Breaking Up: Experimental insights into international economic (dis)integration*

Gabriele Camera
Chapman University
University of Bologna

Lukas Hohl
University of Basel

Rolf Weder
University of Basel

February 1, 2019

Abstract

Trade theory suggests that people should embrace economic integration because it promises large gains. But recent events such as "Brexit" suggest a desire for economic *dis*integration. Here we report results of a laboratory investigation of how size and distribution of potential gains from integration influences the actual economic outcomes, and individuals' inclination to belong to an "integrated" economy. The experiment documents the existence of several behavioral challenges to the process of economic integration, which prevent individuals from fully realizing the gains promised by trade theory.

Keywords: experiments, indefinitely repeated games, international integration, social dilemmas.

JEL codes: C70, C90, D03, E02

1 Introduction

At the heart of international trade theory lies the idea that heterogeneous individuals who are strangers to each other can benefit a great deal from trading with each other. This basic idea goes back to David Ricardo (1817), who

* We acknowledge partial research support through the SNSF grant No. 100018.172901.
Correspondence address: Gabriele Camera, Economic Science Institute, Chapman University, One University Dr., Orange, CA 92866; e-mail: camera@chapman.edu.

noted that market integration pushes forward the efficiency frontier compared to a series of smaller, separate entities. Yet, after decades of increasing regional and global economic integration, we are witnessing signs of a pursuit of economic *dis*integration. The U.S. presidential election in 2016 with the subsequent backlash in the support of regional trade agreements as well as of the multilateral trading system guarded by the World Trade Organization is one indication. The choice of a majority of the British people in 2016 to exit the European Union is another glaring example. But also the 2017 independence vote in the Spanish region of Catalonia and the rising Euroscepticism expressed in German and Italian polls in 2018 point in the same direction. The question is why: what pushes individuals to seek economic *dis*integration?

The trade literature suggests that part of the answer might lie in the distribution of the gains from trade. It is theoretically well-established that integration tends to disproportionately benefit certain industries or factors of production while harming others (Stolper and Samuelson, 1941). It has also been argued that traditional analyses might significantly underestimate the gains from economic integration (Desmet et al., forth.). These considerations suggest that distributional effects and low anticipated benefits—economic and even of a political nature—may be contributing factors to the recent disintegration trend, as they could possibly cause a decline in the support for economic integration among individuals exposed to structural adjustment.

Here, we investigate this possibility by creating laboratory economies where individuals, who know the distribution of benefits made possible by integration, must independently decide whether to pursue it or not. As a result, the organization of trade and the realized distribution of benefits are endogenous. In this manner, this study does not offer a test of a particular trade model but, rather, tests a tacit operating principle of trade theory: individuals who are strangers to each other are naturally willing and capable to reap the benefits of trading with each other. Adopting an experimental approach is especially helpful in this case, as it allows us to manipulate size and distribution of the potential gains from integration, while minimizing possible confounding factors—political, social and cultural, for instance.

The study develops, first, by giving context through a discussion of related experiments on cooperation, group formation, and international trade (Section 2). We then proceed by illustrating the trade-theoretic notion that economic integration can potentially benefit everyone (Section 3), and explain how we develop an experimental design that captures this notion (Section 4). In our experiment, subjects belonging to three identical-size groups can choose to confront a repeated cooperative task either under integration or under autarky. For conceptual clarity, integration involves many random counterparts from all groups, while autarky involves a stable partnership within the subject's group. Everyone can theoretically benefit from integration—more can be earned. But this requires mutual cooperation. In other words, integration yields benefits that are proportional to cooperation rates. As a result, the choice to integrate or not depends not only on the potential gains but also on the gains that it *effectively* yields. These gains are endogenous and may fall short of potential. This choice is non-trivial because subjects cannot monitor individual past conducts, so cannot prevent adverse selection into large groups by screening or sanctioning specific individuals.

The treatments vary the size and distribution across groups of the possible gains from integration. This serves as a link to the observations described above that the gains from integration may well differ between individuals and countries. We employ the theory of repeated games to study how variation in these economic fundamentals affects outcomes in our artificial economies

(Section 5). This allows us to formalize the idea that capturing the gains from integration is not an automatic process, and to derive hypotheses which we then test using the experimental data (Section 6).

The empirical analysis reveals that if subjects experience first fixed pairs and then mixed groups, economic integration raises realized efficiency and thus average payoffs. The heterogeneity of potential gains from integration does not affect the efficiency level attained in the experiment as it does not significantly alter the average rate of cooperation. When we ask subjects to express a preference to stay in the large group (full integration) vs. selective integration (excluding one country from the group) or disintegration (leaving the large group) then we observe that disintegration is the predominant preference. After controlling for the subjects' *experienced* gains, the size of the potential gains from integration affects this choice. We use the variation in potential gains across countries and find that when the potential gains are small, the preference to stay out of the large group increases. When, instead, we increase the gains from integration for every country, we find that the desire for disintegration is no longer dominant. We conclude by putting these results into perspective and offer some policy considerations (Section 7).

2 Related experimental literature

There are only a few experiments about international trade. The pioneering article in Noussair et al. (1995) constructs internationally-integrated laboratory markets to test the conformity of prices and trade flows to the prediction of the competitive trade model. It finds support for factor-price equalization and the principle of comparative advantage in the lab; but it also emphasizes that trade is significantly below the theoretically predicted volume.

By contrast, our experiment does not aim to test theorems from a specific trade model. The goal is to investigate the idea at the heart of trade theory: heterogeneous individuals who are strangers to each other are able and willing to capture the gains from trading with each other. In the experiment, an internationally integrated economy emerges as an endogenous outcome when players who may not trust each other and have heterogeneous benefits from trading with each other choose to form a large, cooperative trading group. In a way, our investigation digs deeper in the possible causes for the earlier experimental result that trade volumes are smaller than predicted.

There is abundant experimental evidence suggesting that—with opportunism and coordination issues—larger groups attain outcomes that are less efficient as compared to smaller groups, when the group size varies exogenously. For example, free riding is more pronounced in large than small groups when the cooperation task takes the form of a finitely repeated Voluntary Contributions Mechanism (Isaac and Walker, 1988). This same pattern emerges when the cooperation task is indefinitely repeated, in prisoners's dilemmas (Camera and Casari, 2009) or helping games (Camera et al., 2013). Designs that focus entirely on coordination problems also reveal that efficiency declines as group size exogenously increases (Van Huyck et al., 1990), although this problem can be corrected through initial exposure to small groups, and judicious increase in group size (Weber, 2006). Our study replicates these baseline results—by initially confronting subjects with exogenous variation in group size—and then extends the analysis to the case where group size varies endogenously as a result of independent choices of many individuals whose potential benefits from joining a large group are heterogeneous.

The experimental literature on endogenous group size has focused on investigating the impact of group entry or exit rules on the size choice, as well as the efficiency and cooperation level of the associated outcome. A main finding is that cooperators will readily aggregate into large groups of like-minded individuals, as long as institutions exist that allow the identification of free riders and their isolation. Various kinds of institutions have proved to be effective in supporting the formation of large groups. Experiments on VCM have shown that contributions to a public good increase if individuals can choose to expel low contributors from the group (Cinyabuguma et al., 2005; Güth et al., 2007; Maier-Rigaud et. al., 2010), if low contributors are automatically removed from the group (Croson et al., 2015), and if group entry can be restricted (Ahn et al., 2009). These institutions support the formation of large, efficient groups because they facilitate identification and direct sanctioning of free riders, thus allowing positive assortative matching (or self-selection).

Our study differs from these experimental designs in three dimensions: type of game, absence of self-selection, and composition of the equilibrium set. First, the underlying cooperative task in our study is not the usual VCM game among partners, but embodies a cognitively simpler decision problem (a "helping game") involving pairs of strangers drawn at random from a fixed matching pool. Second, given our focus on international integration, we do not allow self-selection into separate groups of cooperators and defectors: subjects cannot be selective in their inclusion or exclusion choices—excluding free-riders and including cooperators, for example—nor can a subject choose to individually leave or join a group. Subjects independently express a preference for a group configuration, after which a collective decision process takes over, which aggregates the subjects' preferences and determines the outcome. If a large group is formed it includes cooperators and free-riders alike.

Third, in earlier experiments the interaction is of known and short duration, so standard theory predicts inefficient play. By contrast, we employ a design with interaction of unknown and long duration, so the efficient outcome is an equilibrium in groups of any size. This is also a characteristic of the study in Bigoni et al. (forthcoming), on which our design is based. The central innovation consists in relaxing the constraint of that earlier design that the gains from cooperation are equal across all players. In our study, earning prospects are heterogeneous—so some players may benefit more and others less from cooperation in mixed groups.

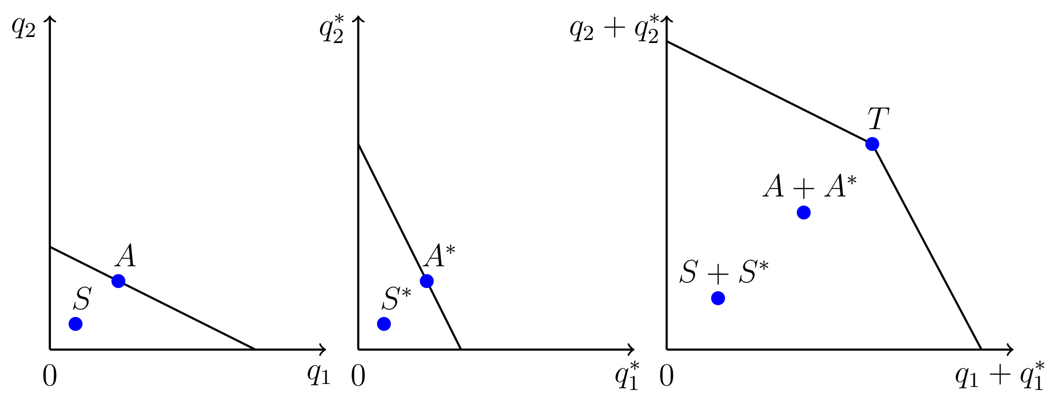
This last dimension allows us to investigate an important open question in the literature on group formation and cooperation, i.e., whether the formation of large, cooperative groups depends on both the size and the distribution of the additional gains that are possible in larger groups. This is an important aspect when thinking about the support and outcomes of international trade because the gains from international trade typically differ across individuals and countries.

3 A trade-theoretic foundation of this study

Consider two countries, Home and Foreign. Individuals produce and consume goods 1 and 2, and can choose to specialize and trade only locally (*autarky*) or with each other (*integration*). Figure 1 illustrates how a trade theorist typically views this problem. The first two panels depict the production possibility frontier (PPF) for Home and Foreign. It traces the maximum production (and consumption) levels, q_1 and q_2, attainable when individuals produce what they are relatively better at, and then trade it locally. Allocations inside the PPF are inefficient; the worst-case scenario is no domestic trade at all, or *self-sufficiency* (S, S^*).

Suppose that in equilibrium there is a unique allocation that maximizes payoffs under autarky, A for Home (A^* for Foreign). Other allocations may exist between S and A that are also an equilibrium but where not all gains from local exchange are exploited. However, allocations A and A^* are only constrained Pareto-efficient: Everyone can improve their welfare by agreeing to trade with everyone else. This would allow to capture the Ricardian gains from international trade as Home is relatively better at producing good 1 than Foreign (Home's PPF is flatter and thus implies lower opportunity costs of producing good 1).

Figure 1: Allocations under self-sufficiency, autarky and integration



The last panel illustrates the integrated-economy PPF, combining the two PPFs, and tracing all possible degrees of specialization—given that each country specializes in the direction of its comparative advantage. If the absolute production capacities of the two countries do not differ much, the trade equilibrium likely occurs at T where each country completely specializes in the production of one good (Home in good 1 and Foreign in good 2).

Pareto-inferior equilibria may exist in between $A + A^*$ and T—where the gains from integration are only partially exploited. Trade theory, however, typically assumes perfectly working decentralized competitive markets and

Pareto efficiency as guiding equilibrium selection. It stipulates that individuals in each country will neither refuse to integrate, nor will end up inside the combined PPF, independently of the size and the distribution of the gains from integration. In short, trade-theory makes two predictions. First, "full cooperation" is as likely under *integration* as it is in *autarky*. Second, if given a choice, individuals in both countries prefer integration.

This means that support for economic integration would be equally strong in both countries, even if one were less efficient at producing both goods. For example, this is the case even if Home's potential gains from integration are larger than Foreign's, for instance because the relative equilibrium price is closer to the opportunity cost of the economically larger foreign country. In reality "full cooperation" may be difficult to achieve, and differences in potential gains from integration could distort the decision to integrate.[1] We develop an experimental design capable of assessing these trade-theoretic predictions.

4 Experimental design

The experimental design adapts to a laboratory setting an indefinitely repeated helping game where the group size is endogenous, as discussed in Bigoni et al. (forthcoming). Twenty-four participants are randomly assigned to three color-differentiated sets, or "countries" (8 per country 1, 2 and 3), for the duration of an experimental session consisting of multiple rounds of play.

A round of play. Interaction in a round takes the form of a "helping game" in pairs drawn from an interaction group of either size $N = 2$ (a "partnership") or $N \geq 12$ (a "mixed group"). Each pair is composed of a producer and a

[1] For instance, Buchanan and Brennan (1985, p.1) note: "the *rules* that coordinate the actions of individuals are important" and may lead to more or less efficient results.

consumer. The producer has a good worth 6 points to him, and can choose to give it to his opponent ("cooperate") or just consume it himself ("defect"). The consumer has no endowment, and earns 3 points if the producer defects. Instead, if the producer cooperates, then the consumer earns $k = 13$ points in a partnership and additional $a = 3$ points in a mixed group (see Table 1). Since the producer earns nothing from cooperating, the return from cooperation in a pair is $k + a$ in mixed groups, and k in partnerships. Both k and a will be treatment variables.

Table 1: Payoffs in a pair

	(a) Partnership ($N = 2$)		(b) Mixed group ($N \geq 12$)	
	Producer's choice		Producer's choice	
	Defect	*Cooperate*	*Defect*	*Cooperate*
Consumer:	3	k	3	$k + a$
Producer:	6	0	6	0

A supergame: Participants play an uncertain number of rounds of this game (a supergame, called "block" in the experiment). At the end of each round a random device determines whether or not a new round will be played (Roth and Murnighan, 1978). The continuation probability is $\beta = 0.75$. In the first round, consumer and producer roles are randomly assigned in each pair and subsequently deterministically alternate (e.g., producer, consumer, producer, ...). Hence, in each round half of the players in the interaction group are producers, and half are consumers. There can be two kinds of interaction groups: same-country groups of size $N = 2$, or *partnerships*, and multi-country groups of size $N \geq 12$, or *mixed groups*. In the latter, players are "strangers" because in every round producers are randomly re-matched with equal probability to any of the consumers and cannot identify past counterparts. In the former, players interact with the same counterpart and simply switch roles.

Either way, identities are undisclosed and since players' roles deterministically alternate they can benefit from cooperating. To facilitate cooperation, at least theoretically (see Section 5), at the end of each round players observe whether each pair in their interaction group has attained the same outcome or not. Producers cannot discriminate based on the consumer's country because the consumer's color is unobservable.

Session. Each session includes five supergames, starting and ending simultaneously for everyone in the session. In each supergame, participants know they will play 18 rounds at which point the game randomly continues as explained above. This ensures a minimal experience with the task across treatments and sessions. Composition and interaction group size differ across supergames: it is exogenously set in supergames 1-4, and it is endogenous in the last supergame. Participants are informed that supergames 1-2 consist of partnerships, while supergames 3-4 consist of mixed groups of size $N = 12$ with four players of each country. These groups are constructed to minimize contagion effects: players are informed that they cannot meet someone they have met in a previous supergame. We also study the reverse order 12-12-2-2.

Before supergame 5 starts, players are *provisionally* assigned to a mixed group with all 24 participants. Then, everyone must express a preference for either (i) leaving the mixed group for a partnership ("leave"); (ii) excluding a country from the mixed group ("exclude"); or (iii) keeping the mixed group as it is ("stay"). Once everyone makes a choice, the computer randomly selects one country, with equal probability. The majority choice in the selected country determines the outcome. There are two possible outcomes:

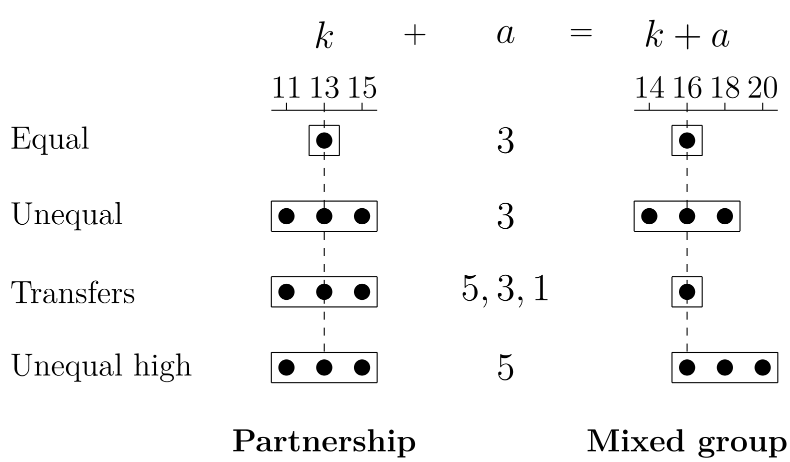
- Everyone interacts in a mixed group of 24.

- One country leaves or is excluded: those players form partnerships with someone not met before; everyone else is in a mixed group of 16.

Players are informed that only one of the five supergames completed is randomly selected for payment at the end of the experiment. The points earned in that supergame are converted into dollars according to a pre-announced conversion rate of USD 0.18.

Treatments. There are four treatments. It is convenient to discuss them by interpreting the consumer's cooperation payoff in partnerships k, and mixed groups, $k + a$, as the payoffs under autarky and economic integration, respectively. Given this interpretation, the ratio a/k is the *potential gain from integration*. The treatments differ in how consumers' cooperation payoffs and gains from integration are distributed across countries. Figure 3 explains how.

Table 2: Distribution of consumers' payoffs from cooperation



Notes: In the experiment players are divided into three equal sets of color green, red and blue, with parameter $k = 11, 13, 15$, respectively, in all treatments except Equal (where all colors have $k = 13$). Here, for convenience, we map colors into country 1, 2 and 3.

The Equal treatment is our baseline: there are no payoff-relevant differences across countries because the cooperation payoff is $k = 13$ in all partnerships and $k + a = 16$ in mixed groups. We need this treatment as a control, and to ensure we can replicate results in previous experiments. In the other two treatments we introduce a mean-preserving spread on the payoff k: it is either 11 (country 1), 13 (country 2) or 15 points (country 3).[2]

In the UNEQUAL treatment, each country has a theoretical gain from integration of 3 points, which however leads to payoff differences within cooperative mixed groups, as $k + a$ is country dependent. In the TRANSFERS treatment, instead, we remove payoff differences within cooperative mixed groups by varying the theoretical gain from integration a across countries: it is 5 points for consumers in country 1, and 1 point for country 3; it remains at 3 points for country 2. Finally, the UNEQUAL-H treatment manipulates the UNEQUAL treatment by more than doubling the gains from integration, $a = 5$ instead of 3 points, to study if and how the size of gains from integration affects behavior.

Remarks. The first four supergames familiarize players with partnerships and mixed groups. The order play 2-2-12-12 facilitates a gradual learning of increasingly complex environments, as compared to 12-12-2-2, and should facilitate cooperation in mixed groups (Weber, 2006). The group choice allows players to control both the degree of heterogeneity in the group, as well as its size, which is novel in the experimental literature on supergames among strangers. The group selection procedure gives equal weight to the majority opinion across countries, as well as the opinion of each player within a country. It is also simple to understand.

[2] In interpreting the results according to trade-theoretic intuitions, we sometimes say that the country with $k = 11$ is *weak*, is *strong* if $k = 15$, and otherwise is *middle*. It follows that in EQUAL all three countries are "middle," while in the other treatments there is one country of each kind, weak, middle, and strong.

In supergame 5, the choice to "leave" is akin to expressing a desire to unilaterally exit an integrated economy. Going into a partnership lowers maximum theoretical earnings, but allows the player to exploit reputational strategies. Choosing to exclude a country does not grant this reputational benefit because the player remains in a mixed group of 16 players. This choice can be seen either as a way to punish players of a specific country forcing them into low-return partnerships, or to have a more advantageous organization of mixed groups removing undesirable counterparts. Either way, the choice to exclude others can be interpreted as indicative of lack of cohesion among the players.[3]

Experimental procedures. The experiment was conducted at the Economic Science Institute's laboratory at Chapman University and involved 768 undergraduate students that were recruited between 2/2017 and 01/2018. We ran 8 sessions per treatment, each with 24 participants. On average, participants were paid USD 31.31, including a show-up fee of USD 7 and the payoff from an incentivized quiz on the instructions that was taken 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 the possibility to follow on individual copies. We used neutral language for the instructions (words like "cooperation" or "help" were never used). The experiment was programmed using the software z-Tree (Fischbacher, 2007). No eye contact was possible between participants. We collected demographic data in an anonymous survey at the end of each session.

[3]The instructions inform participants that they will have an opportunity to alter size and composition of their interaction group in supergame 5, without providing specific details until the end of supergame 4. Doing so minimizes the chance that behavior in the initial four supergames is affected by the intent to avoid a future possible exclusion.

5 Predictions and hypotheses

It is helpful to map the payoffs in Tables 1-2 into the trade-theoretic interpretation in Figure 1. Consider payoffs in the average round of a supergame:

- Full defection corresponds to *self-sufficiency* (point $S + S^*$): the player earns 4.5=(6+3)/2 points (all groups and treatments);

- Full cooperation in a partnership corresponds to *autarky* (point $A + A^*$): the player earns 5.5, 6.5 or 7.5 points, depending on country and treatment; the average is 6.5 in all treatments.

- Full cooperation in a mixed group corresponds to *integration* (point T): the player earns 7, 8, 9 or 10 points, depending on country and treatment; the average is 8 except in Unequal-H where it is 9 points.

Let theoretical *surplus* be measured by the difference between the average round payoff under full cooperation and full defection. Surplus is positive for all players in all treatments, it is on average $6.5 - 4.5 = 2$ points in partnerships and $8 - 4.5 = 3.5$ points in mixed groups of all treatments but Unequal-H (where it is 4.5). Dividing average surplus by its maximum level gives us a measure of theoretical *efficiency*, which ranges from 0 percent under full defection, to 57 percent under full cooperation in partnerships (autarky), to 100 percent under full cooperation in mixed groups (integration). It follows that economic integration is Pareto-dominant, because the theoretical surplus is maximum for every player. The key question is: is full cooperation part of an equilibrium?

Proposition 1. *Full cooperation is a sequential equilibrium in every interaction group and in every treatment.*

Proof. See Appendix A □

The proposition is proved by demonstrating that an informal institution, or social norm, can be constructed that supports the efficient equilibrium even if players ignore the past conduct of their direct counterparts. Specifically, the proof is a version of Kandori (1992, Proposition 1), extended to players with heterogeneous payoffs. It hinges on all players adopting the following trigger strategy: the player always cooperates as a producer, but will forever stop cooperating as soon as some producer defects. Since at the end of each round everyone sees whether or not outcomes differed across meetings, defecting in cooperative equilibrium triggers an immediate and permanent collective sanction: full defection. Full defection is always an equilibrium in the continuation game because defection is always a best response to everyone else defecting. Hence, players have always an incentive to participate in the collective sanction. In the proof of Proposition 1 we derive a condition ensuring that there is no incentive to deviate in equilibrium: if the continuation probability $\beta \geq \beta^* := 6/(a+k-3)$ then the player has no incentive to deviate in cooperative equilibrium. It follows that the threshold discount factor β^* varies across interaction groups, treatments, and countries, as indicated in Table 3.

Table 3: Threshold discount factor β^*

	Equal		Unequal		Transfers		Unequal-H	
	P	MG	P	MG	P	MG	P	MG
Country 1	.60	.46	.75	.55	.75	.46	.75	.46
Country 2	.60	.46	.60	.46	.60	.46	.60	.40
Country 3	.60	.46	.50	.40	.50	.46	.50	.35

Notes: P= Partnership, MG= Mixed group.

The largest value of β^* is 0.75. As the continuation probability is 0.75 in the experiment, full cooperation is part of a sequential equilibrium in any group of any treatment. Yet, multiple equilibria exist, including full defection, which begs the question: should we expect full cooperation in the experiment?

Trade theory suggests a positive answer since full cooperation is Pareto efficient. However, earlier experiments reveal that cooperation rates are rarely close to 100 percent and decline as groups get larger (Camera et al., 2013). Previous experiments indicate that higher threshold discount factors may discourage cooperation (Dal Bó and Fréchette, 2018). In our design variation in the payoff parameters induces variation in thresholds β^* across players (Table 3). If risk aversion plays a role, then we should see that players with lower β^* should be more cooperative as compared to players with higher thresholds. This leads us to the first hypothesis.

Hypothesis 1. *There is a negative association between a subject's cooperation rate and threshold discount factor* β^*.

Experiments have shown that payoff inequality harms the efficient provision of a public good in groups of partners (Tavoni et al., 2011). This observation suggests an alternative to the trade-theoretic prediction of full efficiency. We can test this alternative prediction, since in our strangers setting asymmetries in consumer's payoffs from cooperation induce earnings inequality in cooperative mixed groups (see Table 2). Because efficiency is proportional to the cooperation rate, inefficiency may result if players act uncooperatively as a way to reduce the payoff asymmetries.

Hypothesis 2. *In mixed groups, the cooperation rate is higher when consumers' payoffs are homogeneous as compared to heterogeneous.*

What group choices should we expect in the experiment? The trade-theoretic perspective suggests that no one should choose "leave," because cooperation payoffs are larger in mixed groups. In our design this payoff increment is a/k which, given the theme of our study, is called the *potential gain* *from integration*; it varies between 7% and 45% across countries and treatments (Table 2). An earlier study already suggests that the trade-theoretic prediction is empirically weak. In Bigoni et al. (forthcoming), subjects prefer partnerships to large groups where consumers' cooperation payoffs are 20% larger. But the study leaves open two questions. First, whether the option to exclude others from a large group can reverse or mitigate the preference for partnerships. Second, whether the size of the potential gain from integration affects group choices and how. It is not obvious which of the alternatives will be most attractive in our design, but it is natural to conjecture that the relative attractiveness of partnerships should diminish as the size of potential gains increases. We thus formulate two further hypotheses.

Hypothesis 3. *When countries' potential gains from integration are identical, the distribution of group choices will not differ across countries.*

Hypothesis 4. *Larger potential gains from integration are associated with a lower probability to choose "leave."*

6 Results

Our unit of observation will be a subject in a supergame (unless otherwise noted), which gives us $N = 64$ observations per country, per supergame, per treatment. We will work with three main endogenous variables: a subject's cooperation rate, realized payoff and realized efficiency.

The *cooperation rate* is the relative frequency of cooperative choices of a subject as a producer in a supergame, which ranges from 0 to 100 percent. *Realized payoff* corresponds to the points earned by the subject in the average round of the supergame.[4] *Realized efficiency* measures the average realized

[4] In the experiment, let $c_{it} = 1$ denote a cooperative action by player i in period t (0, if defection or if no choice is taken). Let t_p corresponds to the number of periods in which

payoff in a group relative to the maximum attainable payoff and is directly proportional to the cooperation rate in the group.

We begin by reporting findings about cooperation and realized efficiency in supergames 1-4 where everyone experienced interaction both in partnerships as well as in mixed groups.[5] We then present the results of the group choices expressed by subjects before supergame 5 starts.

6.1 Cooperation and realized efficiency

In this section we report results for supergames 1-4, unless otherwise noted. The first result provides mixed support for Hypothesis 1.

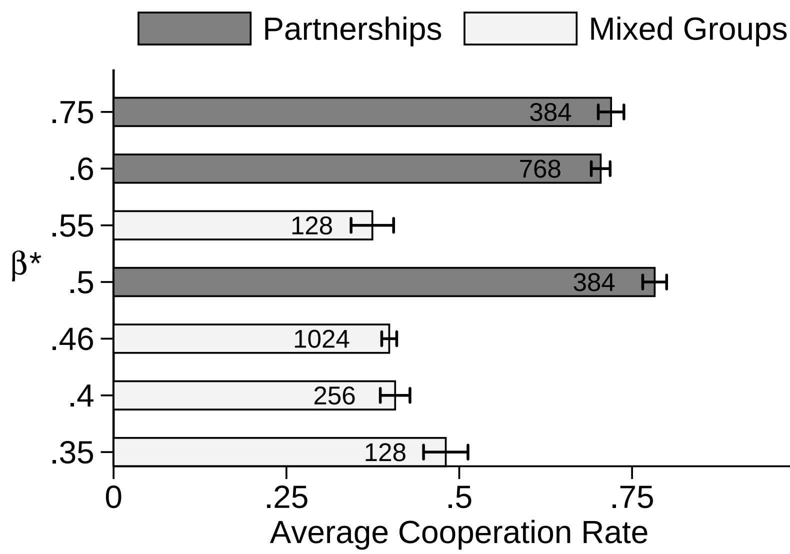
Result 1. *In partnerships, subjects with different threshold discount factors cooperated similarly. In mixed groups, there is a negative association between a subject's threshold discount factor and cooperation rate.*

Evidence is provided by Figure 2 and Table 4. Figure 2 reports average cooperation for each threshold discount factor β^*, by group size for data pooled across all treatments. Partnerships are generally much more cooperative than mixed groups. Cooperation in partnerships varies between 70.5% and 78.3%, when we consider the different β^* thresholds. In mixed groups it varies between 37.4% and 48.1%.

this player was a producer in the supergame. The cooperation rate for this player is $\sum_{t=1}^{t_p} c_{it}/t_p \in [0, 1]$. If we let π_{it} denote the payoff to the player in round t and s be the number of periods of the supergame, then the average realized payoff by player i in the average period of the supergame is $\sum_{t=1}^{s} \pi_{it}/s$.

[5] Subjects have different interactions in supergame 5, depending on realized composition of groups. Data for supergame 5 are presented in the Supplementary Material.

Figure 2: Cooperation by threshold discount factor β^*



Notes: One obs.=one subject in a supergame. Supergames 1-4 only, data pooled from all treatments. *Cooperation:* relative frequency of cooperation of a subject in a supergame (average by β^*). The whiskers identify the mean standard error. The number reported on each bar corresponds to the number of observations.

To establish whether this variation (within a group) across β^* is significant, we ran GLM regressions (see Table 4). The dependent variable is the average cooperation rate of a subject in a supergame. We regress it on the subject's threshold discount factor β^*, standardized, as well as a dummy for each treatment (the EQUAL treatment representing the base value). To soak up order effects, we included a dummy taking value one if the session started with two consecutive supergames of mixed groups (zero, otherwise). To control for any learning effects, we included a dummy taking value one if the subject was in a partnership for the second consecutive supergame (zero, otherwise), and a similar dummy variable for mixed groups. Controls are not reported and include: the duration of the current supergame, the duration of the previous supergame, the number of right answers, the response time in an incentivized comprehension test after the instructions and self-reported sex from a survey after the experiment.

Table 4: How β^* affects an individual's cooperation: marginal effects

Dep. var.: subject's cooperation rate	Partner. (1)	Mixed g. (2)
β^*	-0.036	-0.062**
	(0.026)	(0.030)
Treatment dummies		
Unequal	0.094**	0.012
	(0.043)	(0.051)
Transfer	0.094**	-0.035
	(0.045)	(0.048)
Unequal high	-0.008	-0.005
	(0.040)	(0.048)
Order dummies		
12-12-2-2	0.028	-0.158***
	(0.033)	(0.033)
2^{nd} game	0.112***	-0.070***
	(0.016)	(0.027)
Controls	Yes	Yes
N	1536	1536

Notes: One obs.=one subject in a supergame. Data from supergames 1-4; column (1) data for partnerships, column (2) data for mixed groups. GLM regressions on the average cooperation rate, with standard errors in parentheses robust for clustering at the session level. The regressor β^* is the standardized value of the values reported in Table 3. Equal treatment is the base value. *12-12-2-2*=1 if the session started with two supergames of mixed groups (else, 0). 2^{nd} *game*=1 if it was the second supergame for the given group composition (else, 0). *Controls* include duration of the supergame, duration of the previous supergame, self-reported sex, and two measures of understanding of instructions (response time and wrong answers in the quiz). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. Marginal effects are computed at the mean value of regressors of continuous variables.

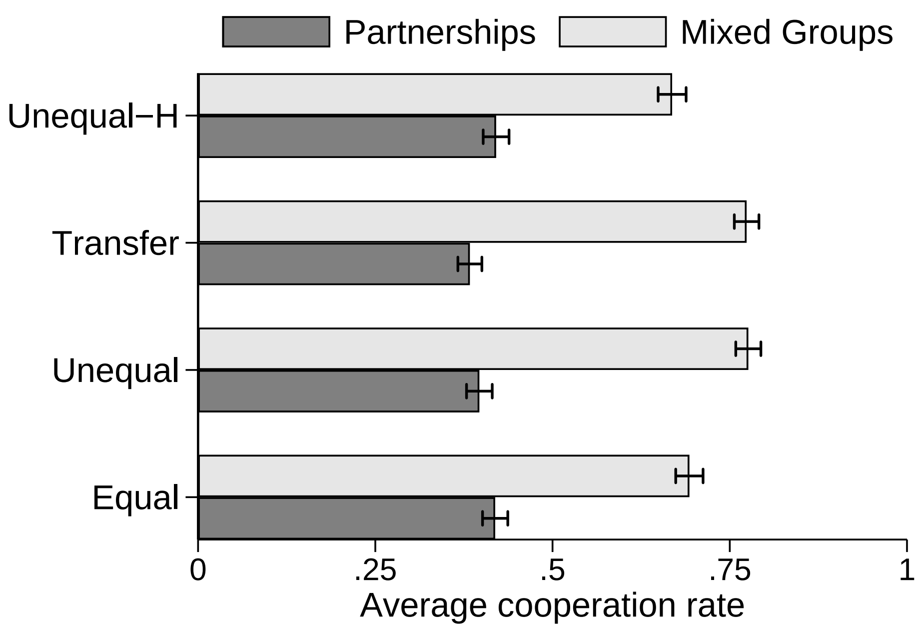
Table 4 reveals that the coefficient on β^* is negative, but significant only in mixed groups. Therefore, Hypothesis 1 cannot be rejected for mixed groups; increasing the threshold discount factor by a standard deviation decreases the subject's cooperation rate by 6 percentage points. A regression in which we use discrete instead of continuous threshold regressor (see Supplementary Material table 11) reveals that the driving factor in mixed groups is subjects with the lowest threshold discount factor.

The regression also shows the presence of order effects. As in previous experiments, we find that cooperation declines as we move from partners to strangers. Yet, this decline is smaller when mixed groups followed partnerships as compared to when mixed groups preceded partnerships (the *12-12-2-2* dummy is negative and highly significant), possibly due to experience factors. Indeed, we find that experience with the game had opposite effects on cooperation depending on the size of the group. It raised cooperation in partnerships (2^{nd} *game* dummy is positive and highly significant), but lowered it in mixed groups (2^{nd} *game* dummy is negative and highly significant). In a way, only partners learned to build trust and to cooperate, while strangers did not. Did the heterogeneity in consumer's cooperation payoffs $k + a$ affect cooperation in mixed groups?

Result 2. *Mixed groups in which consumers have unequal payoffs cooperated similarly to mixed groups in which consumers have identical payoffs.*

Evidence is provided by Figure 3 and Table 5. Figure 3 reports average cooperation by treatment and group size. In mixed groups, it varies between 38.3% and 42% across treatments. To establish whether this variation is significant, we ran a GLM regression (see Table 5).

Figure 3: Cooperation by treatment



Notes: One obs.=one subject in a supergame 1-4. $N = 384$ per treatment and groupsize. *Cooperation:* relative frequency of cooperation of a subject in a supergame (average by treatment). The whiskers identify the mean standard error.

We regress the average cooperation rate in a mixed group on treatment dummies that capture the various configurations of heterogeneity in cooperation payoffs across players in groups. Recall from Table 2 that players in a mixed group have different values $k+a$ only in Unequal and Unequal H, in Transfers there is only inequality across partnerships (the value k), while in Equal (the base) there is no inequality at all. As before we control for order effects, learning effects, and included the standard controls. There is no evidence that heterogeneity in cooperation payoffs affected cooperation in mixed groups; the coefficients on the treatment dummies are all insignificant. Hence, we reject Hypothesis 2. We can also reject the hypothesis that Unequal and

UNEQUAL H are jointly different from the EQUAL treatment (two-sided Wald test, p-value=0.755).

Table 5: Cooperation in mixed groups: marginal effects

Dep. var.: cooperation rate in a mixed group	
Treatment dummies	
Unequal	-0.006
	(0.049)
Transfer	-0.044
	(0.051)
Unequal H	0.029
	(0.047)
Order dummies	
12-12-2-2	-0.152***
	(0.034)
2^{nd} game	-0.071***
	(0.027)
Controls	Yes
N	128

Notes: One obs.=one group in a supergame. Data from mixed groups in supergames 1-4. GLM regression on average cooperation rate, with standard errors in parentheses robust for clustering at the session level. *Treatment dummies:* the EQUAL treatment is the base. *12-12-2-2*=1 if the session started with two supergames of mixed groups (else, 0). 2^{nd} *game*=1 if it was the second supergame of mixed groups (else, 0). *Controls* include duration of the supergame, duration of the previous supergame, average self-reported sex, and two measures of understanding of instructions (average response time and average wrong answers in the quiz). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. Marginal effects are computed at the mean value of regressors of continuous variables.

We conclude this section by studying the impact of the potential gain from integration a/k on realized efficiency. Consumers can potentially earn from 7% to 45% more in a mixed group than in a partnership. The average gain is 23% (3/13) in all treatments, except in UNEQUAL-H, where it jumps to 38% (5/13). Interacting in mixed groups can thus improve average *realized* payoffs even if cooperation drops relative to partnerships. Empirically, this is indeed what we observe if subjects had prior experience with partnerships before moving into mixed groups. We also observe that the order of play—starting with partnerships or mixed groups—matters. This leads to the following result.

Result 3. *Realized efficiency is larger in mixed groups than partnerships, when mixed groups followed partnerships. The reverse holds true when partnerships followed mixed groups.*

Evidence is provided by Table 6. We ran a regression that pools data from all treatments for supergames 1-4. The dependent variable is the average payoff realized in a supergame of a session, which is proportional to realized efficiency in the average group of that supergame. Since subjects interacted either in mixed groups or partnerships, depending on the supergame, we interact the *Group* covariate with the *Order* covariate to trace if the observation is about a partnership or a mixed group, and whether the session has order 2-2-12-12 or the reverse. The case in which the subject is in a partnership in supergames 1-2 is the base. We include treatment dummies (Equal treatment representing the base value) and standard controls. The interaction term *Mixed group* × *2-2-12-12* is positive and significant, while *Mixed group* × *12-12-2-2* is negative and significant. Notice also that the order had no effect on efficiency in partnerships (the coefficient on the *Partnership* × *12-12-2-2* is positive but not statistically significant). The coefficient on the Unequal-H dummy is positive because the payoff from cooperation in mixed groups is by design 2 points larger than in the other treatments.

Table 6: Realized payoff comparison: partnerships vs. mixed groups

Dep. var.: Realized payoff	
12-12-2-2	0.051
	(0.095)
Mixed Group	0.426***
	(0.120)
Mixed Group × 12-12-2-2	-0.753***
	(0.175)
2^{nd} game	0.055
	(0.043)
Treatment dummies	
Unequal	-0.002
	(0.113)
Transfer	-0.069
	(0.112)
Unequal-H	0.299**
	(0.117)
Controls	Yes
Constant	5.726***
	(0.188)
N	128
R^2	0.376
adj R^2	0.317

Notes: One obs.=one supergame in a session, N=32 per group size and treatment. Data from supergames 1-4. Linear regression on average realized payoff, with standard errors in parentheses robust for clustering at the session level. *Mixed group*=1 if it was a mixed group (else, 0), *12-12-2-2*=1 if the session started with two supergames of mixed groups (else, 0). *Treatment dummies:* EQUAL treatment represents the base value. *Controls* include duration of the supergame, duration of the previous supergame, self-reported sex, and two measures of understanding of instructions (response time and wrong answers in the quiz). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

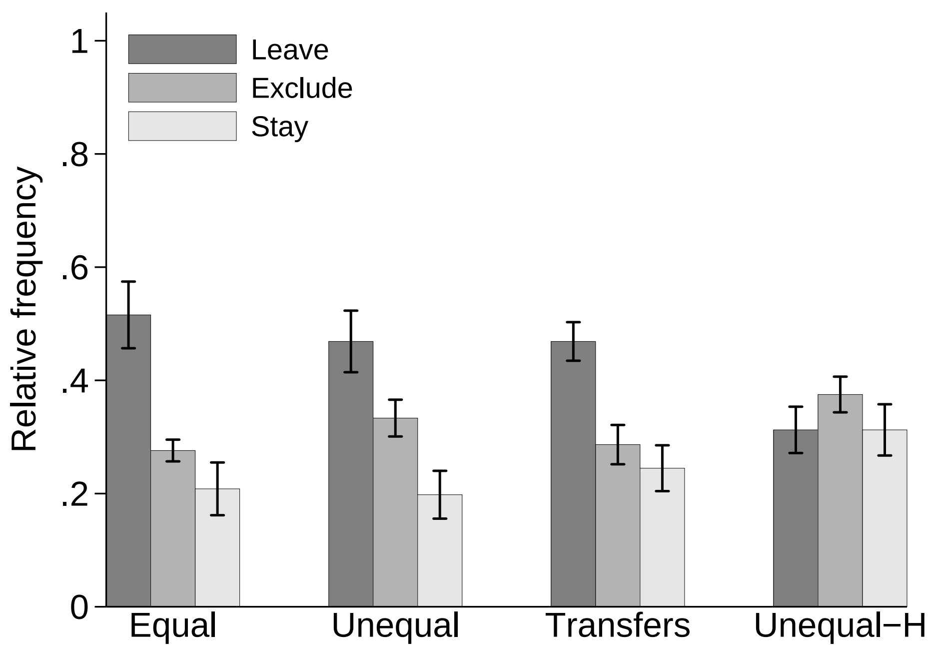
6.2 Group choices

We start by investigating how the potential gain from integration a/k affected group choices of individuals. Recall that in most treatments the average potential gain from integration is 0.23, a value that increases to 0.38 only in the UNEQUAL-H treatment. A first question is thus whether we observe differences in group choices between UNEQUAL-H and the other treatments.

Result 4. *In* UNEQUAL-H*, group choices are equally distributed. In all other treatments, "leave" prevails, followed by "exclude" and "stay".*

This result is in line with Hypothesis 4. Evidence comes from Figure 4 and a series of ranksum tests. The figure reports the distribution of the relative frequency of "leave," "exclude" and "stay," by treatment. The unit of observation is the relative frequency of a given choice in a session (N=8 per each possible choice, per treatment).

Figure 4: Overall distribution of group choices in a session



Notes: One obs.= one session ($N = 8$ per treatment). Relative frequency of choice "leave", "exclude" and "stay." The average potential gain from integration a/k is 23%, in EQUAL, UNEQUAL and TRANSFER, and 38% in UNEQUAL H. The whiskers identify the mean standard error.

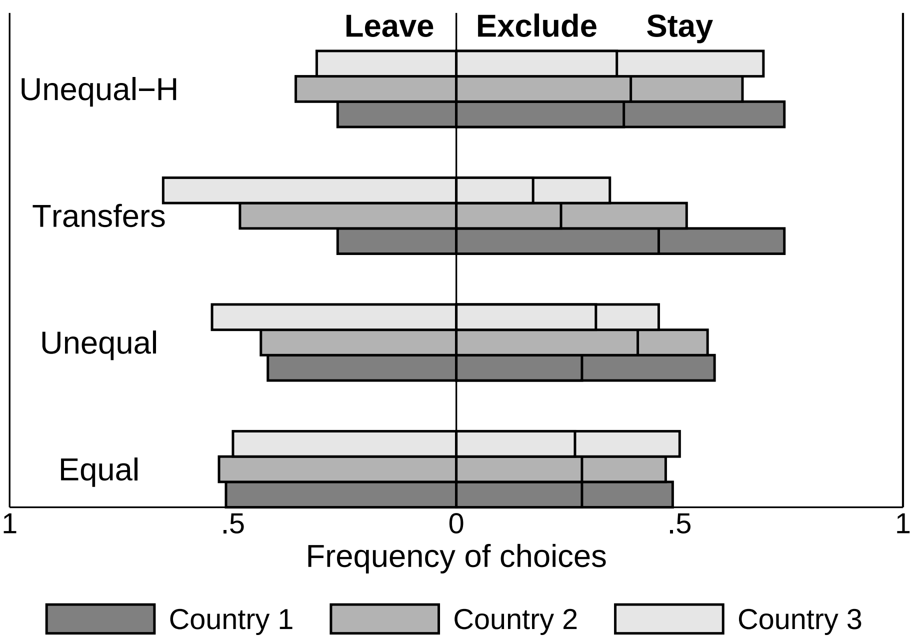
To assess the statistical significance of these observations, we use two-sided Wilcoxon-Mann Whitney ranksum test with exact statistics. The difference between the frequency of "leave" and "exclude" is significant in treatments where the potential gain from integration was on average 0.23 but not when it rose to 0.38 (p-values=0.008, 0.046, 0.005, and 0.202 respectively, for Equal, Unequal, Transfer, and Unequal H, N1=N2=8). A similar result holds for the difference in frequency between "leave" and "stay" choices (p-values=0.006, 0.002, 0.001, and 1.00 , N1=N2=8). The frequency of choice "exclude" and "stay" is significantly different in Equal and Unequal but not in Transfer and Unequal H (p-values=0.042, 0.044, 0.589, and 0.20, N1=N2=8). This is evidence that subjects had a disposition to leave but this attitude changed when we raised the average potential gain from integration by 65% (from 0.23 to 0.38) in line with Hypothesis 4.

We now disaggregate the data at the subject level to investigate whether the size of potential gains from integration a/k affected the subject's group choice. We begin by investigating whether payoff-irrelevant differences across subjects play a role. It is indeed possible that separating subjects into different-color groups might itself affect group choices by creating some form of group identity. To test this hypothesis we study choices in the Equal treatment where the potential gains from integration are identical across subjects.

Result 5. *When there are no disparities in potential gains from integration, subjects from different countries make similar group choices.*

Evidence comes from Figure 5 and Table 7. The figure reports the distribution of the relative frequency of "leave," "exclude" and "stay," by treatment. The unit of observation is one subject in a country (N=64 per each possible choice, per treatment). The figure allows a direct comparison of group choices when differences in potential gains from integration are and are not present in the session. Subjects assigned to different countries made identical group choices in the EQUAL treatment. Hence, it suggests that the payoff-irrelevant heterogeneity in colors we introduced in the design had no effect on group choices.

Figure 5: The frequency of group choices by country and treatment



Notes: Distribution of choices for "leave" (left), "exclude" and "stay" (right). One obs. is one subject in a session (N=96 per treatment).

To test this hypothesis, we ran the multinomial logit regression in Table 7. One observation is one subject in a session of the EQUAL treatment. The outcome variable is the three group choices, 1, 2 and 3 for "leave," "exit" and "stay" respectively. The dummies *Country 1*, 2 (the base) and 3 identify the color of the subject. We include three continuous variables that control for the variation in outcomes experienced by subjects in supergames 1-4. The predictor variable *Realized gains from integration* is the realized net gain/loss in mixed groups relative to partnerships (the potential gain is identical for everyone, $a/k = 3/13$). The regressors *Full Cooperation in Partnerships* and *Cooperation Bias in Mixed Groups* account for experience in supergames 1-4. The first regressor takes the value one if the subject experienced full cooperation in at least one partnership, and zero otherwise. The second regressor, is the difference in cooperation rate between the subject and her opponents in Mixed Groups, which can range from -1 to 1 thus identifying the cooperation bias of this subject (cooperator or defector).[6] Finally, we include an order dummy, and standard controls.

Group choices are similar across countries; none of the coefficients on the *Country* dummies is significant and we cannot reject the hypothesis that the coefficients on the Country 1 and Country 3 dummies are statistical similar (Wald tests, p-values 0.419, 0.440, 0.765 for choice 1, 2 and 3 respectively). Hence, Hypothesis 3 cannot be rejected. We conclude that the separation of subjects into different-color sets (i.e., countries) had no behavioral effect on group choices.

Although the *potential* gains from integration are identical for everyone in the EQUAL treatment, there was variation in realized gains. The gains realized in the past affected choices, lowering the probability to choose "leave," as opposed to interacting in a mixed group. One standard deviation in realized gains decreases the probability of choosing "Leave" by 20 percentage points; it significantly and equally increases the probability of each of the other two choices (Wald test, p-value=0.892).

[6] To be specific, it is 1 for subjects who always cooperated and always met defectors in mixed groups; it is -1 in the opposite scenario; it is 0 if the subject cooperated with the same frequency of her opponents. In the regression we standardize this measure. See (Bigoni et al., forthcoming).

Table 7: Group choices in the EQUAL treatment: marginal effects

Dep. variable=1 if option chosen (else, 0)	Leave	Exclude	Stay
Country 1	-0.042	0.030	0.012
	(0.125)	(0.130)	(0.053)
Country 3	0.033	-0.033	0.000
	(0.114)	(0.098)	(0.057)
Realized gains from integration	-0.199***	0.105**	0.094*
	(0.057)	(0.048)	(0.051)
Full Cooperation in Partnerships	0.121*	-0.079	-0.042
	(0.073)	(0.061)	(0.060)
Cooperation Bias in Mixed Groups	-0.011	0.035	-0.024
	(0.049)	(0.036)	(0.030)
Controls	Yes	Yes	Yes

Notes: One obs.=one subject in a session in the EQUAL treatment ($N = 192$). Multinomial logit regressions on preferences for the three choices "Leave", "Exclude" and "Stay". Robust standard errors in parentheses are adjusted for clustering at the session level. *Realized gains form integration:* standardized percentage gains from integration realized in supergames 1-4. *Full Cooperation in Partnerships*=1 if the subject experienced full cooperation in at least one partnership (0 otherwise). *Cooperation Bias in Mixed Groups*: (standardized) difference in cooperation rate between the subject and her opponents in Mixed Groups (supergames 1-4). *Controls* not reported: total number of rounds played in supergames 1-4, order of groups (partnerships and then mixed groups or reverse), self-reported sex, and two measures of understanding of instructions (response time and wrong answers in the quiz). Marginal effects are computed at the regressors' mean value (at zero for dummy variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Finally, success with reciprocal interaction had a strong drawing power on subjects. Those who successfully coordinated on efficient play in partnerships selected "Leave" more frequently than those who did not (see the coefficient on the regressor *Full Cooperation in Partnerships*). Since defectors by definition could not experience efficient play, this is an indicator that free-riders had an inclination to interact with strangers—a result that confirms the findings in (Bigoni et al., forthcoming).

The open question is whether the group choice is affected by the *potential* gains from integration. This question cannot be addressed in Bigoni et al. (forthcoming), because there the potential gain from integration is 20% across all players and all treatments ($a = 3$ and $k = 15$ pts.); that study reveals that the majority of players selects partnerships. Can this result be reversed if moving away from partnerships enables larger potential gains? In our design the potential gains from integration vary between approximately 7% (a/k =1/15 for country 3 in Transfer) all the way up to 45% (5/11 for country 1 in Unequal-H and Transfer). Figure 5 suggests that group choices choice are affected by the potential gains from integration a/k. Recall that Country 1 subjects had the highest potential gain, while Country 3 the lowest. In Figure 5, Country 3 subjects choose "leave" more often than the rest, while Country 1 subjects choose either "stay" or "exclude" more often. Table 7 suggest that many factors—including the subject's individual experience in supergames 1-4—could influence these choices, in addition to disparities in potential payoffs. To look into this issue, we ran an additional econometric analysis summarized in the following result.

Result 6. *Group choices are elastic to the potential gain of integration. The probability to choose "Leave" falls as a/k increases.*

Evidence comes from the multinomial logit regression in Table 8. One observation is one individual in a session. The outcome variable is the three group choices, 1, 2 and 3 for "leave," "exit" and "stay" respectively. There are two continuous predictor variables involving the gains from integration: *Realized gains from integration* by the subject in supergames 1-4, and *Potential gains from integration*, i.e., a/k. As before, we also include the regressors *Full Cooperation in Partnerships* and *Cooperation Bias in Mixed Groups* to trace the impact of the subject's experience in supergames 1-4.

It is possible that the magnitude of disparities in potential gains from integration may itself affect group choices. For this reason we include the regressor *Disparity*, which measures the relative difference between the potential gains at the top and the bottom of the distribution of a/k. This measure is 1 when all subjects have identical potential gains from integration a/k (Equal treatment), 1.36 when the potential gains differences are small (Unequal and Unequal H treatments), and 6.82 when the differences are much larger (Transfers treatment); see Table 2. Finally, we include an order dummy, and standard controls.

The *Realized* regressor soaks up the effect of the gains from integration realized in supergames 1-4 on group choices. As expected, greater gains induce a shift of probability mass from "Leave" towards "Exclude" and "Stay." A one-standard deviation increase in the coefficient on the *Realized* regressor reduces the probability to "Leave" by 27.1 percentage points, while it equally increases the probabilities to select "Exclude" and "Stay" (Wald test, p-value=0.987)

The coefficients on the *Potential* regressor reveal that the theoretical gains from integration also help predict group choices. A standard deviation increment in potential gains from integration a/k significantly alters probability to choose "leave" and "exclude", reducing the first by 5.5 percentage points and raising the latter by 3.5 percentage points. The effect on the probability to "Stay" is positive but not significant. These results are in line with Hypothesis 4. The effect of potential gains from integration reinforces the effect of realized gains, but it is of second order importance: for each possible group choice, the coefficients on *Realized* and *Potential* are significantly different (two-sided Wald test, p-values < 0.001 and $0.010, 0.008$, respectively for "Leave," "Exclude," and "Stay").

There is no evidence that the magnitude of disparities in potential gains from integration played a role in selecting group choices; the coefficient on the

Disparity regressor is statistically indistinguishable from zero.

Table 8: Group choices in all treatments: marginal effects

Dep. variable: =1 if option chosen (else, 0)	Leave	Exclude	Stay
Gains from integration			
Realized	-0.271***	0.135***	0.136***
	(0.047)	(0.028)	(0.034)
Potential	-0.055**	0.035*	0.019
	(0.026)	(0.021)	(0.019)
Full Cooperation in Partnerships	0.152***	-0.066*	-0.087***
	(0.041)	(0.038)	(0.034)
Cooperation Bias in Mixed Groups	-0.054	0.061**	-0.007
	(0.039)	(0.028)	(0.020)
Disparity	-0.006	-0.001	0.007
	(0.010)	(0.008)	(0.010)
Controls	Yes	Yes	Yes

Notes: One obs.=one subject in a session ($N = 768$). Multinomial logit regressions on preferences for the three choices "Leave", "Exclude" and "Stay". Robust standard errors in parentheses are adjusted for clustering at the session level. *Realized:* Standardized percentage gains from integration a subject realized in supergames 1-4. *Potential:* Standardized potential gains from integration a/k for a subject. *Disparity:* measures the ratio of the largest to smallest value a/k in a mixed group. *Full Cooperation in Partnerships*=1 if the subjects experienced full cooperation in at least one partnership (0 otherwise). *Cooperation Bias in Mixed Groups*: (standardized) difference in cooperation rate between the subject and her opponents in Mixed Groups (supergames 1-4). *Controls* not reported: total number of rounds played in supergames 1-4, order of groups (partnerships and then mixed groups or reverse), self-reported sex, and two measures of understanding of instructions (response time and wrong answers in the quiz). Marginal effects are computed at the regressors' mean value (at zero for dummy variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Experiencing efficient play in partnerships raised the probability to select "Leave" by 15 percentage points (see the coefficient on the regressor *Full Cooperation in Partnerships*), while lowering the probability of the other two alternatives by an equal amount (two-sided Wald test, p-value 0.715).

The fact that subjects who do not seek a partnership are split between "exclude" and "stay" requires additional explanation. Excluding someone from a large mixed group lowers their potential payoff, without raising the payoff of those who remain. It also raises the group size from 12 to 16 (instead of 24), thus making "exclude" suboptimal for someone interested in coordinating on efficient play, but also for a free-rider—who benefits from the anonymity granted by larger groups. We next explain that "Exclude" is selected by cooperators who benefited to some extent from integration and intended to minimize free-riding by self-selecting into mixed groups of like-minded cooperators.

Result 7. *The cooperation rate of opponents affected exclusion choices. The probability to target a country for exclusion is inversely related to its relative cooperation rate.*

Support for this result comes from Tables 8 and 9. Table 8 shows that the probability to choose "Exclude" is 6 percentage points higher for a cooperative subject, i.e., someone who cooperated more than her counterparts in mixed groups; see the coefficient on the *Cooperation Bias in Mixed Groups* regressor. These subjects sought to reap the gains from integration by excluding from a mixed group the country that, according to their personal experience, had the least cooperative members. This evidence comes from the logit regressions in Table 9, where we pool data for all those subjects who choose "Exclude" (all treatments). One observation is one individual in a session. The dependent variable of regression 1 takes the value 1 if subjects from country 2 and 3 chose to exclude country 1 subjects; instead, it takes the value 0 if they chose to exclude the other country (3 and 2, respectively). Regressions 2 and 3 study the choice to exclude countries 2 and 3, in a similar manner.[7]

We estimate the effect of a country's relative uncooperativeness in mixed groups on the choice to exclude that country by subtracting the average co-

[7] Table 15 in Supplementary Materials reports the frequency distribution of exclusion choices in all treatments and across countries.

operation rate of opponents from the two other countries. The *Relative uncooperativeness* of country i regressor is the average cooperation rate in mixed groups of opponents from the non-excluded country minus that of the excluded country $i = 1, 2, 3$. We also add the *Other country* dummy to control for possible country-specific differences in the frequency of exclusion. For example, when we study the choice to exclude country 1 (column 1) the dummy takes value 1 if the subject is from country 3, and zero if from country 2. As before, we include an order dummy, and standard controls.

Table 9: "Exclude" choices: marginal effects

Dep. var.: Exclude a country=1 (exclude other country =0)	Country 1 (1)	Country 2 (2)	Country 3 (3)
Relative uncooperativeness	0.134***	0.220***	0.143***
	(0.037)	(0.046)	(0.043)
Other country	0.002	0.037	-0.115*
	(0.069)	(0.071)	(0.065)
Controls	Yes	Yes	Yes
N	155	160	173

Notes: One obs.=one subject in a session. Logit regressions on "Exclude" choices. Column (1): subjects from country 2 and 3 who chose "Exclude"; Column (2): subjects from country 1 and 3 that chose "Exclude"; Column (3): subjects from country 1 and 2 that chose "Exclude". *Relative uncooperativeness*: difference in average cooperation rate of opponents from non-excluded and excluded country (in mixed groups). *Other country*=1 if subject belongs to country 3 (in columns 1 and 2) or country 2 (column 3), and 0 if subject belongs to country 2 (in columns 1 and 2) or country 1 (column 3). *Controls* not reported: total number of rounds played in supergames 1-4, order of groups (partnerships and then mixed groups or reverse), self-reported sex, and two measures of understanding of instructions (response time and wrong answers in the quiz). Robust standard errors in parentheses are adjusted for clustering at the session level. Marginal effects are computed at the regressors' mean value (at zero for dummy variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

There is clear evidence that subjects who choose to exclude a country from the mixed group conditioned their choice on observed cooperation. The smaller the relative cooperation of the country, the greater the probability to exclude it (see the relevant coefficients on the *Relative uncooperativeness* regressor). We find no country-specific bias to exclude any given country; the coefficients on *Other country* are all indistinguishable from zero.

7 Discussion

From a trade theory perspective, phenomena such as Brexit present a puzzle. People should embrace and not avoid economic integration, because they can economically benefit from it. But underlying this view is a focus on the final allocation and not on the process that enables it—which is left in the background. The point of our study is to show that this process should be moved to the foreground to better understand the events we are witnessing. For economic integration to succeed, unrelated individuals must be able to build trust, to overcome coordination problems, and to manage the uncooperative tendencies that naturally emerge in heterogeneous and large groups (Results 1-3). Because of these problems, in our experiment individuals remained hesitant to "integrate" by forming a large trading group, even if doing so held the promise of large economic gains (Results 4-7). As a consequence, a lot of money was left on the table.

We have documented that three factors significantly influence individuals' desire to economically integrate. First, their perception of the size of *potential* gains. Predictably, subjects in the experiment were less likely to seek isolation ("leave") when they *realized* larger gains from integration. But once we account for this experience factor, we also observe that the allure of isolation falls with the size of *potential* gains. This pattern appears to be independent of whether these gains are absolute or relative.

A second factor is a trust-building process that is preparatory to economic integration. In the experiment, fixed pairs allowed subjects to easily build the trust necessary to attain full cooperation, but mixed groups did not. This "trust differential" is one of the main factors that is responsible for the allure of isolation as compared to integration. Most subjects developed an effective norm of cooperation in isolated partnership, but were utterly incapable to do so when we merged these partnerships into mixed groups of strangers.

The third factor is the availability of institutions to manage uncooperative tendencies and short-run opportunistic temptations. About half of the subjects who chose to "integrate" in mixed groups, also sought to exclude others from it. This choice is inconsistent with full integration and is motivated by the desire to mitigate opportunistic behavior within the group. Cooperators tried to self-select into more productive groups by excluding participants perceived to be the least cooperative.

If we are willing to entertain the hypothesis that these laboratory results reflect a principle of behavior that also underlies external decision processes, what can we infer about the current tendencies of economic disintegration? The experiment suggests four kinds of considerations. First, the need to strengthen institutions that promote international cooperation. This and other experiments reveal that long-run cooperation is inherently difficult among strangers, as it requires that mutual trust be established and maintained. This may explain why keeping together economic unions and multilateral trade systems seems so challenging, even if they are potentially very profitable. The institutional requirements established by the EU for new members, and the sanctioning of uncooperative members in the WTO may be ways to mitigate these challenges.

A second consideration concerns the role played by smaller countries in supporting economic integration. International trade theory indicates that small countries typically gain more from economic integration, and large countries less, both absolutely and relatively. Under this interpretation, country 3 is large in our experiment, and the data reveal that country 3 individuals exhibited only a lukewarm desire for economic integration. This suggests that it might be particularly important for small countries to take the initiative, for example by putting maximum engagement in the WTO.

A third aspect concerns the importance of combining an accurate calculation of the anticipated gains from economic integration with a clear communication of these benefits to the public. In the experiment individuals who knew to have the largest potential benefits from integration were less likely to choose "leave," all else equal, i.e., after we control for their realized gains. 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 (Rossi-Hansberg, 2017). If so, this determination should be made transparent to policymakers and practitioners, and also to the public.

An additional consideration concerns the delicate role played by cross-country redistribution policies, for example the "cohesion" policies in the EU. As the countries with a high per capita income are typically net contributors to cohesion and other redistributive policies, stronger countries transfer some of their gains from economic integration to weaker countries. According to our experiment, while this process increases the desire for the weak countries to operate in an integrated market, it also weakens the support for integration in the strong countries stirring the opposite sentiment. A cohesion policy thus must be carefully designed, as it runs the risk of backfiring.

Finally, we can think of the benefits from integration introduced in the experiment as embedding also a "political" component. If so, the loss in political sovereignty that is generally associated with economic integration may significantly reduce the overall benefits, thus contributing to explain the disintegration trends the world has been witnessing .

References

Abreu, D., D. Pierce and E. Stacchetti . 1990. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica*, 58, 1041-1063.

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.

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

Buchanan, J., and Brennan, G.(1985). *The Reason of Rules*, New York: Cambridge University Press.

Camera, G., and Casari, M. (2009). Cooperation among strangers under the shadow of the future. *American Economic Review*, 99 (3), 979-1005.

Camera, G., Casari, M., and Bigoni, M. (2013). Money and trust among strangers. *Proceedings of the National Academy of Sciences*, 110 (37), 14889-14893.

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.

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. and G. Fréchette. 2018. On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56 (1), 60-114.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2015). The geography of development. *Journal of Political Economy*, forthcoming.

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

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.

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.

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

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

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

Page, T., Putterman, L., and Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *Economic Journal*, 115 (506), 1032-1053.

Ricardo, D. (1817). *On the principles of political economy and taxation*, London: John Murray.

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 (pp. 197-205). Cham: Springer.

Roth, E., and K. Murnighan (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17, 189-98.

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

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.

R. A. Weber, 2006. Managing Growth to Achieve Efficient Coordination in Large Groups. *Am. Econ. Rev.*, 96(1), 114-126.

Tavoni, Alessandro, Astrid Dannenberg, Giorgos Kallis, and Andreas Löschel. 2011. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (29), 11825-11829.

A Appendix

A.1 Proof of proposition 1

Here we prove that full cooperation is a sequential equilibrium in every group 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 5. For convenience let the defection payoffs be, respectively, d and $d-l$ to a producer and a consumer. Given this notation, a necessary and sufficient condition is reported in the following lemma:

Lemma 1. *Fix an interaction group. Let $a+k$ denote the smallest cooperation payoff in that group. If the continuation probability*

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

then full cooperation is a sequential equilibrium.

To prove it, note that producers cannot discriminate consumers according to their country, because in mixed groups this information is not observed. So consider the cooperation payoff to someone with the lowest return from cooperation, i.e., consider $a=0$ and the smallest k This player, as well as every other player, deterministically alternates between producer and consumer roles over the course of the supergame, earning k every other round, in equilibrium. Let $s=0,1$ denote the role of the player at the start of a round, where 0 is for a producer and 1 for a consumer. Payoffs are independent of the size N of the group. In cooperative equilibrium the player nets

$$v_0 := \frac{\beta(a+k)}{1-\beta^2} \qquad \text{and} \qquad v_1 := \frac{a+k}{1-\beta^2},$$

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

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

It is immediate that off-equilibrium a producer has no incentive to deviate from the sanctioning rule, 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, by refusing to help some consumer.[8] This inequality can be rearranged as in Proposition 1. Note that $\beta^* < 1$ because $a + k - (2d - l) > 0$ by assumption for all player types. The Lemma exploits the fact that the lowerbound probability β consistent with cooperation is a decreasing function of the player's return from cooperation $a + k$. Hence $\beta \geq \beta^*$ ensures that players who have higher returns from cooperation in the group also have an incentive to cooperate. Proposition 1 follows from observing that in the experiment $\beta = 0.75$ and the most stringent requirement comes from fixed pairs composed of weak country players, in which case $a = 0$ and $k = 11$, so $\beta^* = 0.75$.

Table 10: Cooperation payoffs and gains from integration across treatments

		Mixed group payoff $k+a$ (Integration)	
		16	16 ± 2
Partnership payoff k (Autarky)	13	EQUAL Autarky: same payoffs Integration: same payoffs Gains from integration: 3	—
	13 ± 2	TRANSFERS Autarky: different payoffs Integration: same payoffs Gains from integration: 3 ± 2	UNEQUAL Autarky: different payoffs Integration: different payoffs Gains from integration: 3

Notes: The plus and minus indicate the points difference for, respectively, the Strong (+) and Weak (−) country. In the experiment 1 point = US$ 0.18. The cooperation payoff k is always equal within a partnership and can be unequal across partnerships. The cooperation payoffs $k + a$ can be unequal within a mixed group and are equally distributed across mixed groups.Gains from integration corresponds to a.

[8]Though in the experiment discounting starts on round $t = 18$, because only at that point the random termination rule starts, one can demonstrates that the incentives to cooperate monotonically decline until round t. It follows that one can simply study the incentives to cooperate in equilibrium using payoffs associated with the beginning of that round t. Those payoffs correspond to v_s above. The details of this demonstration are provided in Bigoni et al. (forthcoming).