASELINE, IGNORANCE, SANCTIONS, and REVEAL), separate for the majority and the minority. In complement to Table 2, the table reports *norm change*, *within-group* proposals and links, *inter-group* proposals and links, *conformity*, *economic mobility*, and *in-group bias*.

Table C-2 Earnings and allocations by gender

Average earnings per period in PART 2, average earnings per period in PART 3, and average allocation to in-group in PART 4. Standard deviations in parenthesis.

	Treatments							
	BASELINE		IGNORANCE		SANCTIONS		REVEAL	
	Major	*Minor*	*Major*	*Minor*	*Major*	Minor	*Major*	*Minor*
Fixed norms								
Female	15.21	10.97	15.02	10.05	16.72	12.11	14.31	11.54
	(2.32)	(3.52)	(4.42)	(3.79)	(1.97)	(1.74)	(3.46)	(3.28)
Male	15.98	10.01	15.02	11.04	16.85	8.68	14.61	10.73
	(2.08)	(4.51)	(4.33)	(2.73)	(1.70)	(3.96)	(3.18)	(3.92)
Flexible norms								
Female	24.29	22.43	20.22	14.14	22.30	14.10	20.98	15.45
	(5.10)	(5.17)	(5.78)	(6.12)	(5.02)	(6.10)	(4.74)	(6.34)
Male	26.65	24.71	23.60	19.10	24.13	11.25	21.90	13.98
	(4.36)	(3.91)	(5.13)	(6.33)	(4.00)	(5.55)	(5.28)	(5.66)
Allocation								
Female	6.19	5.89	6.20	6.63	6.83	6.63	6.35	5.90
	(1.86)	(1.63)	(2.60)	(2.44)	(2.16)	(2.41)	(1.83)	(2.12)
Male	6.85	6.35	7.07	6.50	5.61	6.70	6.11	6.93
	(2.39)	(2.08)	(2.64)	(1.87)	(1.91)	(3.40)	(2.02)	(2.35)

Table C-3 Summary statistics in the coordination network game

Average fractions (percentages) for each of the main variables. Standard deviations in parenthesis. Observations are disaggregated by majority and minority; pooled for the game with *fixed* social norms and separated by treatment for the game with *flexible* social norms.

	Fixed norms		Flexible norms							
			BASELINE		IGNORANCE		SANCTIONS		REVEAL	
	Major	*Minor*	*Major*	*Minor*	*Major*	Minor	*Major*	*Minor*	*Major*	*Minor*
Norm change	n/a	n/a	0.01	0.95	0.01	0.65	0.01	0.59	0.02	0.70
			(0.09)	(0.21)	(0.08)	(0.48)	(0.08)	(0.49)	(0.13)	(0.46)
WITHIN-GROUP										
proposals	0.94	0.92	0.95	0.91	0.96	0.88	0.98	0.78	0.96	0.81
	(0.17)	(0.24)	(0.16)	(0.23)	(0.15)	(0.29)	(0.09)	(0.32)	(0.15)	(0.32)
links	0.88	0.85	0.91	0.85	0.93	0.80	0.97	0.68	0.92	0.71
	(0.23)	(0.29)	(0.20)	(0.30)	(0.20)	(0.34)	(0.11)	(0.38)	(0.19)	(0.37)
INTER-GROUP										
proposals	0.15	0.16	0.82	0.91	0.50	0.59	0.57	0.63	0.55	0.66
	(0.29)	(0.29)	(0.20)	(0.23)	(0.06)	(0.16)	(0.33)	(0.57)	(0.41)	(0.43)
links	0.04		0.77		0.42		0.51		0.48	
	(0.14)		(0.33)		(0.44)		(0.43)		(0.40)	
Conformity	0.98	0.90	0.99	0.98	0.99	0.96	0.99	0.91	0.99	0.94
	(0.15)	(0.31)	(0.06)	(0.16)	(0.09)	(0.20)	(0.06)	(0.28)	(0.10)	(0.23)
Efficiency	0.86	0.76	1.42	1.68	1.18	1.15	1.28	0.91	1.18	1.06
	(0.26)	(0.35)	(0.33)	(0.57)	(0.39)	(0.71)	(0.35)	(0.67)	(0.37)	(0.66)
In-group bias	n/a	n/a	0.36	0.22	0.39	0.37	0.31	0.48	0.26	0.32
			(0.42)	(0.37)	(0.43)	(0.38)	(0.40)	(0.45)	(0.38)	(0.43)

C.2 Norm choices by treatment

In this section, I further explore the choices of norm-change of the minority across treatments. The first consideration is to see in which period minority players adopt the mainstream norm for the first time. Across the four treatments the median period in which norm change starts is period 1. Figure 10A illustrates the fraction of participants that becomes mainstream in each of the 10 periods in the game with flexible norms, by treatment. The bars on period 0 represent the fraction of participants that never change norms, which are positive for all but BASELINE. By period 3, all participants in BASELINE had already changed norms at least once, while the latest period in which someone changed norms for the first time is 5 in REVEAL, 8 in IGNORANCE, and 10 in SANCTIONS.

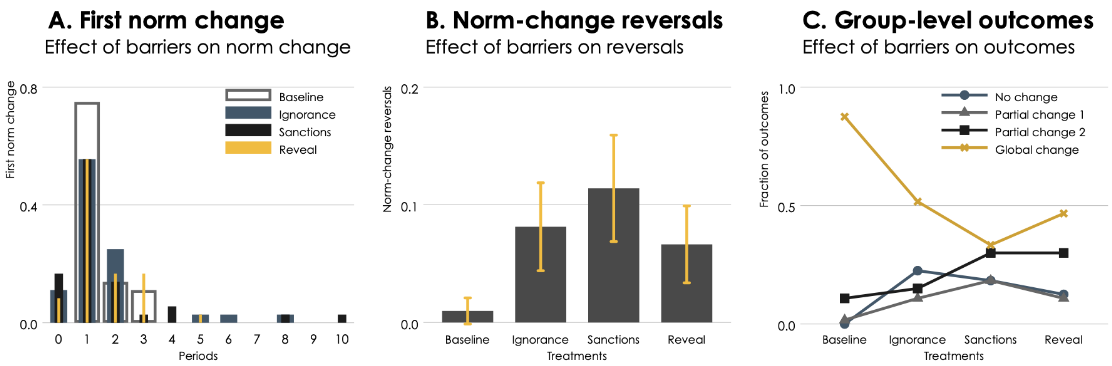


Figure 10 Assumption choices

Panel A illustrates the fraction of minority members that change norms for the first time per period. Panel B shows the fraction of participants who reverse their norm-change choice by adopting back their original social norm in a subsequent period. In both panels choices are disaggregated by treatment. Panel C reports the frequency of outcomes by treatment.

The second consideration with respect to the norm-change choice is whether minority players reverse it. That is, after having adopted the social norm of the majority, I look at the fraction of adopters who choose back the minority's original social norm. This is illustrated in Figure 10B. I run non-parametric Wilcoxon-Mann-Whitney tests to evaluate the underlying distributions of choice reversals between treatments. The pairwise comparison shows that there is only a statistically significant difference between BASELINE, where only 1% of the adopters reverse, and the other three treatments; IGNORANCE where 8% reverse ($p < 0.001$), SANCTIONS where 11% ($p < 0.001$) and REVEAL where 7% reverse ($p < 0.001$).

Finally, I show in more detail the frequency of outcomes as a consequence of the norm choices of the minority (see Figure 10C). The *global change outcome* is most frequently observed in BASELINE and least in SANCTIONS. On the other extreme, the *no change outcome* does not occur in BASELINE, is most frequent in IGNORANCE and least in REVEAL. Finally, the *partial change outcome* is most frequent in SANCTIONS when two in the minority become mainstream (which can lead to economic mobility), and also when only one minority participant adopts the majority's norms (which leaves all three players worse-off than in the *no change outcome*).

Thus, the analysis on norm choices suggest that the reason why the average norm change in SANCTIONS is lower than BASELINE, and the frequency of the *partial change outcome* is highest in SANCTIONS than in any other treatment, is possibly due to: (i) minority players waiting to change norms in SANCTIONS even to the last period, and (ii) minority players being somewhat more likely to reverse the norm-change choice in SANCTIONS than in the other treatments.

C.3 Network choices by treatment

After the norm-change choices have been made, participants make network choices. Specifically, each participant makes linking proposals to both in-group and out-group members. This is illustrated in Figure 11, which depicts connectivity choices towards the in-group (Panel A) and towards the out-group (Panel B), for the minority (filled bars) and the majority (empty bars). The top of each range bar reports the proportion of proposals made, the bottom reports the proportion of links formed, and the height of each bar is the proportion of unreciprocated proposals (i.e., failed links).

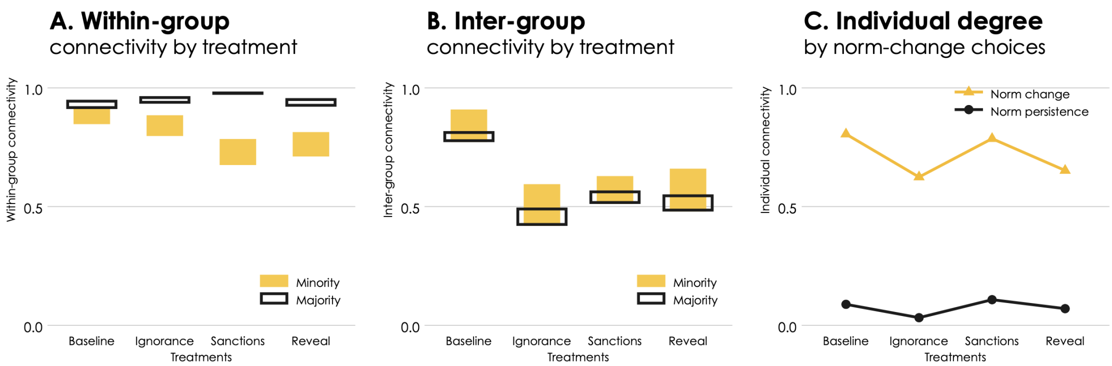


Figure 11 Network connectivity: Proposals and Links

Panel A illustrates the average level of connectivity within groups (minority with minority and majority with majority). Panel B illustrates the average level of connectivity across groups (minority with majority). Panel C illustrates the average level of connectivity of the minority, for players that change norms and those that persist on their group norm. In all panels choices are disaggregated by treatment.

Finally, Figure 11C reports the average level of connectivity of minority players who become mainstream (change norms) versus minority players who reject the mainstream norm, for each treatment. Notably, connectivity is mainly driven by norm adoption, even when the choice is not visible, as in IGNORANCE and REVEAL.

D Instructions

Below I include a sample of the instructions participants read for each part in treatment BASELINE. Each set of instructions has comprehension questions that participants were required to answer correctly before continuing. I also specify when differences with instructions for the other treatments when necessary.

Instructions were originally in Spanish.

Instructions Part 1

Welcome to this study

You will receive a minimum pay of **10,000 pesos** for your participation in this study. Please read these instructions carefully to find out how you can earn **additional money**.

All interactions take place through the computers. Please do not talk or communicate with the other participants in any other way.

Please raise your hand if you have any question, and an experimentalist will come and answer your question privately.

This study is **anonymous**. Therefore, your identity will not be revealed to the other participants nor theirs to you.

This study has **4 parts**. You will now only read the instructions for Part 1. Once Part 1 is over, you will read the instructions for Part 2, and so on. In all four parts you will interact with the same participants.

You will participate with other 6 participants in the different parts of the study. 4 participants will be assigned to group ● and 3 participants will be assigned to group ▲. This is explained in the next screens.

In this experiment you can earn points depending on your choices and the choices made by the other participants. The amount of points you can earn is explained in the instructions for each part. At the end of the study, we will convert the total number of points you have earned into pesos using the following exchange rate: **1 point = 800 pesos**. You will receive your earnings in cash.

[*page break*]

Part 1: Your Group

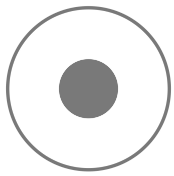
Here we explain the visual information of your group (the way it will be displayed on the screen). This information will be useful for the subsequent parts of the experiment. In this part you will also make a decision together with the other participants assigned to your group: you will choose a name for your group. This is explained as follows:

1. Group image

You have been assigned to **group** ● (internal symbol) and your **appearance** is ○ (external symbol).

There are 4 players in your group ●, you included. There are 3 players in group ▲.

Each player in your group will be displayed on the screen using the following image:



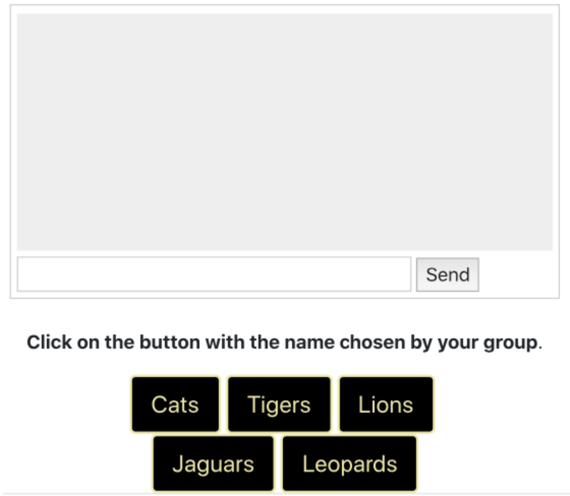
In the chat box on the right hand side you can communicate with the other participants in group ● and choose a name for your group, *by clicking on the button with the name you prefer as a group*. **To choose a name, all 4 players in your group must click on the same button**.

2. Group name

In the chat box below, you will be labeled as **Player 1**. It is not allowed to use any offensive language. And to maintain anonymity, it is not allowed to send any information that can be used by the other participants to identify who you are.

Write on the chat box to communicate with the other players in your group.

[*page break*]



Instructions Part 2

This is Part 2

In the following screens you will see the instructions of Part 2. At the end of the instructions you will see a summary of the most important information. This **summary** will be available in each screen during Part 2.

Part 2 has **10 rounds**. As in Part 1, once Part 2 has ended, you will read the instructions for Part 3. In Part 2 you can interact with the same 6 other players from Part 1, those in group ● (Cats) and those in group ▲ (Dogs).

Each of the 10 rounds in Part 2 have three stages: **Connections** (stage 1), **Actions** (stage 2), and **Results** (stage 3). The choices you make in each stage are explained in the following screens.

You will receive payment for **one of the 10 rounds** of Part 2. At the end of the study, the computer will randomly choose which round will be used for payment.

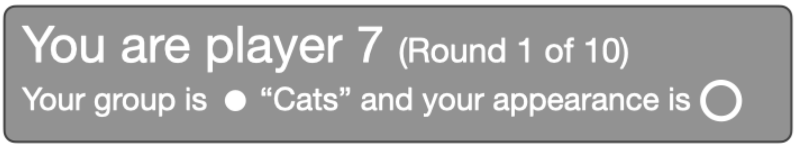
[*page break*]

Decisions in Part 2

Groups, Appearances & Labels

In Part 2, you will participate with 6 other players (the same as in Part 1). Each player is randomly assigned to a group ● “Cats” or ▲ “Dogs” (internal symbol), and will also have an appearance ○ or △ (external symbol). Each player will also have a numeric label from 1 to 7. You will see the group, appearance and label of each player on the screen (see Figure 1).

On the header of each screen you will see a summary of this information as in the example below:



In each round there will be 4 players in group ● “Cats” and 3 in group ▲ “Dogs”. A player’s group and appearance **will not change** between rounds during Part 2, while his numeric label and position in the screen will be **randomly changed in each round**.

For instance, the player in the header above is in group ● “Cats” and has the numeric label 7 in round 1. In round 2, the same player will be in the same group than before but his label may be any number from 1 to 7, and his position may be anywhere on the screen. The same is the case for all other players.

Stage 1: Connections

In the **Connections stage** you can propose a connection to those players you want to participate with in Stage 2. You can propose a connection to any player no matter their group or appearance. If you propose a connection to a player who **also** proposes to you, there is an **active connection** between the two of you. In Stage 2 you will **only participate** with the players with whom you have an active connection.

You will see your connections on the screen (see Figure 2). There is a *light arrow* leaving from you for each connection you proposed, and a ***dark arrow*** heading towards you for each connection proposed to you. **If both** arrows are present, you have an **active connection** with that other player. If only one of you proposes a connection but the other does not, there is no active connection between the two of you. **Your connections are illustrated with thicker lines and those of others are thinner.*

Stage 2: Actions

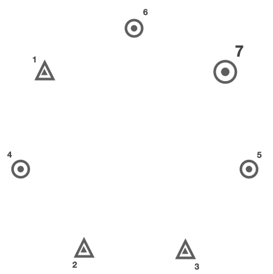


Fig. 1: Groups

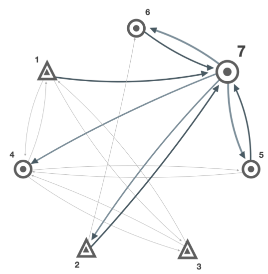


Fig. 2: Connections

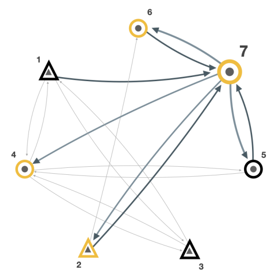


Fig. 3: Actions

In the **Actions stage** you will pick one of two actions: mustard or black. You only pick one action for the stage, not one for each connection. You will get points for each of your active connections choosing **the same action as you**. You will **not** get any points for an active connection choosing **a different action** than you, nor from a player choosing the same action as you, if he **is not** an active connection.

The action chosen by each player is illustrated by the color of his appearance (see Figure 3).

[*page break*]

Comprehension Questions

Before you continue, please answer the following comprehension questions:

1. You have been assigned to a group: ● or ▲ (inner symbol). How often does your group change?
   - It is fixed and does not change
   - The computer changes it in each round
   - I can change it in each round
2. You have been given an appearance: ○ or △ (external symbol). How often does your appearance change?
   - It is fixed and does not change
   - The computer changes it in each round
   - I can change it in each round by changing my group

3. You will be assigned a label from 1 to 7. How often does your label change?
   - It is fixed and does not change
   - The computer changes it in each round
   - I can change it in each round
4. In Stage 2 (Actions) you only participate with your active connections. When is a connection active?
   - When I propose a connection to another player regardless of he/she proposing a connection to me
   - When another player proposes a connection to me regardless of me proposing a connection to him/her
   - When I propose a connection to a player who also proposes a connection to me
5. In Figure 2 above, how many active connections does player 7 have?
   - 5
   - 4
   - 3

[*page break*]

Points in Part 2

Stage 3: Results

In stage 3 for each round, you will see how many points you get in that round. The points are calculated by the gains of coordinating with your active connections minus the costs of your proposed connections. A coordination is an instance where you and another player with whom you have an active connection both chose the same action: mustard or black. The computer will do the calculations for you in each round, but below you can see how the points are calculated in 3 simple steps:

Step 1: Gains from coordination

Gains in each round depend on (i) your group, (ii) the action you choose, and (iii) the number of your active connections choosing the same (including yourself). You do not earn anything with a player if you do not have an active connection together and choose the same action.

If your group is ● "Cats" and you:

- choose mustard, you get **6 points** for each active connection also choosing mustard (including yourself) and **0 points** for any active connection choosing black.
- choose black, you get **4 points** for each active connection also choosing black (including yourself) and **0 points** for any active connection choosing mustard.

If your group is ▲ "Dogs" and you:

- choose black, you get **6 points** for each active connection also choosing black (including yourself) and **0 points** for any active connection choosing mustard.
- choose mustard, you get **4 points** for each active connection also choosing mustard (including yourself) and **0 points** for any active connection choosing black.

Step 2: Costs from proposals

Your costs in each round depend **only on the number of connections you propose**. You pay a cost of **2 points** for each proposal you make, independently of whether the other player also proposes a connection to you or not.

Step 3: Total points

Your points are calculated as the **gains from coordinating with your active connections** minus the **cost of each connection you propose**. The examples illustrate this:

*Example 1: Player 6 in group ● **"Cats"** chose mustard.*

1. Earnings: Player 6 formed **two** active connections (with players 2 and 3) and have **two** coordinations, (with player 2 and with himself). The two coordinations are multiplied by **6** (because his group is ● "Cats" and he chose mustard).

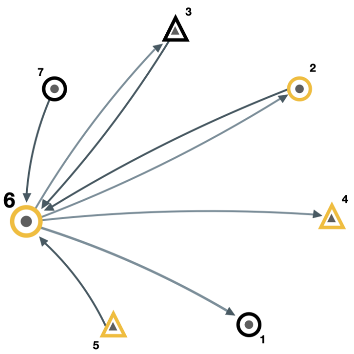


Fig. Example 1

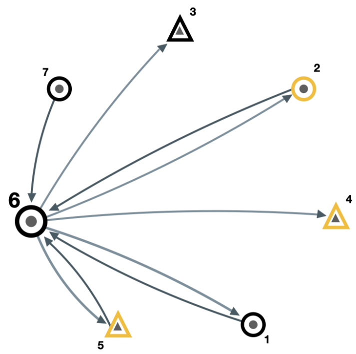


Fig. Example 2

Player 6 did not coordinate with player 3, therefore he does not earn points with that active connection. Player 6 coordinated with players 4 and 5, but did not form an active connection with either of them, so he cannot earn points with them. Player 6's earnings are: **12 points.**

2. Costs: Player 6 proposed **four** connections (to players 1, 2, 3 and 4), the cost of each proposal is 2 points, so he pays **8 points**.

3. Total Points: Player 6 gets 12 in gains minus 8 in costs: 12 - 8 = **4 points** in this round.

*Example 2: Player 6 in group ● **"Cats"** chose black.*

1. Earnings: Player 6 formed **three** active connections (with players 1, 2 and 5) and have **two** coordinations, (with player 1 and with himself). The two coordinations are multiplied by **4** (because his group is ● "Cats" and he chose black).

Player 6 did not coordinate with players 2 and 5, therefore he does not earn points with those active connection. Player 6 coordinated with players 3 and 7, but did not form an active connection with either of them, so he cannot earn points with them. Player 6's earnings are: **8 points.**

2. Costs: Player 6 proposed **five** connections (to players 1, 2, 3, 4 and 5), the cost of each proposal is 2 points, so he pays **10 points**.

3. Total Points: Player 6 gets 8 in gains minus 10 in costs: 8 - 10 = **-2 points** in this round.

[*page break*]

Comprehension Questions

Before you continue, please answer the following comprehension questions:

1. Imagine your group is ● "Cats", you chose action mustard and you have one active connection with a player who chose action mustard. What are the total points you get *from this connection* (not including what you get from coordinating with yourself)?
   - I gain 6 and pay the cost of 2 = 4 points in total
   - I gain 4 and pay the cost of 2 = 2 points in total
   - I gain 0 and pay the cost of 2 = -2 points in total
2. Imagine your group is ● "Cats", you chose action mustard and you have one active connection with a player who chose action black. What are the total points you get *from this connection* (not including what you get from coordinating with yourself)?
   - I gain 6 and pay the cost of 2 = 4 points in total
   - I gain 4 and pay the cost of 2 = 2 points in total
   - I gain 0 and pay the cost of 2 = -2 points in total

[*page break*]

Summary

The summary of the instructions (below) will be available in each round. To display it just click on the button **Show Summary** at the bottom of the screen.

Instructions Part 2

Groups, Appearances & Labels

In Part 2 you will interact with 6 other players. At the beginning of the experiment each of you was assigned to one of two groups: ● "Cats" or ▲ "Dogs" (internal symbol), and was assigned an appearance: ○ or △ (external symbol) and a numeric label (between 1 and 7). Your group and appearance are fixed for all 10 rounds in Part 2 while your numeric label and your position on the screen will change in each round. **There are 3 stages in each round.**

Stage 1: Connections

In stage 1, you can propose connections to the other players. A connection with another player is **active** if both you and the other player propose a connection to each other. Each connection you propose will cost you 2 points, independently of whether the other player also proposes a connection to you or not.

Stage 2: Actions

In stage 2, you will choose an action: mustard or black. You will earn points depending on your group, the action you choose, and the number of your active connections choosing **the same action** as you:

If your group is ● "Cats" and you:

- choose mustard, you get **6 points** for each active connection also choosing mustard (including yourself) and **0 points** for any active connection choosing black.
- choose black, you get **4 points** for each active connection also choosing black (including yourself) and **0 points** for any active connection choosing mustard.

If your group is ▲ "Dogs" and you:

- choose black, you get **6 points** for each active connection also choosing black (including yourself) and **0 points** for any active connection choosing mustard.
- choose mustard, you get **4 points** for each active connection also choosing mustard (including yourself) and **0 points** for any active connection choosing black.

Stage 3: Results

In stage 3, you will see how many points you get in that round. The points are calculated by the gains of coordinating with your active connections minus the costs of your proposed connections. The border of each player's appearance (○ or △) shows the action that player chose.

Earnings in Part 2

You will participate for 10 rounds in Part 2. At the end of the experiment the computer **will randomly select one of the 10 rounds**. The number of points you got in the selected round will be added to the total number of points used to determine your payment.

[*page break*]

Instructions Part 3

Welcome to Part 3

You have concluded Part 2 and will now begin Part 3:

First, you will read the instructions for **Part 3**.

Then, you will participate in Part 3 for **10** rounds.

You will be paid for **one of the 10 rounds** in Part 3. As in Part 2, at the end of the study, the computer will randomly choose which round will be used for payment.

[*page break*]

Decisions in Part 3

In Part 3 you will participate for 10 rounds with the same 6 other players from Part 2. While, each round in Part 2 had 3 stages, each round in Part 3 will have 4 stages: **Group choice** (Stage 1), **Connections** (Stage 2), **Actions** (Stage 3) and **Results** (Stage 4).

The Connections, Actions and Results stages are the same as in Part 2. The **Group Choice** stage is **new** and is explained below.

Stage 1: Group Choice

At the beginning of each round, **before Stage 1**, all players will be assigned to the same group they belonged to in Part 2. If you were ● "Cats" in Part 2, you will begin each round in Part 3 as a ●. If you were ▲ "Dogs" in Part 2, you will begin each round in Part 3 as a ▲.

Then, in the **Group Choice** stage, each player will decide if he wants to stay in his group or if he wants to change it.

BASELINE: **The group symbol will indicate the group each player has chosen. A player's appearance will not change even if he changes his group.**

IGNORANCE: **The group symbol will not be visible regardless of whether a player changes or stays in his group. A player's appearance will not change even if he changes his group.**

SANCTIONS: **The group symbol will indicate the group each player has chosen. A player's appearance will not change even if he changes his group. Staying in the same group is free, but changing group has a fixed cost of 6 points, plus 2 points for each player in your group that decides not to change**

REVEAL: **The group symbol will indicate the group each player has chosen. But it will only be visible if the player has chosen to reveal it (see Stage 2). A player's appearance will not change even if he changes his group.**

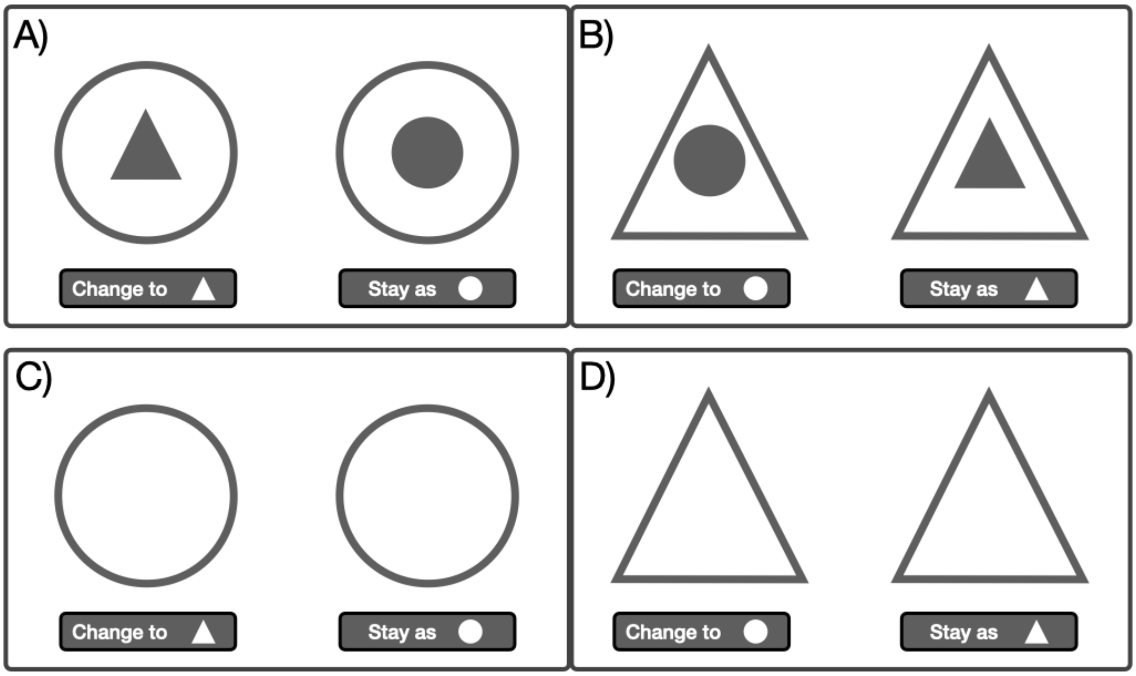
The example below shows the choices a player can make depending on his group at the beginning of each round. The image above each button indicates how other players will see him in the remaining three stages of that round (Connections, Actions and Results), depending on whether he changes or stays in his group.

Points in Part 3:

If a player in group ● "Cats" stays in the same group, he will continue earning points as a ● "Cats". That is, 6 points for each coordination on mustard and 4 points for each coordination on black.

However, if a player in group ● "Cats" changes to group ▲ "Dogs", he will earn points as a ▲ "Dogs". That is, 6 points for each coordination on black and 4 points for each coordination on mustard.

Similarly, if a player in group ▲ "Dogs" changes to group ● "Cats", he will earn points as a ● "Cats". That is, 6 points for each coordination on mustard and 4 points for each coordination on black.



Group choice for a player belonging to group ● "Cats" (Panel A) or to group ▲ "Dogs" (Panel B). And for players in treatment ignorance or reveal (Panel C and D).

But if he chooses to stay in group ▲ "Cats", he will continue earning points as a ▲ "Cats". That is, 6 points for each coordination on black and 4 points for each coordination on mustard.

Stage 2: Group revelation (*only in treatment* REVEAL)

After the **Group Choice** stage, each player will decide whether he wants to reveal the group he has chosen, so others see it or not. Not revealing your group is free, while revealing your group has a cost.

The cost has two elements. First, a fixed part of 6 points for revealing, independently of whether you chose to change or stay in your group. In addition, if you chose to change your group, there is a variable part that increases by 2 points for each person in your group that chooses not to change.

Stage 2: Connections

This Stage is the same as in Part 2. Each player can propose to others and pays a cost of 2 points for each connection proposed.

Stage 3: Actions

This Stage is the same as in Part 2. Each player chooses one of two actions: mustard or black and earns points by the number of his/her active connections choosing the same action.

Stage 4: Results

This Stage is the same as in Part 2. Each player is informed of the number of total points he gets in the round. Points are calculated by the gains from coordination (which depend on the group chosen in the current round, the action chosen, and the number of coordinations with active connections) minus the costs of proposing connections to others.

Comprehension Questions

Before you continue, please answer the following comprehension questions:

1. You have been assigned to a group: ● or ▲ (inner symbol). How often does your group change?
    - It is fixed and does not change
    - The computer changes it in each round
    - I can change it in each round
2. You have been given an appearance: ○ or △ (external symbol). How often does your appearance change?
    - It is fixed and does not change
    - The computer changes it in each round
    - I can change it in each round by changing my group
3. You will be assigned a label from 1 to 7. How often does your label change?
    - It is fixed and does not change
    - The computer changes it in each round

- I can change it in each round

4. Imagine you began in group ● "Cats", and you changed to group ▲ "Dogs". Then you chose action mustard and you have one active connection with a player who chose action mustard. What are the total points you get *from this connection* (not including what you get from coordinating with yourself)?

   - I gain 6 and pay the cost of 2 = 4 points in total
   - I gain 4 and pay the cost of 2 = 2 points in total
   - I gain 0 and pay the cost of 2 = -2 points in total

5. Imagine you began in group ● "Cats", and you stayed in your group. Then you chose action mustard and you have one active connection with a player who chose action mustard. What are the total points you get *from this connection* (not including what you get from coordinating with yourself)?

   - I gain 6 and pay the cost of 2 = 4 points in total
   - I gain 4 and pay the cost of 2 = 2 points in total
   - I gain 0 and pay the cost of 2 = -2 points in total

6. After the Group choice (Stage 1) what will other players see about you in the remaining three stages?

   - They can see the group I choose and my new appearance
   - They can see the group I choose and my appearance from Part 2
   - They cannot see the group I choose only my appearance from Part 2

[*page break*]

Summary

The summary of the instructions (below) will be available in each round. To display it just click on the button **Show Summary** at the bottom of the screen.

Instructions Part 3

In Part 3 you will interact with the same 6 players from Part 2. There are 4 stages in each round.

Stage 1: Group Choice

At the beginning of each round, **before Stage 1**, you and all other players will be assigned to the same group they belonged to in Part 2. In Stage 1 you will choose a group. You can stay in the same group you have been assigned to or you can change groups. **The group symbol will indicate the group each player has chosen. A player's appearance will not change when he changes his group.**

Stage 2: Connections

In this stage, you can propose connections to the other players at a cost of 2 points.

Stage 3: Actions

In this stage, you choose an action: mustard or black. You will earn points depending on the group you choose, the action you choose, and the number of your active connections choosing **the same action** as you:

If your choosen group is ● "Cats" and you:

- choose mustard, you get **6 points** for each active connection also choosing mustard (including yourself) and **0 points** for any active connection choosing black.
- choose black, you get **4 points** for each active connection also choosing black (including yourself) and **0 points** for any active connection choosing mustard.

If your chosen group is ▲ "Dogs" and you:

- choose black, you get **6 points** for each active connection also choosing black (including yourself) and **0 points** for any active connection choosing mustard.
- choose mustard, you get **4 points** for each active connection also choosing mustard (including yourself) and **0 points** for any active connection choosing black.

Stage 4: Results

In this stage, you will see how many points you get in that round. Points are calculated by the gains of coordinating with your active connections minus the costs of the connections you have proposed.

Earnings in Part 3

You will participate for 10 rounds in Part 3. At the end of the experiment the computer **will randomly select one of the 10 rounds**. The number of points you got in the selected round will be added to the total number of points used to determine your payment.

[*page break*]

Instructions Part 4

Allocation

This is the last part of the experiment and it is played for a single round. In this round, you will be endowed with 10 points and you have to decide how you want to distribute the points between two players who have been randomly matched with you: Players **A** and **B**.

Player **A** is one of the players assigned to group ● "Cats" in Parts 1 and 2. Player **B** is one of the players assigned to group ▲ "Dogs" in Parts 1 and 2.

The computer will randomly choose the allocation decision of one of the 7 players who have participated together in the experiment, and will implement the payment of that decision. For instance, if your decision is chosen, players **A** and **B** matched with you will receive the points you have allocated to them. None of the other 5 players (including yourself) will receive any points in this part. Similarly, if another player's decision is chosen and you are one of the players **A** or **B** matched with him, you will receive the points allocated to you.

Decision: You have 10 points to allocate between players **A** and **B**. Press the button with the allocation you prefer.

- 0 to A and 10 to B
- 1 to A and 9 to B

- 2 to A and 8 to B
- 3 to A and 7 to B
- 4 to A and 6 to B
- 5 to A and 5 to B
- 6 to A and 4 to B
- 7 to A and 3 to B
- 8 to A and 2 to B
- 9 to A and 1 to B
- 10 to A and 0 to B

[*page break*]

Your participation in this experiment ends here.

In the next screen you will be informed of your earnings. After that, please remain seated until you are called to receive your payment.

To conclude, please press the button indicating your gender:

- Female
- Male

[*End of the instructions.*]