9-20-2022

Inequality as a Barrier to Economic Integration? An Experiment

Gabriele Camera
Chapman University, camera@chapman.edu

Lukas Hohl
University of Basel

Rolf Weder
University of Basel

Inequality as a Barrier to Economic Integration?An Experiment*

Gabriele Camera
Chapman University

Lukas Hohl
University of Basel
Federal Finance Administration

Rolf Weder
University of Basel

September 20, 2022

Abstract

International economic theory suggests that people should embrace economic integration because it promises large gains. But policy reversals such as Brexit indicate a desire for economic *dis*integration. Here we report results of an experiment of how size and cross-country distribution of gains from integration influence individuals' inclination to cooperate to reap its intended benefits and to embrace or reject integration. The design considers an indefinitely repeated helping game with multiple equilibria and strategic uncertainty. The data reveal that inequality of *potential* gains neither affected behavior nor reduced support for economic integration. However, integration may lead to disappointing, unequally distributed welfare gains, undermining support for the policy. This suggests that to better assess integration policies, we should account for the spillover effects of integration on behavior. Miscalculating this behavioral aspect may undermine the intended development goals and motivate calls for dramatic policy-reversals.

Keywords: economic opportunity, endogenous institutions, globalization, indefinitely repeated games, social dilemmas.
JEL codes: C70, C90, F02

---

* The authors thank Editors L. Gangadharan and J. Duffy, two anonymous reviewers for many helpful comments, N. Schmitt, the ESI lab manager, seminar participants at: CSU Fullerton, Universities of Basel, Hamburg, Miami, USC, UC Santa Barbara, Canadian Economics Association 2019 meetings, World Bank ABCDE 2019 conference, 2019 GSE Summer Forum (Pompeu Fabra), 2019 SSES Conference (Geneva), 2018 and 2019 ESA meetings (Berlin, Los Angeles). The authors acknowledge partial research support through the SNSF grant No. 100018.172901; G. Camera acknowledges partial research support through an IFREE Small Grant. Gabriele Camera, Economic Science Institute, Chapman University, One University Dr., Orange, CA 92866; e-mail: camera@chapman.edu. Lukas Hohl, Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland; e-mail: lukas.hohl@unibas.ch. Rolf Weder, Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland; e-mail: rolf.weder@unibas.ch.

1 Introduction

A large body of research asserts that integration of markets generates large overall gains, due to specialization, increased productivity, expanded product variety, and pro-competitive gains and innovation (Costinot and Rodríguez-Clare, 2018; Feenstra, 2018; Rossi-Hansberg, 2017). Yet, after decades of increasing economic integration, there are signs of a desire to scale it back. The backlash in the support of regional and multilateral trade agreements is one indication. The choice of a majority of the British people in 2016 to exit the European Union is another. The question is why: what pushes individuals to limit or scale back economic integration, given that it is costly?[1]

Here, we use an experiment to study a possible contributing factor: cross-country inequality of economic opportunity. What motivates this angle of inquiry are related observations from the trade and experimental literatures. There is evidence that, though the estimated gains from integration are positive, they are also unevenly distributed (Fort et al., 2018; Hakobyan and McLaren, 2016); this may reduce support for it (Stiglitz, 2020, p. 288), as economic inequality can distort decision-making (Cappelen et al., 2014), induce short-sighted conduct (Haushofer and Fehr, 2014), and act as a barrier to cooperation in short-term interactions (Tavoni et al., 2011). Further, the literature on economic integration recognizes that as local markets are scaled up trust and enforcement frictions emerge making it necessary to rely on self-enforcing trading arrangements (Rodrik, 2000); but informal norms of cooperation do not easily scale up (Camera et al., 2013a). We thus hypothesize that cross-country inequality in *prospective* gains might degrade coordination on co-

[1]Brexit may cost up to 10 percent of UK per capita GDP (Sampson, 2017); Chakraborty et al. (2017) estimate the cost of financial disintegration in Europe at half percent of GDP.

operative arrangements to the point where *actual* gains become unattractive, in which case pursuing an integration policy would undermine its economic development objectives.[2]

The experiment is based on an indefinitely-repeated helping game (a supergame) played in consumer-producer pairs with alternating roles. The producer can either cooperate or defect, and is the only decision-maker. By cooperating, she suffers a small cost and the consumer obtains a large *economic benefit*. We induce heterogeneous cooperation valuations by randomly assigning 24 participants to three types (or, countries) each with a different benefit amount (cross-country inequality of opportunity). During a familiarization phase, participants interact in fixed pairs with a single counterpart of their same type—an "isolated" economy. They also interact in mixed groups of 12 strangers where counterparts can be of any type and change at random—an "integrated" economy. There are *potential* gains from integration because cooperation benefits are largest—and possibly heterogeneous—in mixed groups where, however, identities and past behaviors are unobservable. After this phase, we provisionally form a mixed group with all 24 players, and ask participants to express a preference for staying in it (economic integration), scaling it down to 16 by excluding a type, or leave it for a fixed pair (isolation). A majority preference rule determines the economy's configuration. Based on the theory of infinitely repeated games, a self-enforcing norm of conduct supports full cooperation (or, efficient play) in every economy, although many other equilibria exist.

[2]These considerations seem also relevant for smaller-scale integration phenomena. Universities often confront the problem of reorganizing separate (and differently compensated) units by merging them into a larger one, which may lack cohesiveness and perform poorly. Or, consider vertical mergers, when the contribution to value-creation of each firm is heterogeneous, and corporate settings where personnel from different units work in teams.

The larger size, inability to rely on reciprocity and reputation, and the lack of explicit coordination devices create significant strategic uncertainty in mixed groups as compared to fixed pairs. This makes explicit a tradeoff between frictions to cooperation and potential gains from integration because *actual* gains depend on how strangers behave relative to partners. In pursuing integration, participants must weigh its primary economic effects (greater cooperation benefits) against its possible secondary behavioral effects (lower cooperation rate). This is a way to operationalize in the lab the fundamental problem discussed in Rodrik (2000): in the international trading arena cooperation is highly profitable but not easily attainable because individuals must be willing—not only able—to cooperate with each other and face additional frictions that are absent within local borders (e.g., cultural, legal).

To assess the impact of inequality of economic opportunity, we randomly assign subjects to treatments manipulating the baseline distribution of potential gains from integration. Baseline gains are homogeneous so integration does not alter pre-existing inequalities. In two treatments, integration either removes or increases pre-existing inequalities, using mean-preserving spreads of baseline gains. In a final treatment, we increase baseline gains by 70%.

We report four main results. First, participants struggled to capture the possible benefits of integration. A possible reason is that trust and enforcement frictions undermined subjects' ability to develop self-enforcing norms of cooperation (the survey in Bigoni et al., 2020, supports this view). Second, inequality of economic opportunity does not appear to impair cooperation as redistributing or increasing *potential* gains from integration did not improve cooperation. Third, the size of *realized* gains played a primary role in the choice to integrate (not the distribution or size of *prospective* gains). Fourth, though everyone could benefit, integration created few "winners" and many "losers," which induced a negative sentiment toward it. This corroborates the notion that integration may produce outcomes that diverge from those predicted by standard models owing to frictions that limit individuals' ability to realize the gains from trade (see Antràs and Costinot, 2010), and confirms the classic Stolper-Samuelson Theorem—opening up international trade may create losers and winners (Stolper and Samuelson, 1941). We show that these results partly depend on strategic uncertainty preventing coordination on cooperation in mixed groups. In a treatment where participants could coordinate their strategies using a chat box, cooperation in mixed groups significantly improved, albeit an aversion to forming large heterogeneous groups remained.

Our experiment can offer insights into whether distributional effects of economic integration agreements should be part of the policy discussion. It allows us to establish the importance of distributional considerations while minimizing possible confounding factors—political, social and cultural, for instance. An insight is that researchers and policymakers should weigh the primary economic effects of economic integration against its secondary behavioral effects. The promised productivity gains (greater cooperation benefits) may be greatly reduced if integration has adverse spillover effects on cooperative attitudes. Human behavior is not invariant to economic processes and the trading environment, so miscalculating this secondary effect of integration may lead to overestimate its potential benefits. In the experiment, the policy of integration undermined economic development and pushed subjects to reject it; this echoes the message in Dal Bó et al. (2018), where mistaken beliefs about the behavioral effects of a policy switch induced selection of a policy that, though theoretically superior, was empirically inferior. Our results contribute to the growing literature on endogenous institutions for cooperation (see the survey in Dannenberg and Gallier, 2020), and suggest that to better assess outcomes of integration policies, research and policy should account for possible spillover effects of integration on behavior, especially those due to cross-country distributional effects. Miscalculating this behavioral aspect might undermine the intended development and motivate calls for dramatic policy-reversals.

2 Related experimental literature

This study is primarily related to two research themes in the experimental literature: endogenous group formation and cooperation in repeated games. Experiments on endogenous group formation typically consider a public goods game with a fixed number of decision rounds, where theory rules out efficient play. A main finding is that if free-riders can be identified and isolated, then this promotes the formation of large cooperative groups. Typically this requires institutions for monitoring individual conduct, unilaterally joining or leaving groups, or sanctioning low contributors, which are either exogenous (e.g., Ahn et al., 2009; Baland et al., 2017; Cinyabuguma et al., 2005; Croson et al., 2015; Güth et al., 2007; Maier-Rigaud et al., 2010) or endogenous (.e.g, Dannenberg et al., 2019; Gürerk et al., 2006). Manipulating the power structure, by choosing to give to someone the freedom to distribute group earnings also promotes the formation of cooperative groups (Nash et al., 2012). We take the endogenous institutions angle, but neither assume redistributive powers nor institutions for self-selection. A collective decision process allows players to combine pre-existing subgroups into a larger one and random assignment prevents self-selection.[3] Forming a large group opens the door to large gains

[3]Our group formation rule randomly lets one subset of players (one type) select group configuration. Initial random assignment to types implies that if participants have heterogeneous inclinations towards cooperation, then neither choosing "isolation" nor "integration" allows self-selection into a cooperative coalition. As we are interested in economic integration, this ensures that, if we think of the three types as countries, individuals cannot leave their coun-

but magnifies the incentives to free ride as participants interact as strangers, who cannot identify, exclude or prevent free-riders from joining the group. A way to mitigate free-riding problems is to collectively choose to avoid the group, giving up potential gains to interact with a fixed partner.

To situate our design in the literature about cooperation in repeated games, consider two design features: the number of decision rounds (finite/indefinite) and the type of interaction (partners/strangers). A fixed number of decision rounds is common. Here, standard theory predicts inefficient play and, indeed, empirically cooperation is short-lived (Isaac and Walker, 1988a). Instead, as we wish to study cooperation as a long-run phenomenon, we work with an *uncertain* number of decision rounds, which support efficient play as an equilibrium. Fixed counterparts (partners) are also common. Instead, as a way to introduce frictions to impersonal trade, we assume anonymous counterparts that change at random (strangers).

Our design fits into a broader research agenda about institutions for long-run cooperation in groups of strangers and, in particular, four recent studies. The two-person stage game, indefinite repetition and random re-matching process follows Camera et al. (2013a), which studies how cooperation reacts to *exogenous* variation in group size, when payoffs are invariant to group size. Without a monetary institution in place, cooperation fell as groups got larger. Unlike our design, players are homogeneous, their roles randomly alternate, group size is exogenous and does not affect cooperation payoffs. Bigoni et al. (2019) studies endogenous group formation with homogeneous players (vs. heterogeneous in our design) when cooperation benefits identically increase by 20% in groups of 12 or 24 strangers compared to groups of 2 partners (vs. het-

try, or create a new entity comprising arbitrary regions from different countries. Countries can only aggregate into a bigger economic entity, or choose the status quo.

erogeneous increase by 20%-50%). Without a monetary institution in place, participants seldom selected groups of strangers and, if they did, performed poorly. Camera et al. (2020a) studies if fairness and inequality motivations affect cooperation in ex-ante homogeneous economies with 4 players with randomly changing roles. As the game progresses this causes ex-post earnings inequality in the group. Participants conditioned cooperation on their own past realized roles and, when ex-post inequality was made visible, discriminated against those who are better off. Camera and Hohl (2021) studies social identity in repeated games, to understand if categorization alone can influence norms of cooperation. There are three types of group affiliates who either have different cooperation benefits (as we do), or have no payoff-relevant differences. Unlike our study—where individuals cannot discriminate between "insiders" and "outsiders"—cooperation *can be* conditioned on the counterpart's group affiliation but not on the behavior of group affiliates, so players can adopt discriminatory strategies but cannot base them on group reputation. Subjects cooperated similarly with insiders and outsiders.

Unlike the above experiments, we study cooperation among strangers who face *unequal economic prospects*, something largely unexplored in the literature on supergames. Our players are ex-ante heterogeneous in their cooperation benefits and in how these benefits increase with group size. Standard results suggest that this heterogeneity should not interfere with coordination on efficient play.[4] Our findings support this theoretical assessment—inequality of economic prospects did not instigate more opportunism, although it reduced the desire to form large groups.

[4]This means that in all treatments and all groups a strategy always exists that, if adopted by the entire group, can support the efficient outcome as an equilibrium. According to this strategy, which is discussed in Section 4, a player should fully cooperate in equilibrium, and should choose to form mixed groups if given the chance.

Our study also contributes to an experimental literature about the effect of heterogeneity on voluntary contributions in public goods games. In some experiments, heterogeneity is introduced in participants' endowments (e.g., see Buckley and Croson, 2006; Chan et al., 1999; De Geest and Kingsley, 2021; Hargreaves Heap et al., 2016; Isaac and Walker, 1988b; Sadrieh and Verbon, 2006), and its effects on cooperation vary from none to negative. A meta analysis of linear public good experiments (Zelmer, 2003) reports a negative and significant effect of heterogeneous endowments. In other experiments, participants have different valuations for cooperation. Here, too, heterogeneity has mixed effects. It does not impact cooperation in Reuben and Riedl (2013), it lowers unconditional contributions in Fischbacher et al. (2014), and in Kölle (2015) contributions are lower in groups with heterogeneous vs. homogeneous valuations, with low valuation types contributing less than high-valuation types. Heterogeneity disrupts cooperative play even when punishment and reward mechanisms are available (Nikiforakis et al., 2012; Reuben and Riedl, 2009), and even if players can coordinate via free-form communication (Dekel et al., 2017; Gangadharan et al., 2017).

These mechanisms lose effectiveness as subjects tend to prioritize equality over efficiency, leading to under-punishment, counter-punishment, or norms that side-line efficiency in favor of redistribution. As in these experiments players interact with few partners whose behavior and type can be observed, reciprocity and the ability to discriminate—which are impossible in our design—might play a role in how heterogeneity affects cooperation. Our study is also related to public goods experiments about the trade-off between equity and efficiency, where heterogeneous players vote on redistribution or contributions rules (e.g., see Balafoutas et al., 2013; Gallier et al., 2017). Unlike those studies, we do not study collective choice about redistribution of income, although we do study what happens when subjects can *informally* coordinate their strategies using a chat box. Finally, we contribute to expand the use of experimental methods in international economics, following in the footsteps of the pioneering market experiment in Noussair et al. (1995) that tested the competitive trade model, and the social dilemma experiment in Barrett and Dannenberg (2012) about international collective action.

3 Experimental design

We construct an intertemporal trading game based on Camera et al. (2013a) and Bigoni et al. (2019), where trust and enforcement problems create incentives for short-sighted conduct. Twenty-four participants are randomly assigned to three types $i = 1, 2, 3$ (8 per type) for the entire session. The experiment has many rounds. In each round subjects interact in pairs, which may be fixed or randomly changing in every round. This is explained below, where we discuss the Neutral treatment, which is our baseline.

Interaction in a round. In each round all subjects are in a pair. Each pair comprises a *producer* and a *consumer*. The producer chooses whether to engage in a transaction (cooperate) or not (defect); the consumer has no action to take. By defecting, the producer obtains 6 points while the consumer earns 3 (1 point= USD 0.18); see Table 1. By cooperating, the producer earns nothing, while consumer i earns $9 + 2i + y$ points, where y depends on whether pairs are fixed or not (more later). Hence, cooperation creates $2i + y > 0$ *surplus*, which is entirely earned by the consumer. This payoff matrix is known to all players. However, the producer cannot observe the consumer's type, while the consumer can see the producer's type.

This design has two implications. First, since surplus increases in the consumer's type i, we say that type 1 is "disadvantaged" relative to others, while type 3 is "advantaged."[5] Interpreting a type as a country, this implies cross-country *inequality of opportunity* due to the 4-point gap in cooperation payoffs between advantaged and disadvantaged consumers.

Table 1: Payoffs in a Pair (type $i = 1, 2, 3$).

	Producer's Choice	
	Defect	*Cooperate*
Consumer of type i:	3	$9 + 2i + y$
Producer (any type):	6	0

Second, for cooperation to be incentive-compatible, producers must have a prospect of future benefits. This is done by letting participants interact for at least 18 rounds and with alternating roles (producer, consumer, ..., and vice-versa). Participants are informed of this—they know their role and action set switches in every round.[6] Starting with round 18, we use a random stopping rule: a new round is played with probability $\beta = 0.75$ using computer randomization (see instructions and post-instruction quiz in Appendix B). This random sequence has mean duration and standard deviation of 21 (=17+1/0.25) and 3.5 rounds, respectively, and is called a "supergame."

[5] In the experiment types are color-coded (green, red, and blue). A related experiment uses Neutral data (Phase 1, only) to study if the assignment to types artificially induces group-identity that constitutes a psychological basis for intergroup discrimination (see Camera and Hohl, 2021, footnote 9). No evidence of group effects from the initial categorization into color-coded types is detected.

[6] This is unlike most cooperation experiments, which use a synchronous cooperation task (e.g., a PD). Our task is asynchronous—a repeated helping game with alternating decision-makers who experience a time delay between cost of and benefit from cooperation. This creates a sharper distinction between isolated and integrated economies as it removes strategic uncertainty in fixed pairs (see the discussion in Bigoni et al., 2019). It also emphasizes the *intertemporal* nature of cooperation, thus capturing the spirit of the dynamic macroeconomics and trade literature, where the benefits of trade derive from an intertemporal exchange of goods and services.

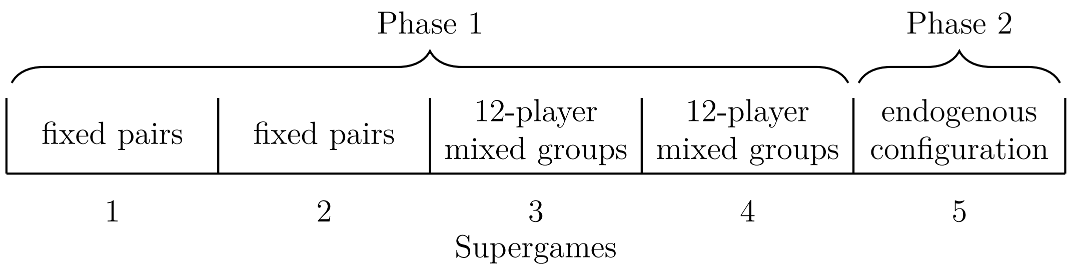
A supergame. At the start of a supergame, participants are randomly assigned to a matching group, or *economy*. We study two economy configurations, *isolated* and *integrated*. An isolated (autarkic) economy is small, consisting of a *fixed pair* where players are of the same type and $y = 0$. An integrated economy consists of a heterogeneous *mixed group* with four players per type (12 in all) and $y = 3$. Here, in each round players are randomly rematched with uniform probability and as strangers so that the producer cannot see ID, type and past actions of the counterpart; hence, cooperation cannot be conditioned on the consumer's type. At the end of each round players are informed if their economy attained full cooperation or not.[7] As compared to isolated economies, integrated economies preclude reciprocity and reputation-building but offer larger cooperation benefits because $y = 3$ (instead of zero), a value we refer to as the *potential gains from integration*. This design captures the key tradeoff of economic integration that motivates our study: it has the potential to benefit everyone, but it also induces trust, enforcement and coordination frictions due to larger size, no reputation building and anonymity.

A session. There are two phases; see Fig. 1. Phase 1 familiarizes participants with isolated economies in supergames 1-2, and integrated in 3-4 (we also study the reverse order). Matching across supergames is pre-arranged to minimize spillover effects: participants are informed that they will not meet counterparts from previous supergames. Phase 1 offers subjects an experiential basis on which they can base their endogenous configuration choices in Phase 2, which is when they must choose size and composition of the economy for supergame 5, the last one. Before this supergame starts, participants are

[7]This information is anonymized, see the results screen in the instructions in Appendix B, column "Same Outcome In All Pairs." We provided it to ensure that the economy size did not affect the condition for existence of the efficient equilibrium–as discussed in Section 4.

informed that they are all provisionally included into a 24-player mixed group.

Figure 1: A Session.



Notes: Four sessions had Phase 1 as above, and four the reverse order (each treatment). Our matching process ensured the same consumer-producer pair could not be formed in different supergames except for a mixed group of Phase 2; details are in Supp. Mat.

Then, everyone privately expresses a preference for either (i) maintaining the provisional configuration ("stay"); (ii) reducing the mixed group to 16 players by matching one type into fixed pairs ("exclude"); or (iii) leaving the mixed group for a fixed pair ("leave"). After choosing, a computer randomly selects a player type with equal probability. The majority choice within that type determines the economy configuration in supergame 5. Two outcomes are possible: (i) a 24-player mixed group (economic integration) or, (ii) a 16-player mixed group comprising two types, and four fixed pairs of the remaining type.[8] The payoff structure is the same as in Phase 1. This selection procedure is democratic from an ex-ante perspective, as it gives equal weight to each individual and type.[9] The payoff structure ensures that integration is socially

[8]Suppose type 1 players are selected to determine the economy configuration. If their majority chose to exclude type 3, we form a mixed group with types 1 and 2, and four fixed pairs of type 3. If the majority chose to leave, we form a mixed group with types 2 and 3, and four fixed pairs of type 1. The computer resolves ties via a coin flip. The instructions explain that participants would have an opportunity to alter size and composition of the economy in supergame 5, in a manner specified at the end of supergame 4. Given the default assignment to a single mixed group consisting of all 24 session participants, a "leave" choice can be interpreted choosing economic disintegration.

[9]Counting all choices in the session would have been also problematic for two reasons. It

efficient as it maximizes potential earnings for everyone.

Treatments. Subjects are randomly assigned to four treatments, which differ in just one aspect: the distribution of y in mixed groups (see Table 2), i.e., the distribution of the potential gains from integration. The value of y is always zero in fixed pairs of all treatments.

In our baseline, potential gains from integration are identical for everyone, $y = 3$ in mixed groups. We call this baseline the Neutral treatment because economic integration increases potential surplus without altering the pre-existing inequality of economic opportunity: advantaged consumers can always earn four more points than the disadvantaged (in mixed groups and in fixed pairs).

Table 2: Main Treatments—Cooperation Payoffs in Mixed Groups.

	Potential Gain from Integration (to type 1,2,3)			Cooperation Payoffs (to type 1,2,3)			Potential Surplus	Inequality of Opportunity
Neutral	3	3	3	14	16	18	baseline	baseline
Converge	5	3	1	16	16	16	baseline	removed
Diverge	1	3	5	12	16	20	baseline	doubled
Neutral+	5	5	5	16	18	20	increased	baseline

Notes: *Potential gains from integration* corresponds to the value y in mixed groups; this value is type-dependent in some treatments. The *Cooperation Payoff* is the cooperation payoff to a consumer of type i, which is the sum of the payoff in fixed pairs, $9 + 2i$, plus the potential gain from integration, y. Inequality of opportunity is the difference in cooperation payoffs for consumers of types 3 and 1.

The Converge treatment redistributes potential gains top to bottom without altering potential surplus: $y = 5, 3, 1$ for type $i = 1, 2, 3$ in mixed groups. Here, integration removes inequality of opportunity as cooperation payoffs con-

would have precluded the possibility of "leaving" ever being the majority choice, and it would have increased the complexity of determining an outcome (e.g., type 1 wants to exclude 2, type 2 wants to exclude 3, and 3 wants to exclude 1).

verge to 16 points for everyone. The opposite Diverge treatment redistributes potential gains iniquitously, $y = 1, 3, 5$. Here, integration leads to a greater divergence of cooperation payoffs, 8 points instead of 4 in fixed pairs. Finally, the Neutral+ treatment boosts y to 5 points for everyone in mixed groups, a 67% increase over the 3-points baseline.

Summing up, the baseline distribution of potential gains from integration is manipulated either via mean-preserving spreads or by shifting it to the right. In Converge and Diverge, potential surplus from integration is redistributed to remove or increase inequality of opportunity. In Neutral+, potential surplus rises across the board. As the structure of fixed pairs is treatment-invariant, these isolated economies offer a stable reference point for subjects in evaluating the performance of integrated economies.

As explained below, all economies exhibit equilibrium multiplicity. This creates significant strategic uncertainty in mixed groups. Previous work suggests this is an obstacle to coordination on cooperation and, consequently, the formation of large groups (Bigoni et al., 2019). We thus add a final treatment, Neutral-Chat were subjects can explicitly coordinate their actions through a chat-box, to study if the possibility to overcome strategic uncertainty problems affects the results on group formation.

Experimental procedures. The experiment was conducted at the Economic Science Institute's laboratory at Chapman University and involved 864 undergraduates recruited between 2/2017 and 10/2019. Subjects were recruited using the online proprietary system developed at the Economic Science Institute. We ran 8 sessions per treatment except for Neutral-Chat, for which we ran only four session. Each session has 24 participants, which are randomly assigned to a type. As a result there are 192 subjects per treatment except for Neutral-Chat, which has 96 subjects. No subject had previous experience with this game, 43% were males. On average, participants were paid about USD 32, including a show-up fee of USD 7 and about USD 2 from an incentivized quiz to test subjects' understanding of the instructions before the start of the experiment. The average duration of a session was 1 hour and 40 minutes. Instructions were recorded in advance and played aloud at the beginning of a session, participants had also individual paper copies on their desks (see Appendix B). We used neutral language for the instructions (words like "cooperation" or "help" were never used). The instructions informed players that only one of the five supergames would be randomly selected for payment, with public random draw at the end of the experiment. The points earned in that supergame were converted into dollars according to a pre-announced conversion rate of USD 0.18 per point. After the instructions, subjects were given the incentivized quiz to test their understanding of the instructions (see Appendix B). The experiment was programmed using the software z-Tree (Fischbacher, 2007). No eye contact was possible between participants and supergames started and ended simultaneously for everyone. We collected demographic data via post-session anonymous surveys.

4 Theoretical predictions

Here we demonstrate that full cooperation is an equilibrium in all treatments and all economies. Based on this, we then formulate three testable hypotheses. The focus on full cooperation is motivated by the following observations:

Observation 1. *Full cooperation is Pareto optimal in all economies.*

Observation 2. *Economic integration maximizes theoretical efficiency.*

To see this, consider two reference outcomes: *full cooperation* (every producer always cooperates) and *full defection* (no producer ever cooperates). Outcomes in between, where some but not all producers cooperate all the time, exhibit *partial cooperation.* Defection gives 6 points to a producer and 3 to a consumer. Consumers earn $9 + 2i + y$ points if cooperation occurs, while producers earn nothing. Hence, full defection generates a total profit of 6+3=9 points in every pair, of any economy, of any treatment. Instead, full cooperation generates $9 + 2i + y$ points in a pair where the consumer is of type i, where $y = 0$ in isolated economies, and $y > 0$ in integrated economies. The implications are, first, full cooperation is Pareto optimal as it creates $2i + y \geq 2$ surplus in all pairs. Second, since $y > 0$ only in mixed groups, economic integration expands the theoretical efficiency frontier relative to isolated economies. Hence, forming a mixed group composed of all three player types—choosing to economically integrate, that is —is the only way to maximize potential surplus and, hence, theoretical efficiency in Phase 2. Third, integrated economies are inherently more efficient, in that they can create as much surplus as isolated economies with a lower cooperation rate. This implies that full cooperation is not necessary for integration to be economically beneficial. However, there is a danger: partial cooperation might create "winners" and "losers," because, depending on the distribution of potential gains, some type may end up benefitting from integration while another may suffer.

Summing up, in our design full cooperation corresponds to efficient play. Subjects can maximize payoffs by fully cooperating in every economy, and choosing a mixed group composed of all three player types, when given the option. The question is thus: is efficient play part of an equilibrium?

Proposition 1. *Full cooperation is a sequential equilibrium in all economies and all treatments.*

Proof. See Appendix A □

The proof relies on a version of Kandori (1992, Proposition 1), extended to the case of heterogeneous players. Consider the following trigger strategy: a player always cooperates as a producer, but will forever defect if some producer defects. At the end of each round players see whether their economy coordinated on full cooperation, or not. Hence, defecting in cooperative equilibrium can be used to trigger an immediate and permanent sanction: always defect. This sanction is an equilibrium in the continuation game since defection is a best response to everyone else defecting; so, it is incentive-compatible for a player to follow this sanction if everyone is expected to do the same. A producer of type i has no incentive to deviate in equilibrium if $\beta \geq \beta^* := 6/(6 + 2i + y)$, which—given the experimental parameters—is satisfied in all treatments and economy configurations. Hence, full cooperation is part of a sequential equilibrium in our laboratory economies. By design, the threshold value β^* is lowest in mixed groups because cooperation payoffs are the largest (see Table A1 in Appendix A), indicating that, all else equal, integrated economies provide the strongest monetary incentive to cooperate.

Hypotheses. Many equilibria exist in our setup and standard theory cannot address selection issues without additional assumptions. A common assumption is that players are rational and seek to maximize their payoffs. If so, they should attempt to coordinate on full cooperation in all economies, and should select full integration in all treatments. To see why, note that the economic incentives to coordinate on efficient play are aligned across subjects (full cooperation is the Pareto-dominant equilibrium) and are strongest in mixed groups, where cooperation generates the highest benefit to *every* player. Hence, theory suggests that full cooperation should be a focal outcome in every economy, and that subjects should seek economic integration when given a choice. Moreover, behavior should not be affected by the distribution of cooperation payoffs in mixed groups. This leads us to put forward three hypotheses:

H 1. *Treatments should not alter cooperation rates in mixed groups relative to the baseline.*

H 2. *Treatments should not alter the distribution of economy configuration choices relative to the baseline.*

H 3. *Integration should be the majority choice in all treatments.*

Empirically, strangers tend to cooperate below the theoretical maximum (Bigoni et al., 2019; Camera et al., 2013a). Hence, H1 states that cooperation rates in mixed groups should be the same across treatments, without predicting full cooperation. This hypothesis immediately extends to fixed pairs, as their payoff structure is treatment-invariant. Similarly, H2-H3 state that economy configuration choices should not vary based on treatment, and economic integration should be the most frequent for all player types.

Our hypotheses hinge on assuming purely rational, self-interested players. Alternative predictions could be formulated based on theoretical platforms that incorporate other assumptions, such as fairness and inequality aversion in Fehr and Schmidt (1999), Bolton and Ockenfels (2000), or Rabin (1993). These models give rise to an explicit trade-off between efficiency and equity, which is relevant to our model, as full cooperation supports payoff heterogeneity. If fairness motivations dominate efficiency motivations, then it is conceivable that disadvantaged players might be unwilling to fully cooperate, to reach a more balanced payoff distribution; by occasionally breaking the cooperation norm a player raises her payoff at the expense of the average counterpart. Advantaged players with strong aversion to unequal payoffs might tolerate some free-riding, cooperating even after suffering a defection to avoid sanctioning counterparts who, on average, are less fortunate. These behaviors address inequality concerns at the expense of efficiency, crowding-out the gains from cooperation. Hence, an alternative conjecture is that manipulating the distribution of potential gains might affect subjects' motivation to fully cooperate in mixed groups, and to seek economic integration.[10]

Hence, the alternate hypothesis to H1 is that cooperation in mixed groups will be affected by treatment, although we cannot predict a specific direction. If cooperation is affected, then this might alter the economic incentives to interact in mixed groups and affect the motivation to seek integration. Hence, the alternate hypotheses to H2-H3 are that treatments will affect the distribution of economy choices, and "stay" might not be the majority choice when we move away from the baseline treatment.

5 Results

Here we first focus on the four main treatments, studying cooperation and efficiency in Phase 1, and then the economy configuration choices in Phase 2. We conduct analyses at the session level, economy level, and individual level.[11] In a final section, we then consider the impact of strategic uncertainty, studying the Neutral-Chat treatment. Appendix B contains technical details.

[10]Avoiding integration reduces potential surplus, while conditioning cooperation on (own) type lowers the player's payoff in equilibrium because uncooperative actions trigger defections; off equilibrium, unconditional defection is a best response to others defecting. Hence, players who are income-maximizing or who seek to maximize efficiency have no economic reason to coordinate on less than full cooperation and to avoid integration.

[11]For session-level analysis we have 8 independent observations per treatment, used to perform statistical tests. For economy-level analysis we have 192 observations for fixed pairs and 32 observations for mixed groups, in Phase 1, per treatment. Here, the only truly independent observations are from supergame 1 (2 for mixed groups and 12 for fixed pairs) while the others may be correlated within a session; overall, in Phase 1 of each session we have 4 (not independent) mixed groups and 24 (not independent) fixed pairs. These possible correlations are mitigated by our strangers matching protocol, but are not entirely eliminated. We address possible interdependencies by means of regression analysis with cluster-robust standard errors, with clustering at the session level. Similar considerations apply to subject-level analysis, where we have 192 observations per treatment.

5.1 Cooperation in Phase 1

Here we focus on the average cooperation rate in an economy (calculated in Appendix B) because, together with the economy configuration, is what determines the economy's performance. To explain, let *profit* denote the points earned by a subject in the average round of a supergame; it depends on: (i) the subject's cooperation rate, (ii) the choices of the subject's producer counterparts, and (iii) the payoff matrix applying to the economy. Per-capita profits in the economy are directly proportional to average cooperation, going from a minimum of 4.5 points (no cooperation—any economy) to between 5.5 and 7.5 points (full cooperation—fixed pair, depending on type of players), to 8 or 9 points (full cooperation—mixed group, 9 only in Neutral+).

Result 1. *Redistributing or increasing potential gains did not alter cooperation rates in mixed groups, which remained as low as in the Neutral baseline.*

Cooperation rates are similarly low in all treatments. The mean cooperation rate in a mixed group is 0.397, 0.384, 0.333, and 0.421 in, respectively, Neutral, Converge, Diverge and Neutral+ (with s.d., respectively, 0.198, 0.185, 0.176, and 0.192). Furthermore, cooperation rates decline as the supergame progresses, in all treatments; see Fig. B1 in Appendix B.

Using a session as the independent unit of observation, we fail to reject the null hypothesis of identical cooperation rates for any treatment pair, at the 10% significance level (two-sided Wilcoxon rank-sum tests with exact statistics N1=N2=8); this holds true also if we consider only endogenously formed mixed groups in Phase 2; see Appendix B. A GLM regression that controls for individual characteristics confirms this view; we can only detect a difference at the 10% level between Diverge and Neutral+; see the marginal effects reported in col. 1 of Table B3 in Appendix B. This assessment is robust to considering only Phase 2 data (see col. 2 of Table B3 in Appendix B), and is also robust to considering only the first round of play in mixed groups.[12]

Based on this evidence we cannot reject H1 for mixed groups. Removing or increasing inequality of opportunity did not impact overall cooperation rates. There is also no significant effect when we equally increased the economic incentive to cooperate in mixed groups by about 70%, in Neutral+. The largest difference in cooperation (and, hence, efficiency) is 0.065 (Neutral vs. Diverge). By means of comparison consider that in a public good experiment, Kölle (2015) finds that social efficiency falls by 0.17 when moving from groups with a symmetric to an asymmetric payoff matrix. A natural question is thus how small a treatment effect we could detect given our design.[13] We thus perform a two-sample mean power analysis using one mixed group in Phase 1 as the unit of observation ($N = 64$ for our balanced sample, 32 per treatment). We ran one-sided tests where the control group mean and s.d. comes from Neutral, while the s.d. of the treated group comes from the other three treatments. The null hypothesis is that mean cooperation should be unaffected by treatment. The alternative hypothesis is that cooperation should increase in Converge and Neutral+ (less payoff inequality and larger payoffs, respectively), and fall in Diverge (more payoff inequality). Given a 10% significance level, the smallest detectable effect sizes are about 0.12 and 0.10 for a power of 0.9 and 0.8, respectively (0.14 and 0.12 if two-sided tests). This suggests that our design has enough power to detect economically meaningful distributional effects—an increase in cooperation of at least 10 percentage points—and not minor effects.

[12] In Phase 1, round 1 cooperation rates are 0.46, 0.60, 0.44 and 0.55, for Neutral, Converge, Diverge and Neutral+. Using two-sided Wilcoxon rank-sum tests, we can only reject the hypothesis that round 1 cooperation rates are similar in Converge vs. Diverge (p-value=0.0995, N1=N2=8). Using a logit panel regression, we cannot reject the null of identical round 1 cooperation rates for any treatment comparison; see Appendix B.

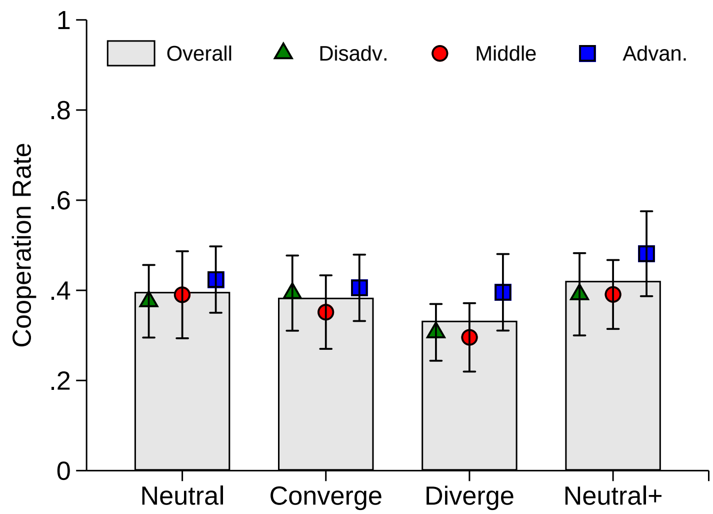
[13] We thank an anonymous Referee for suggesting this additional analysis.

It is natural to ask if the lack of treatment effects stems from data aggregation hiding possible offsetting impacts on the different types. We can reject the hypothesis that treatments altered behavior of player types in mixed groups.

Result 2. *Redistributing or increasing potential gains did not alter cooperation rates of any player type in mixed groups, relative to the* Neutral *baseline.*

Fig. 2 reports mean cooperation rates in mixed groups of Phase 1, by player type, and overall.

Figure 2: Cooperation by Player Type in Mixed Groups (Phase 1)



Notes: One obs.=one mixed group. The markers identify the mean, the whiskers identify the 95% confidence interval. Moving left to right, mean cooperation rates of disadvantaged types are 0.38, 0.39, 0.31, 0.40, for middle are 0.39, 0.35, 0.30, 0.39, and for advantaged are 0.42, 0.41, 0.40, 0.48.

We fail to reject the null that cooperation of a given type is identical for all three relevant comparisons at the 10% significance level (two-sided Wilcoxon rank-sum tests with exact statistics, N1=N2=8, one observation = one session). A GLM regression confirms that neither redistributing nor increasing potential gains affected the conduct of any player type, relative to NEUTRAL; see the treatment coefficients in Table B5 in Appendix B. Hence, we cannot reject H1 even when focusing on specific player types: reshuffling potential gains to remove inequality of opportunity, increasing it, or boosting potential gains for all did not affect cooperation in mixed groups.

Within each treatment, different types cooperated similarly except in Diverge. Here, potential earnings strongly diverged and advantaged players (type 3) cooperated significantly more than everyone else in their mixed group by about 8 to 10 percentage points (see Table B7 in Appendix B). These findings are robust to inclusion of Phase 2 data (see Tables B6 and B8). This suggests that inequality factors *can* influence cooperation choices within the economy, and for this to happen inequality must be sufficiently pronounced. An important question is whether subjects benefitted from interacting in mixed groups instead of fixed pairs. This is discussed in the next two results.

Result 3. *Subjects cooperated less in mixed groups than in fixed pairs.*

Mean cooperation rates in fixed pairs of Phase 1 are 0.776, 0.773, 0.656, and 0.668 in, respectively, Neutral, Converge, Diverge and Neutral+. Using a session as the independent unit of observation, we can only reject the null hypothesis of identical cooperation rates for Neutral vs. Neutral+, and Converge vs. Neutral+ (two-sided Wilcoxon rank-sum tests with exact statistics, N1=N2=8, p-values=0.0225, 0.0207); no statistically significant difference emerges when we consider only Phase 2 data (Table B1 in Appendix B). Using a GLM regression that controls for individual characteristics, we cannot reject the null hypothesis of identical cooperation rates across treatments, except for Neutral vs. Diverge (Phase 1 data only, see Table B9 in Appendix B). Finally, note that, unlike mixed groups, fixed pairs exhibit a stable cooperation trend in all treatments; see Fig. B1 in Appendix B.

Comparing fixed pairs to mixed groups in Phase 1, we reject the null of identical cooperation at the 1% significance level in all treatments (Wilcoxon signed rank tests on matched observations about mean cooperation in fixed pairs and in mixed groups, one observation is one subject in a session, N=192 per treatment). This shows that subjects coordinated on similarly high cooperation rates in fixed pairs (for the most part), and similarly low rates in mixed groups, in all treatments. The outcome of these cooperation differences is described next.

Result 4. *Economic integration did not improve economic outcomes in Phase 1, except in Neutral+, and created winners and losers in all treatments. Realized gains from integration were generally positive for the disadvantaged, and negative for the advantaged.*

In our Neutral baseline, *potential* gains from integration are 1.5 points per-capita, as each consumer can earn 3 more points relative to fixed pairs. Average potential gains are still 1.5 points in Converge and Diverge, but vary depending on the player type (from 2.5 to 0.5 points, see Table 2). In Neutral+ they are 2.5 points for every player.

Realized gains from integration in Phase 1 are well below potential in all treatments, and in some cases even negative. The difference between average profits in mixed groups and fixed pairs is -0.187, -0.207, -0.0956, and 0.501 in, respectively, Neutral, Converge, Diverge and Neutral+. We reject the null of identical payoffs in fixed pairs and mixed groups at the 1% level in Neutral and Neutral+, and at the 5% level in Converge (Wilcoxon signed rank tests on matched observations about mean cooperation in fixed pairs and in mixed groups, in Phase 1, one observation is one subject in a session, N=192 per treatment). This assessment is confirmed by linear regressions that control for order of play and individual characteristics (Table B10 in Appendix B).

Table 3: Realized Gains from Economic Integration (Phase 1)

	Neutral	Converge	Diverge	Neutral+
Disadvantaged	+0.39 (0.250)	+0.57 (0.039)	−0.19 (0.109)	+0.98 (0.008)
Middle	−0.16 (0.547)	−0.24 (0.312)	+0.06 (0.742)	+0.72 (0.039)
Advantaged	−0.78 (0.312)	−0.95 (0.023)	−0.15 (0.742)	−0.20 (0.461)
Total	-0.187	-0.207	-0.0956	0.501

Notes: One obs.=one type of player in a session, Phase 1. We report mean realized profits in mixed groups minus fixed pairs, for each player type (first three rows), and as an average for the economy (*Total* row); in parentheses p-values from signed-rank tests on matched observations (mixed groups vs. fixed pairs, N1=N2=8). Maximum theoretical gains are 1.5 and 2.5 points in Neutral and Neutral+; 1.5 for Middle players in all treatments; 2.5 and 0.5 points for Advantaged and Disadvantaged in Diverge, and the reverse in Converge.

Table 3 breaks down realized gains in Phase 1, by player type. A positive value means that integration is empirically beneficial *for that type*. Realized gains are about zero for every type in Neutral and Diverge. They are negative for the advantaged and positive for the disadvantaged in Converge. They are positive for most types only in Neutral+, where cooperation benefits increased for everyone relative to Neutral. We establish the statistical significance of these observations using Wilcoxon signed rank tests on matched observations (fixed pairs and mixed groups). The unit of observation is one type in a session (N=8 per type, per treatment). We reject the null of identical mean profits for the disadvantaged in Converge and Neutral+, the advantaged in Converge and middle players in Neutral+ (p-values in Table 3).

Summing up, integration in Phase 1 was economically unsuccessful except in Neutral+, where small gains were obtained. Moreover, integration separated types into winners and losers, increasing economic heterogeneity.[14]

[14] In all treatments, there are also winners and losers within the same player type. If we interpret a type as a country, then this observation would seem consistent with interna-

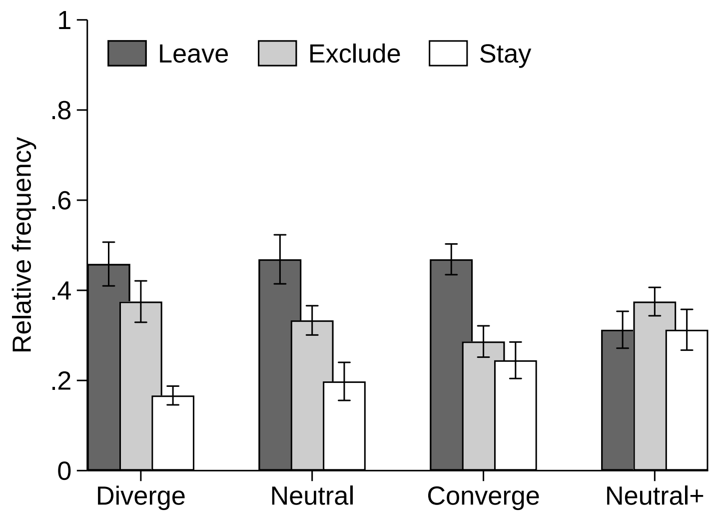
5.2 Economy configuration choices

In Phase 2, participants expressed their preference for the economy to be implemented in the last supergame. We report the following:

Result 5. *Economic integration was not the majority choice, but was chosen more frequently when potential gains grew.*

Fig. 3 reports (relative) frequencies of economy configuration choices.

Figure 3: Economy Configuration Choices (Phase 2).



Notes: One obs.= one session ($N = 8$ per treatment). Relative frequency of choice "leave", "exclude" and "stay." The whiskers identify the mean standard error.

"Exclude" pools the two exclusion choices available to a player, each directed at excluding either one of the two other types. Hence, we have three possible choices: "stay" in the 24-player mixed group (economic integration), scale it down to 16 players by excluding one type, or "leave" to interact in a fixed pair (isolated economy). Three observations stand out. First, in the

tional trade theories. Yet, these differences within a type are due to unequal cooperation rates, with frequent cooperators being "losers" and frequent defectors "winners."

Neutral baseline, choice frequencies are inversely related to the economy size implied by that choice. “Leave” is the most frequent choice (0.47), followed by “Exclude” (0.33) to interact in a 16-player economy, while “Stay” is the least preferred (0.20). This pattern is largely unaffected in treatments where we redistribute potential gains from integration. There is evidence of a treatment effect only when potential gains increase across the board, in Neutral+, where the choices are equally frequent.[15]

The preference for “exclude” over “stay” is interesting because subjects had no experience with 16- and 24-player groups. They knew that scaling the group down to 16 players would not affect their cooperation benefit, but would reduce that of the excluded type—which is socially inefficient. So why did subjects go for “exclude” instead of “stay”? A possible reason is a desire to mitigate strategic uncertainty, which increases as the number of possible counterparts increases.

Redistributing potential gains from integration does not affect the preference for isolation. It simply affects the preference for the size of mixed groups. As potential gains flows from the advantaged to the disadvantaged, the difference between “exclude” and “stay” drops from 0.21 (in Diverge) to 0.05 (in Converge). “Leave” is unaffected. In fact, “leave” remains the majority choice unless potential gains jump for everyone—though doing so did not significantly increase the frequency of “stay” relative to the baseline case.[16]

[15]Using a session as an independent observation, we can reject the null hypothesis of equality of frequencies of “leave” and “stay,” against the alternative that “leave” is more frequent in Converge, Neutral, and Diverge (one-sided signtest, p-values=0.0352, 0.0078, 0.0039 , N1=N2=8) but not in Neutral+ (p-value=0.6367). We can reject the null of equality of frequencies of “leave” and “exclude,” against the alternative that “leave” is more frequent for Neutral and Converge (p-values=0.0352 in each case) but not Diverge and Neutral+ (p-values=0.6563, 0.9648). We reject the null of equal frequencies for “exclude” and “stay,” against the alternative that “exclude” is more frequent in Neutral and Diverge (p-values=0.0625, 0.0156) but not in Converge and Neutral+ (p-values=0.5000, 0.1445).

[16]Using two-sided Wilcoxon-Mann Whitney ranksum test with exact statistics we cannot

Based on this evidence, we can reject H2. The results suggest a strong attractiveness for interaction that can be based on reciprocity, which was unaffected by the redistributive interventions in the experiment. However, we do see a significant shift in preferences away from isolation and towards integration, in Neutral+. Does this mean that a general improvement in *potential* gains from integration is sufficient to make it more attractive?

Result 6. *The economy configuration choices reflected variation in* realized *gains from integration, not in size or distribution of* potential *gains.*

We disentangle the effect of *realized* and *potential* gains from integration with a multinomial logit regression that controls for realized gains. We regress a subject's economy choice on *Realized Gains*, which is standardized. A categorical variable soaks up treatment effects (Neutral is the base case), which is interacted with *Realized Gains*. An *order* dummy and standard controls are included. Marginal effects are summarized in Table 4.

Realized gains have a primary influence on economy configuration choices.[17] One standard deviation increase in realized gains significantly lowers the probability of choosing "leave" by 19 percentage points, while "exclude" or "stay" grow by 7 and 12 points; see the coefficients on *Realized Gains* (columns 2 and 3 are statistically similar, Wald test, p-value=0.560). Subjects rationally responded to economic incentives, basing their choice to seek integration or

reject the null of equal frequency of "leave" choices in any of the first three treatments, at the 10% level. We can reject that the difference between "exclude" and "stay" frequencies is equal in Diverge and Converge (p-value=0.0959, N1=N2=8). We can reject the null of identical "leave" choices in Neutral and Neutral+, but not "stay" (p-values=0.047,0.165).

[17] Participants' realized gains depend on their relative cooperativeness, with cooperators being more likely to have losses from integration and free-riders gains. The reason is that cooperators are more easily exploited by free riders in mixed groups (vs. fixed pairs) because strangers' interaction prevents identification and directly sanctioning of free riders. This issue is studied in Bigoni et al. (2019, p.210-11), which provides evidence on this point and reveals that—as self-selection is ruled out—free riders more frequently select large groups, while cooperators fixed pairs.

isolation on *realized* earnings, not potential earnings. Once we control for this, choices are unaffected by treatment manipulations. All coefficients on treatment dummies are insignificant, meaning that manipulating the distribution or overall size of *potential* gains did not significantly affect choices.

Table 4: Economy Configuration Choices–Marginal Effects

Dep. variable= Economy choice	Leave (1)	Exclude (2)	Stay (3)
Treatment			
Converge	-0.031	-0.019	0.051
	(0.070)	(0.044)	(0.054)
Diverge	0.002	0.030	-0.032
	(0.059)	(0.047)	(0.038)
Neutral+	-0.094	0.028	0.066
	(0.062)	(0.039)	(0.048)
Realized Gain	-0.190***	0.070	0.120***
	(0.044)	(0.053)	(0.044)
Order (12-12-2-2)	-0.133***	0.001	0.132***
	(0.042)	(0.036)	(0.036)
Controls	Yes	Yes	Yes
N	768	768	768

Notes: Multinomial logit regression on preferences for choices of endogenous configuration. One observation = one subject in a session. Robust standard errors (in parentheses) adjusted for clustering at session level. The categorical variable *Treatment* corresponds to the four treatments (*Neutral* is the base case). *Realized Gains* is the standardized difference in earnings between mixed groups and fixed pairs of Phase 1, and is interacted with *Treatment*. Controls at the individual level consisting of the subject's sex and our two standardized measures of understanding of instructions (response time and wrong answers in the post-instruction quiz). Marginal effects are computed at the regressors' mean value (at zero for indicator variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Summing up, economy configuration choices are *not* elastic to the manipulations of *potential* gains from integration that we considered. Seen this way, inequality of *opportunity* or the size of prospective gains do not seem to be the primary reasons behind the choice to seek or avoid integration, in the experiment. A key element was subjects' personal experience with integration in

Phase 1. A negative experience pushed them towards isolation, and a positive towards integration. Except for Neutral+, only a minority of players experience positive realized gains, which explains why in all treatments but Neutral+ the majority choice was to steer clear of economic integration.[18]

Finally, we document choices of the different player types. We ask: once we account for realized gains, was economic integration chosen with similar frequency by the different player types? If not, what biases do we observe?

Result 7. *In each treatment, the choice to "stay" was similar across types.*

We ran multinomial logit regressions using one subject in a session as the observation unit. We regress a subject's group choice on the continuous regressor *Realized Gains*, an *order* dummy, the standard controls used earlier. We use the "type" factor variable to assess differences between player types (Disadvantaged is the base case). The coefficients are estimated separately for each treatment. Marginal effects are in Table 5.

We do not see a significant effect in the probability of choosing "stay." This similarity extends to the other two possible economy configuration choices, in treatments where potential gains were identically distributed. Panels A and D reveal that player types made similar choices in treatments where the potential gains from integration were identically distributed. None of the coefficients is significant, and we cannot reject the null of equal coefficients on Middle and

[18]We can reject the null that cooperative participants have a higher propensity to choose the socially efficient 24-player group. We categorize subjects into three groups (low, moderate, high) based on cooperation relative to their opponent(s), for fixed pairs and mixed groups of Phase 1. This categorical variable is used in a regression model as in Table 4, but omitting realized gains. Relative cooperativeness in fixed pairs does not affect group choices. Instead, Low cooperators in mixed groups are significantly more likely to choose the socially efficient "stay" choice, while high cooperators are significantly less likely to do so. This evidence is consistent with the results in Bigoni et al. (2019), suggesting that high cooperators avoid mixed groups to coordinate on cooperation with a fixed partner, while low cooperators seek large groups where free-riding cannot be directly sanctioned.

Advantaged, for any of the three possible choices (Wald tests). Some differences in the other two choices, "leave" and "exclude" emerge in treatments where potential gains were unequally distributed. Panels B and C show that Advantaged players selected "leave" more frequently than the Disadvantaged, and "exclude" less frequently. This "isolation bias" reflects the distribution of potential gains to some extent. Advantaged players selected "leave" more frequently than the Disadvantaged, respectively by about 18 and 9 percentage point in Converge and in Diverge. There is a corresponding difference of the opposite sign in the frequency of "exclude," which is interesting because it reveals that all types shared a similar inclination regarding "stay."[19]

Table 5: Inequality of Opportunity and Economy Configuration Choices

Dep. var. =	**Panel A: Neutral**			**Panel B: Converge**		
1 if vote	Leave	Exclude	Stay	Leave	Exclude	Stay
Middle	-0.079	0.121	-0.042	0.108	-0.165	0.056
	(0.111)	(0.098)	(0.039)	(0.155)	(0.104)	(0.105)
Advantaged	-0.031	0.068	-0.037	0.181*	-0.187*	0.006
	(0.106)	(0.079)	(0.029)	(0.105)	(0.101)	(0.086)
	Panel C: Diverge			**Panel D: Neutral+**		
	Leave	Exclude	Stay	Leave	Exclude	Stay
Middle	-0.069	0.109	-0.040	0.057	0.021	-0.079
	(0.108)	(0.110)	(0.083)	(0.126)	(0.107)	(0.049)
Advantaged	0.091**	-0.093**	0.002	-0.125	0.026	0.100
	(0.036)	(0.046)	(0.033)	(0.137)	(0.159)	(0.095)

Notes: Multinomial logit regressions on preferences for choices of endogenous configuration. Marginal Effects. One observation = one subject in a session. Robust standard errors (in parentheses) adjusted for clustering at session level. The categorical variable corresponds to the three types of players (Disadvantaged is the base case). Each panel refers to a separate regression, using data for the specified treatment. For other details see notes to Table 4.

[19]Middle players behave similarly to Disadvantaged. They fall in-between Advantaged and Disadvantaged. We can only reject the null of equal coefficients on Middle and Advantaged for the "exclude" choice in Panel C (Wald test, p-value=0.0139). Table B12 in Appendix B reports the distribution of choices by type.

Summing up, the preference for economic integration is unaffected by disparities in *potential* gains from integration. All player types selected full economic integration ("stay") with a similar frequency, independent of size and distribution of potential gains. We can also rule out that players of different types made different group choices in the Neutral and Neutral+ treatments, where potential gains from integration are equally distributed. Instead, when potential gains from economic integration are unequally distributed (the Converge and Diverge treatments) we observe that Advantaged players choose differently than the Disadvantaged. They exhibit a lower frequency of "exclude" choice and a higher frequency of "leave" choice. We interpret this evidence as suggesting that, under unequal distributions, "leave" became a more attractive alternative to players who had an economic advantage to start with.

5.3 The role of strategic uncertainty

In the experiment full cooperation requires tacit coordination on a risky and complex dynamic strategy. It is thus possible that our results on group choices depend on the strategic uncertainty present in mixed groups preventing cooperation and, hence, gains from integration.[20] Here we reduce strategic uncertainty by introducing pre-play communication in free-form. This was shown to improve contributions in public-goods experiments among partners (e.g., Chan et al., 1999; Dekel et al., 2017; Gangadharan et al., 2017; Isaac and Walker, 1988b; Tavoni et al., 2011) and to facilitate efficient play in indefinitely repeated games among strangers (e.g., Camera et al., 2013b, 2020b). The Neutral-Chat treatment pursues this idea, introducing costless pre-play free-form communication in Neutral. Before each Phase 1 supergame, participants can communicate with each other using a chat-box, for up to two

[20]We thank an anonymous referee for suggesting this line of inquiry.

minutes. No communication is possible in Phase 2, to avoid explicit coordination on group selection. We ran four sessions with the order 2-2-12-12, which we compare to the four similar Neutral sessions.

As Neutral-Chat allows *explicit* coordination on a strategy, we expect higher cooperation in mixed groups, hence positive realized gains from integration in Phase 1. This experience should support the belief of high cooperation in future mixed groups, encouraging the selection of fully integrated economies in Phase 2. Given this belief, "leave" is economically suboptimal for an individual, while "exclude" simply hurt others without providing a direct personal economic benefit (although reducing the economy size might mitigate strategic uncertainty—an indirect benefit). This suggests two testable hypotheses:

H 4. *In Phase 1, mixed groups cooperate more with than without pre-play communication.*

H 5. *In Phase 2, "stay" is more frequently selected with than without pre-play communication.*

Neither hypothesis can be rejected. Average cooperation rates in mixed groups of Phase 1 are 0.939 in Neutral-Chat vs. 0.516 in Neutral, a highly significant difference (GLM regression, col. 1, Table B13 in Appendix B). The ability to explicitly coordinate strategies promoted efficient play, which led to nearly universal gains from integration: 95 out of 96 subjects had gains in Neutral-Chat, 1.61 points on average, while only 55 out of 96 gained in Neutral, 0.42 points on average. The four-fold increase in average gains from integration motivated a more frequent choice of mixed groups. In Neutral-Chat the frequency of choice of "leave" was .08, "exclude" 0.41, and "stay" 0.51, while the opposite pattern is observed in Neutral (0.47, 0.33, and 0.20). The differences in frequencies are statistically significant for "leave" and "stay" (two-sided Wilcoxon-Mann Whitney ranksum test with exact statistics, p-values=0.002,0.002, N1=N2=4). This change in attitudes towards economic integration led to the formation of 24-player groups in 2 of the 4 sessions of Neutral-Chat as opposed to none in Neutral.

Did subjects choose rationally or were their expectations about behavior in mixed groups mistaken? When we pool together all mixed groups of Phase 2 (N=4 per treatment), average cooperation is 0.749 in Neutral-Chat vs. 0.399 in Neutral, a statistically significant increase (col. 2 in Table B13). As a result, 61 out of 80 subjects who participated in a mixed group of Phase 2 earned more than they did in fixed pairs of Phase 1 (average gain = 0.94 points). This contrasts with only 29 out of 64 subjects in the four comparable Neutral sessions for an average gain of -0.03 points.

To understand if this change in preference for economic integration is entirely explained by the greater gains experienced in Phase 1, we run a multinomial logit regression that controls for realized gains; see Table B14 in Appendix B. There is a large and significant treatment effect, suggesting that free-form communication contributed to shift subjects' attitudes toward forming large heterogeneous groups. That is to say, the benefit of the chat box went beyond simply increasing the expectation of high payoffs in mixed groups. Still, there is no clear majority for "stay." In fact, many subjects chose to exclude some player type from the mixed group. Out of 96 subjects, 24 chose to exclude the disadvantaged, while 7 and 8, respectively, chose to exclude the other two types of players. Since excluding a player type reduces the heterogeneity in the group, without altering the payoff matrix, this suggests that resolving strategic uncertainty issues did not remove the aversion to interacting in large heterogeneous groups as some subjects tried to minimize it.

6 Discussion

What may induce a country to avoid an open-trade policy? From the perspective of international economics, it is primarily the possibility that the economic gains envisioned for the average citizen may come at the expense of absolute losses for *some* of them (Stolper and Samuelson, 1941). Our experiment reveals that this may happen even if *everyone* has a potential absolute gain because attaining the gains from trade is not a mechanical process, but requires cooperation on a large scale. In the experiment, every participant could theoretically earn more in mixed groups than in fixed pairs, economic integration was potentially beneficial, but these gains did not materialize because participants did not coordinate on a cooperative strategy. This difficulty together with heterogeneous potential gains created winners and losers from economic integration, which ultimately undermined support for it.

This suggests the importance of accounting for behavioral aspects that may influence individuals' perception of economic integration. One may object that subjects in the experiment did not understand the potential gains from integration. Although possible, this is an unlikely explanation for our results. The instructions informed subjects of the cooperation benefits in mixed groups vs. fixed pairs and our regressions control for understanding of instructions using a post-instruction quiz, which showed how to calculate earnings under full cooperation and full defection in any group.

The experiment provides several insights about factors that might preclude potential gains from integration to be attained outside of the lab. First, the scale of interaction is a primary obstacle to cooperation, due to coordination problems, while inequality of economic *opportunity* does not appear to directly affect cooperation. In the experiment, cooperation did not improve in mixed groups when its benefits where equalized or increased across the board (Results 1-2) and was always low (Result 3). Second, inequality of opportunity affects the *choice* to aggregate into larger groups because low cooperation splits the group into absolute winners and losers (Result 4-7). To the extent that expanding trade beyond local boundaries requires greater reliance on self-enforcing agreements, integration is more likely to deliver its promised benefits—and thus to be supported—when it can leverage institutions that facilitate coordination and mitigate strategic uncertainty, and foster an overall trust between countries. Seen this way, the experiment suggests that the backlash against integration can hardly be ascribed to a communication failure of policymakers and economists, say, because the potential benefits are unequal across countries or not made transparent to the public. It suggests that the distribution of gains *realized* in those countries that did pursue an integration policy affected subsequent attitudes toward that policy.[21]

What explains the efficiency failure from economic integration in the experiment? As we move from fixed pairs to mixed groups: (i) we go from partner to strangers matching, (ii) the average cooperation benefit increases from 13 to 16 points, and (iii) we shift from a homogeneous to a heterogeneous economy where cooperation benefits differ. Each of these changes might influence cooperation in its own way. It could fall as we go from partners to strangers or as we shift from homogeneous to heterogeneous economies, while it could rise as its average benefit increases. To disentangle these three potential effects consider the following. In previous experiments with homogeneous

[21]Fetzer (2019) notes that the austerity-induced welfare reform in the UK may have created losses that partly made economic integration seem responsible for their fate. Fetzer and Wang (2020) offer a measure of the economic cost of Brexit, in the very short-run. They find sizable and unevenly distributed costs across the 382 U.K districts, and classify about 168 districts as losers, 78 as winners, and the remaining as neither. Interestingly, they also find more pronounced losses in districts that more strongly supported Leave.

players cooperation significantly declined as interaction shifted from partners to strangers matching, when cooperation benefits were constant (Camera et al., 2013a), and also when they increased (Bigoni et al., 2019). Moreover, in our experiment cooperation did not improve in Neutral+, when the average cooperation benefit increased while keeping constant inequality in benefits and strangers matching. Camera and Hohl (2021) compares homogeneous to heterogeneous mixed groups similar to ours by varying the distribution of cooperation benefits, while keeping constant mean benefits and strangers matching. Using their data, we carried out an analysis (see section B.9 in Appendix B) revealing that the mean-invariant redistribution of benefits did not significantly affect cooperation. The survey of related experiments in Bigoni et al. (2020, Section 3) shows that when other aspects of the design change (e.g., continuation probability, role alternation, information), we still see lower cooperation in strangers settings. Overall, this suggests that integration failed to improve efficiency primarily due to the shift from partners to strangers matching, not due to the induced heterogeneity in prospective payoffs.

Supposing that these laboratory results reflect a principle of behavior that also underlies external decision processes, what policy considerations can we make given the current disintegration tendencies? First, it has been argued that standard models of trade in goods and services may underestimate the gains from trade by not taking into account the dynamics of innovation in integrated markets (Desmet et al., forth.; Ossa, 2015). This may suggest that, as a matter of policy, advertising more optimistic prospects to the public might itself improve attitudes toward integration. The experiment does not support this view if we narrowly consider the results from Neutral+ where cooperation did not improve relative to the baseline even if prospective integration benefits were much better. As a result, the 24-player group was infrequently selected.

The main barrier to cooperation and integration seems the significant strategic uncertainty under which participants had to make decisions (see Bigoni et al., 2019, and our Neutral-Chat treatment): cooperation required tacit coordination on a complicated dynamic strategy involving community sanctions. Institutions designed to foster coordination, open communication and a spirit of cohesion in the international arena thus seem fundamental. It is also possible that in the experiment the potential integration benefits were not large enough to dominate the underlying strategic uncertainty; it is an open question if more robust economic prospects would be alone sufficient to overcome barriers to cooperation and integration.

Second, the experiment reveals that although inequality in prospective economic benefits might not directly influence integration attitudes, it indirectly does so in uncooperative environments by creating sharp economic imbalances—absolute losers alongside winners. This suggests that there is scope for cross-country redistributive fiscal policies to limit strong negative effects on some groups caused by integration. This last aspect is in line with findings on Brexit (Fetzer, 2019), and is especially relevant if we think of integration benefits in the experiment as embedding also a "political" component. If so, the loss in political sovereignty that is generally associated with economic integration should also be accounted for to calculate the benefits *perceived* by some countries but—due to different cultural views— not others. This may also contribute to explain the disintegration trends we have been witnessing.

There are limitations of our study that must be considered in assessing the external validity of our experiment. Three, in particular, stand out: the binary-choice design, the lack of prices, and the exogenous distribution of the potential gains from integration. In our design each interaction involves a binary choice over two *extreme* cooperation levels: either full or none. However, human and commercial interactions are more complex, and allow a much wider range of choice. For example, partners might find it easy to fully cooperate from the start of an interaction, while strangers might not, starting from a low, but non-zero, level which is then raised to is full potential once trust gets established. If so, then our study likely overestimates the negative impact of economic integration on cooperation. We also excluded prices, which are key determinants of international trade flows in field economies and, hence, the benefits of economic integration. Consumers in our design could not entice strangers to cooperate by offering some of the surplus created by their cooperative action. In a way, prices are fixed, which might have contributed to reduce cooperation rates among strangers and, consequently, the desirability of economic integration. Finally, the determination of winners and losers from integration is not fully endogenous in our design, being affected by the assumed heterogeneity in payoffs and potential gains from integration. This might create gains (or losses) that are artificially low—relative to a fully endogenous determination—or, conversely, artificially high. As a result, this would end up either understating or overstating the true backlash against integration, we would expect in the field. These features of our design are limitations from an external validity perspective, and natural candidates for further investigation.

References

Ahn, T., Isaac, M., and Salmon, T. (2009). Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics*, 93(1), 336-351.

Antràs, P. and Costinot, A. (2010). Intermediation and Economic Integration. *American Economic Review: Papers & Proceedings* 100, 424-428.

Antràs, P. and E. Rossi-Hansberg. (2009). Organizations and Trade. *Annual Review of Economics* 1(1), 43-64.

Balafoutas, L., M. Kocher, L. Putterman, and M. Sutter. 2013. Equality, equity and incentives: an experiment. *European Economic Review* 60, 35-51.

Baland, J.M., Gangadharan, L., Maitra, P., and Somanathan, R. (2017). Repayment and exclusion in a microfinance experiment, *Journal of Economic Behavior & Organization*, 137, 176-190.

Barrett, S. and Dannenberg, A. (2012) Climate negotiations under scientific uncertainty. *Proceedings of the National Academy of Sciences* 17372-17376.

Bigoni, M., Camera, G., and Casari M. (2019) Partners or Strangers? Cooperation, monetary trade, and the choice of scale of interaction. *American Economic Journal: Microeconomics*, 11(2), 195-227.

Bigoni, M., Camera, G., and Casari M. (2020). Cooperation among strangers with and without a monetary system. In C. Monica Capra, Rachel Croson, Mary Rigdon and Tanya Rosenblat (eds), *Handbook of Experimental Game Theory*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, 213-240.

Bolton, G.E, Ockenfels, A., 2000. ERC: a theory of equity, reciprocity and competition. *American Economic Review* 90, 166-193.

Buckley, E., and Croson, R. 2006. Income and wealth heterogeneity in the voluntary provision of linear public goods. *Journal of Public Economics*, 90 (4-5), 935-955.

Camera, G., and Hohl, L. (2021). Group-identity and long-run cooperation: an experiment. *Journal of Economic Behavior & Organization*, 188, 903-915.

Camera, G., Casari, M., and Bigoni, M. (2013a). Money and trust among strangers. *Proc. of the Nat. Ac. of Sciences of the U.S.*, 110(37), 14889-14893.

Camera, G., Casari M., and Bigoni, M. (2013b) Experimental Markets with Frictions. *Journal of Economic Surveys* 27(3), 536-553.

Camera, G., Deck, C., and Porter, D. (2020a). Do Economic Inequalities Affect Long-Run Cooperation and Prosperity? *Experimental Economics.*

Camera, G., D. Goldberg, and A. Weiss (2020b). Endogenous market formation and monetary trade: an experiment. *Journal of the European Economic Association* 18(3), 1553-1588

Cappelen, A. W., Eichele, T., Hugdahl, K., Specht, K., Sørensen, E., and Tungodden, B. (2014). Equity theory and fair inequality: A neuroeconomic study. *Proc. of the Nat. Acad. of Science of the U.S.* 111(43), 15368-15372.

Chakraborty, I., Hai, R., Holter, H.A., and Stepanchuk, S. (2017). The Real Effects of Financial (Dis)integration: A Multi-Country Equilibrium Analysis of Europe. *Journal of Monetary Economics*, 85, 28-45.

Chan, K.S., Mestelman, S., Moir, R., and Muller, R.A. 1999. Heterogeneity and the voluntary provision of public goods. Experimental Economics, 2(1), 5-30.

Cinyabuguma, M., Page, T., and Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89(8), 1421-1435.

Costinot A., and Rodríguez-Clare A. (2018). The US Gains From Trade: Valuation Using the Demand for Foreign Factor Services *Journal of Economic Perspectives*, 32(2), 3-24.

Croson, R., Fatas, E., Neugebauer, T., and Morales, J. (2015). Excludability: A laboratory study on forced ranking in team production. *Journal of Economic Behavior & Organization*, 114, 13-26.

Dal Bó, P., A. Foster, and L. Puttermann (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review*, 100. 2205 - 2229.

Dal B´, E., Dal Bó, P., and Eyster, E. (2018). The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects. *The Review of Economic Studies*, 85, 964-998.

Dannenberg, A., C. Haita-Falah, and S. Zitzelsberger. 2019. Voting on the threat of exclusion in a public goods experiment. *Experimental Economics* 23, 84-109.

Dannenberg, A. and C. Gallier (2020). The choice of institutions to solve cooperation problems: a survey of experimental research. *Experimental Economics*, 23. 716- 749.

De Geest, L. R., and D. C. Kingsley (2021). Norm enforcement with incomplete information. *Journal of Economic Behavior & Organization*, 189, 403-430.

Dekel, S., Fischer, S., and Zultan, R. (2017). Punishment and reward institutions with harmed minorities. *Journal of Public Economics*, 146(c), 87-96.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy* 126(3), 903-983.

Feenstra, R. (2018). Alternative Sources of the Gains from International Trade: Variety, Creative Destruction, and Markups. *Journal of Economic Perspectives*, 32 (2), 25-46.

Fehr, E. and K. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817-868.

Fetzer, T. (2019). Did Austerity Cause Brexit? American Economic Review 109(11), 3849-3886.

Fetzer T., and S. Wang. 2020. Measuring the Regional Economic Cost of Brexit: Evidence Up to 2019. CAGE working paper no. 486.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171-178.

Fischbacher, U., S. Schudy, and S. Teyssier. 2014. Heterogeneous reactions to heterogeneity in returns from public goods. *Social Choice and Welfare* 43, 195-217.

Fort, T. C., J. R. Pierce, and P. K. Schott (2018). New Perspectives on the Decline of US Manufacturing Employment. *Journal of Economic Perspectives*, 32(2), 47-72.

Gallier, C., M. Kesternich, and B. Sturm. 2017. Voting for burden sharing rules in public goods games. *Environmental and Resource Economics* 67, 535-557.

Gangadharan, L., Nikiforakis, N., and Villeval, M.C. (2017). Normative Conflict and the Limit to Self-Governance in Heterogeneous Populations. *European Economic Review*, 100, 143-156.

Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2006). The competitive advantage of sanctioning institutions. *Science* 312, 107-111.

Güth, W., Levati, V., Sutter, M., and Van der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5), 1023-1042.

Hakobyan, S., and J. McLaren (2016). Looking for Local Labor Market Effects of NAFTA. *Review of Economics and Statistics* 98(4), 728-41.

Hargreaves Heap, S. P., A. Ramalingam, and B.V. Stoddard (2016). Endowment inequality in public goods games: A re-examination. *Economics Letters*, 146, 4-7.

Haushofer, J. and Fehr, E. (2014). On the psychology of poverty. *Science*, 344(6186), 862-867.

Isaac, R. M., and Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, 179-199.

Isaac, R. M., and Walker, J. M. (1988). Communication and free-riding behavior: the voluntary contribution mechanism. *Economic Inquiry*, 26, 585-608.

Kandori, M. (1992). Social Norms and Community Enforcement, *Review of Economic Studies*, 59(1), 63-80.

Kölle, F. (2015). Heterogeneity and cooperation: The role of capability and valuation on public goods provision. *Journal of Economic Behavior & Organization*, 109, 120-134.

Maier-Rigaud, F. P., Martinsson, P., and Staffiero, G. (2010). Ostracism and the provision of a public good: experimental evidence. *Journal of Economic Behavior & Organization*, 73(3), 387-395.

Nash, J. F., R. Nagel, A. Ockenfels, & R. Selten, 2012. The agencies method for coalition formation in experimental games. *Proceedings of the National Academy of Sciences*, 109(50), 20358-20363.

Nikiforakis, N., Noussair, C.N., and Wilkening, T. 2012. Normative conflict and feuds: the limits of self-enforcement. *Journal of Public Economics*, 96(9-10), 797-807.

Noussair, C., Plott, C. R., and Riezman, R. G. (1995). An Experimental Investigation of the Patterns of International Trade. *American Economic Review* 85(3), 462-491.

Ossa. R. 2015. Why trade matters after all? *Journal of International Economics* 97(2), 266-277

Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83, 1281-1302.

Reuben, E., and Riedl, A. 2013. Enforcement of Contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77, 122-137.

Rodrik, D. (2000). How far will international economic integration go? *Journal of Economic Perspectives*, 14 (1), 177-186.

Rodrik, D. (2018). What Do Trade Agreements Really Do? *Journal of Economic Perspectives*, 32 (2), 73-90.

Rossi-Hansberg E. (2017). *200 Years of Ricardian Theory: The Missing Dynamics.* In: Jones, R. W., and Weder, R. (Eds.), 200 Years of Ricardian Trade Theory, 197-205, Cham: Springer.

Reuben, E. and A. Riedl. 2009. Public goods provision and sanctioning in privileged groups. *Journal of Conflict Resolution* 53, 72-93.

Reuben, E., and Riedl, A. 2013. Enforcement of Contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77, 122-137.

Sadrieh, A., and H.A. Verbon. 2006. Inequality, cooperation, and growth: an experimental study. *European Economic Review* 50, 1197-1222.

Sampson, T. (2017) Brexit: The Economics of International Disintegration. *Journal of Economic Perspectives* 31(4), 163-184.

Stiglitz, J. E. (2020). *Rewriting the Rules of the European Economy*, New York: W.W. Norton & Company (in Collaboration with Carter Dougherty and the Foundation For European Progressive Studies).

Stolper, W. F., and Samuelson, P. A. (1941). Protection and real wages. *The Review of Economic Studies*, 9(1), 58-73.

Tavoni, A., Dannenberg, A., Kallis, G., and Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. of the Nat. Acad. of Sciences of the U.S.*, 108(29), 11825-11829.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80(1), 234-248.

Weber, R. A. (2006). Managing Growth to Achieve Efficient Coordination in Large Groups. *American Economic Review*, 96(1), 114-126.

Zelmer, J. 2003. Linear public goods experiments: a meta-analysis. *Experimental Economics* 6, 299-310.

A Appendix

A.1 Proof of proposition 1

Here we prove that full cooperation is a sequential equilibrium in every economy and treatment. We say that a norm of cooperation is being followed in the group whenever all players adopt the trigger strategy discussed in Section 4. For clarity, let the defection payoffs be, respectively, $d = 6$ and $d - l = 3$ to a producer and a consumer. Let $k_i := 9 + 2i + y$ denote the cooperation payoff to a consumer of type $i = 1, 2, 3$, with $y = 0, a_i$ in a fixed pair and mixed group, respectively. Here, the gain from integration a explicitly depends on the type of player, as it happens in some treatments. A necessary and sufficient condition for full cooperation to be an equilibrium is reported in the following lemma:

Lemma 1. *Fix an economy. Let k denote the smallest cooperation payoff in that economy. If the continuation probability*

$$\beta \geq \beta^* := \frac{d}{k - d + l} \in (0, 1),$$

then full cooperation is a sequential equilibrium.

Study the payoff to a type i player. Under full cooperation, she earns k_i every other round as a consumer (zero, as a producer). Let $s = 0, 1$ denote the role of the player at the start of a round, where 0=producer and 1=consumer. The type of counterpart does not affect the player's payoff—only their action as a producer. The equilibrium payoff is

$$v_0 := \frac{\beta k_i}{1 - \beta^2} \qquad \text{and} \qquad v_1 := \frac{k_i}{1 - \beta^2}.$$

To understand v_0 note that in equilibrium every player always cooperates as a producer. Hence, a player of type i who is a producer earns 0 in the current period, and k_i next period (as a consumer), a value discounted by β. As this two-period cycle is indefinitely repeated, we obtain v_0. The explanation is similar for v_1.

Off-equilibrium there is full defection so the payoff corresponds to the one associated to infinite repetition of the static Nash equilibrium, denoted

$$\hat{v}_0 := \frac{d + \beta(d - l)}{1 - \beta^2} \qquad \text{and} \qquad \hat{v}_1 := \frac{d - l + \beta d}{1 - \beta^2}.$$

Full defection payoffs do not depend on the type i, unlike equilibrium payoffs. It is immediate that off-equilibrium a producer has no incentive to deviate from following the sanctioning rule (always defect), because defecting is the unique best response to every other producer defecting in every round. Hence, we only need to show that $v_0 \geq \hat{v}_0$, i.e., in equilibrium the player has no incentive to defect as a producer, by refusing to help some consumer.[22] This inequality can be rearranged as $\beta \geq \beta_i^* = 6/(6+2i+y)$ for the case of fixed pairs and mixed groups, and the Lemma automatically follows. Note that $\beta_i^* < 1$ because $k_i - (2d - l) > 0$ by assumption for all player types in all economies. The Lemma exploits the fact that the lowerbound probability β consistent with cooperation is a decreasing function of the player's return from cooperation k_i. Hence β_i^* decreases in i; players of "higher" type have higher returns from cooperation, hence a greater economic incentive to cooperate; see Table A1. Proposition 1 follows from observing that in the experiment $\beta = 0.75$ and the most stringent requirement for existence of equilibrium comes from fixed pairs composed of type 1 players, in which case $y = 0$ and $k_i = 11$; here, $\beta_1^* = 0.75$, which is the smallest lowerbound threshold.

Table A1: Threshold continuation probability β^*.

	Isolated econ.			Integrated econ.		
Treatment	i =1	2	3	i =1	2	3
Neutral	.75	.60	.50	.55	.46	.40
Converge	.75	.60	.50	.46	.46	.46
Diverge	.75	.60	.50	.67	.46	.35
Neutral+	.75	.60	.50	.46	.40	.35

[22]Though in the experiment discounting starts on round $T = 18$, the round in which the random termination rule started, one can demonstrates that the incentives to cooperate monotonically decline until round t. It follows that by studying the incentives to cooperate in equilibrium using payoffs associated with the beginning of round T ensures those incentives are satisfied in all $t < T$. In round $t = T$ payoffs correspond to v_s above. The details of this demonstration are provided in Bigoni et al. (2019).