Social information 'nudges': an experiment with multiple group references

Shaun HARGREAVES HEAP, Abhijit RAMALINGAM and David ROJO ARJONA

Shaun Hargreaves Heap: Department of Political Economy, King's College London, Strand, London WC2R 2LS, United Kingdom; s.hargreavesheap@kcl.ac.uk.

Abhijit Ramalingam: School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia, Norwich NR4 7TJ, United Kingdom; a.ramalingam@uea.ac.uk

David Rojo Arjona: Department of Economics, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom; dra9@le.ac.uk (Corresponding author)

JEL classifications: C91, D03, D60

The authors thank Pablo Brañas-Garza, David Cooper, Phil Grossman, Tim Salmon, Roberto Weber, two anonymous referees and seminar participants at the University of East Anglia and EWEBE 2014 for helpful advice, comments and suggestions. Hargreaves Heap's work was supported by the Economic and Social Science Research Council through the Network for Integrated Behavioural Science (Grant reference ES/K002201/1). Funding from the School of Economics, University of East Anglia and the Department of Political Economy, King's College London is gratefully acknowledged.

Abstract

Social information 'nudges' concerning how others perform typically boost individual performances in experiments with one group reference point. However, in many natural settings, sometimes due to policy, there are several such group reference points. We address the complications that such multiple group social information might introduce through an experiment. The boost to average performance is significant and comparable to the one group case. Between-group inequality does not change. Individual inequality falls, however, because the boost is largest among the pre-'nudge' very poor performers. Finally, the boost to average performance is highest when individuals freely choose their group affiliations.

1. Introduction

People often respond positively to information about the performance of others in real effort task experiments (e.g. see Falk and Ichino, 2006, Mas and Moretti, 2009, and Azmat and Iriberri, 2010).[1] The provision of such information has, as a result, become a popular candidate public policy 'nudge': e.g., to encourage tax returns and changes in energy consumption (see Sunstein and Thaler, 2008, and Halpern, 2015). The evidence, however, typically comes from experiments where there is one reference group and, in natural settings, there are often several. For example, a particular ethnic group's average income level typically becomes known to members of that group when information is provided about the average income of all ethnic groups. Likewise, a UK school's performance on national tests is known through a league table that contains information on other Schools' performance. We consider whether this difference matters. In particular, we report on a real effort task experiment where there are *two*, exclusive group reference points: one refers to the performance of the subject's own group and the other refers to that of another group.[2] This enables us to consider 3 questions where the existing evidence is negligible or inconclusive.

I. Does the boost to average individual performance when there is a single group reference point generalize to where there are two such reference points?
II. How does the provision of multiple group reference points affect the level of inequality both between individuals and between groups?
III. Does the mechanism for generating individual group affiliation affect the influence of multiple group reference points on performance?

The first question is important because there are reasons for supposing that the influence of social comparison is more complicated in multiple than single group settings. There is not only the difference between an individual and his or her group's behavior ('within-group') that can occur when there is only the own group reference; there is also the difference

between the individual's group and the other group ('between-group'). The dilemma over whether it is better to be a 'big fish in a small pond or a small fish in a big pond' suggests both types of social comparison might motivate individuals. If the existence of several group references complicates social comparison in this and possibly other ways, then it may, in turn, alter how the provision of such group reference points affects behavior.

Our second question is important because policy makers are commonly concerned with how interventions affect both output and the level of inequality in society. There is mixed evidence on this from the single group experiments. In most studies, everyone's performance improves (e.g. Azmat and Iriberri, 2010), but in others there is also convergence of individual behavior on the group reference point (see Abeler *et al,* 2011*)*, in part because the boost in performance is largely among those performing below the norm (see Mas and Moretti, 2009).[3] Both responses are understandable depending on whether the provision of own group information triggers competitive or norm-conforming preferences. Since the presence of two exclusive groups might as easily prime competitive instincts as those of group conformism, it is an open question how the balance between these two motives will be affected.

There is an additional aspect to the question of how inequality is affected by social information. Groups are often hierarchically differentiated by their performance (as in ethnic group incomes and school league tables) and policy makers can be interested in how interventions affect group differences as much as individual ones. Will members of a poor ethnic group, for instance, respond fatalistically to the information that their average income is below that of other ethnic groups? The information may seem to reveal that odds are so stacked against them as members of this group that effort appears relatively pointless (as happens, for instance, when a counter-culture develops and, with respect to decisions over educational efforts, in theories of statistical discrimination and signaling, see Phelps, 1972 and Spence, 1973). Alternatively, the evidence of what others can achieve may spur efforts to

close the gap through some 'underdog' or emulation effect. Which is the case obviously matters for public policy both in relation to the extent of the boost to performance (question I) and the effect on group inequalities (part of question II).

We tackle the first two questions through an experiment where subjects engage in a real effort task (the Gill and Prowse, 2012, slider task) twice, first as *isolated* (as explained below) individuals with no social information (Task 1) and then again (Task 2) as isolated individuals with either no social information (Baseline 1) or an own group reference point for performance (Baseline 2) or an own and another group reference point (Treatment 1). Baseline 1 supplies the control for any changes in performance due to experience. Thus the comparison between the change in performance between Task 1 and Task 2 in Baseline 1 and Baseline 2 gives the effect of a single group reference point and the comparison of the change in Baseline 2 and Treatment 1 gives the effect or two rather than one reference point.

The reference point(s) is (are) information about the performance of people in their (and the other group) at an earlier date. It is *not* generated endogenously by subjects in the experimental session. This is an important part of our experimental design, not only because it gives control over the reference point, but also because this is the way, in practice, that social information is often generated. For example, members of an ethnic group typically receive information through the provision by statistical agencies on average income levels for their and other groups in the past; and this year's Year 12 UK students receive information about what the previous year's Year 12 students did in their GCSEs.[4]

There are two group reference points in Baseline 2 and Treatment 1 (but those in Baseline 2 only know their own and do not know there is another). One group reference point for performance is higher than the other to capture group stratification. In both Baseline 2 and Treatment 1 subjects are randomly allocated to groups. The random process reflects how many groups are formed, particularly natural ones (e.g., ethnic groups). But, it does not

capture how all groups are formed. For this reason in Treatments 2, 3 and 4, subjects select their reference group from the same two options as are in Treatment 1, but under different choice procedures. We discuss these in more detail below. It is through the comparison of the effects of social information across Treatments 1, 2, 3 and 4 that we address question III. The question is important for public policy reasons both in anticipating the effects of policy under different conditions for group formation and when giving policy advice, where this is possible, over how group affiliation should occur. For example, in the case of school league table information, do the effects of this public policy depend on whether the allocation of individuals to schools is random or arises through parental/individual choice? If they do, then an important additional dimension of policy may be the mechanism for allocating individuals to schools.

We discuss the background to these questions and formulate specific hypotheses in the next section. Section 3 explains the experimental design in more detail. Section 4 gives the results. We find that the provision of multiple group reference points boosts almost everyone's performance and that it tends to reduce the variance in performance because the boost is biggest for those, who in the absence of group references, are the very worst performers. There are no significant differences in either of these respects between those in the 'High' performance group and those in the 'Low' one. Thus, the provision of multiple group reference points raises average performance, it has no effect on between-group inequality and reduces within-group inequality. In short, this is a 'nudge' that raises performance without incurring any cost in terms of increasing inequality; indeed it lowers inequality in one dimension. We find that the biggest boost to performance occurs when group affiliation is freely chosen; and the weakest effects on performance and on inequality occur when group affiliation arises from a mechanism of delegated choice. Section 5 discusses the results and we conclude in Section 6.

2. Background and hypotheses

SOCIAL COMPARISON WITH SINGLE AND MULTIPLE GROUP REFERENCE POINTS

There are a variety of explanations for why social comparison *within a group* might motivate individuals in a single group setting. They can be broadly divided into two according to their implication for individual behavior (and hence behavior in the aggregate). First, there are preference-based arguments for social comparison that predict a boost in everyone's efforts. For instance, people may have competitive preferences: they like to do better than others. They may just like winning (see Parco et al., 2005 and Sheremeta, 2010 on this in contests), or they may value the status that comes from visibly being better than others (e.g. see Charness et al., 2013). They may alternatively have a preference for self-esteem and judgments regarding self-worth often depend on a similar kind of relative comparison for epistemic reasons. In each case, in addition to whatever material return comes from performance, people also gain some extra psychological benefits when there is social information and this boosts everyone's performance.[5] Aggregate performance therefore increases, but there is no clear effect on the variance of individual performances within the group. This is the basis for H1.

H1: The provision of a single group reference point boosts individual performances.

Second, there are norm-based arguments for social comparison that predict behavior will converge on the norm. The desire to conform to a norm may arise from 'group-think' type social pressure within the group or because conformity enables a person to identify more strongly with his or her group (e.g. see Akerlof and Kranton, 2000, in the economics literature; and Asch, 1951, for some famous early psychological experiments on such tendencies to conformism). Those who would otherwise do better than the norm, ease back in this case; and those who would lag behind the norm, redouble their efforts to catch up. The

impact of social information on *overall* performance is, thus, unclear. The clear effect is on the variance in performance: it should fall with social comparison on this account. This is the basis for H2.

> H2: The provision of a single group reference point reduces the inequality of individual performances within the group.

As a preliminary to the 3 questions that motivate our experiment, we test H1 and H2 on the effect of single group reference points. While the evidence from single group reference point experiments is generally clear on H1, it is not on H2 (e.g. see Azmat and Iriberri, 2010, Abeler et al., 2011, Falk and Ichino, 2006, and Mas and Moretti, 2009, where knowledge of peers in various forms seems to boost performance).

The introduction of multiple group reference points could complicate the effect of social comparison in a variety of ways. First, it might play into this debate over the 'within-group' effect. For instance, it might be argued that the contrast that is made possible when there are two groups (and there is exclusive membership of one or the other) primes norm guided behavior because any deviation from a group norm may now be associated with membership of another group. This might be expected from Social Identity Theory (see Tajfel and Turner, 1979). Alternatively, if the two groups appear in competition (as they are perhaps more likely to when they are stratified into 'High' and 'Low' groups), then perhaps within-group competitiveness will be primed.

Second, the presence of multiple group references introduces the possibility of another type of social comparison: a 'between-group' comparison effect (see Frank, 1985).[6] Again, there are conflicting intuitions over this. It is possible that individuals may identify more strongly with a 'High' performing group than with a 'Low' performing one because membership of 'High' group conveys status. So, whatever are the effects of social comparison, they are felt more keenly in the 'High' than in the 'Low' performing group. This will be reinforced if

those in the 'Low' group suffer any form of discouragement. Alternatively those in the 'Low' group may be spurred into special efforts because they are 'underdogs' and the gap between the groups may narrow.

With such conflicting arguments and intuitions over questions I and II, our null hypotheses, H3, H4 and H5, are that multiple rather than single group reference points have no effect on performance and inequality.[7]

> H3: The provision of multiple group reference points has the same effect on individual performances as a single own group reference point.
>
> H4: The provision of multiple group reference points has the same effect on the inequality of individual performances as a single own group reference point.
>
> H5: The provision of multiple group reference points has the same effect on the inequality between group performances as a single own group reference point.

GROUP AFFILIATION MECHANISM

Once there is more than one group reference point, the question arises as to how people become attached to one rather than another and whether this influences the effect of social information 'nudges'. This is our third research question.

The mechanism of affiliation might matter because there is some evidence from single group reference points that people are more likely to be influenced when they identify more closely with their group (see Charness *et al*., 2013). People may select a group on the basis of past performance (or more generally other observable characteristics) with the result that those in the same group identify more closely with fellow members because they share more things in common with other group members than when affiliation is random. This is a selection effect. In addition, the procedure itself may, in liberal societies where freedom of choice is highly

valued, encourage greater identification with the group (see Page *et al*, 2005, for experimental support). H6 follows.

> H6: The boost to performance with multiple group reference points is bigger when individuals freely choose their group affiliations than when they are randomly determined.

While this is a powerful intuition in liberal societies, it is not the only one.[8] For example, identity that turns on ethnicity, race or gender is also important in many liberal democratic societies, but is not chosen. There is also the thought that, if people choose their group for the maximum advantage this gives for social comparison, the self-serving nature of the group choice may in turn undermine its psychological power.

The conditions under which people choose their group affiliation also varies. Free choice in H6 is one possibility, but there are others. For example, some groups have monetary membership fees: political parties often charge a small fee and so do many sporting supporters' clubs and, implicitly, religious organizations when membership comes with expectations about charitable giving. This might affect behavior because there is some evidence that the introduction of a cash nexus into otherwise similar decisions generally affects behavior (e.g., see Vohs, 2006, where it appears to promote individual self-reliance) and, more specifically, there is evidence that a small monetary cost actually lowers the value placed on an object (see Shampanier et al., 2007). Thus, the introduction of a monetary fee for group membership might weaken the attachment of the individual to the group and hence the motivating power of the social comparisons based on membership of that group. This is the basis of H7.

H7: The boost to performance with multiple group reference points is smaller when individuals choose their groups and there is a fee as compared with when there is no fee.

Some choices are also made freely but indirectly through a form of delegation. For example, individuals freely choose between job offers, but once a job offer is accepted they accept that someone will manage their time in that job, including whether it is spent with one work group rather than another. There is some evidence that people are averse to other people making decisions for them (see Falk and Kosfeld, 2006) and the introduction of an element of delegation where such control occurs may weaken the attachment to the group. This is the basis of H8.

H8: The boost to performance with multiple group reference points is smaller when individuals delegate their choice of group to another as compared with when these choices are directly and freely made.

3. Experimental design

DESIGN OVERVIEW

There are 2 Baselines and 4 Treatments. In all, individuals in the experiment engage in a real effort task twice (Task 1 and Task 2), always as *isolated* individuals (as explained below). The difference between the Baselines and between them and the Treatments concerns what subjects are told before performing Task 2 about group affiliation and what members of their (and the other group) have done in the past in this real effort task (= the social information).

In Baseline 1, there is no social information. Subjects are not affiliated with groups and they receive no information about past performance. In Baseline 2, there is social information about own group performance in the past. Subjects are told in Task 2 either that they have been affiliated with a group where members on average achieved 171 in the past on the real

effort task or they have been affiliated with a group where members have on average achieved 108 in the past on the real effort task.[9] The purpose of the two Baselines is, 1) to provide a preliminary check that we obtain the 'usual' boost to performance through the provision of information about own group performance with the specific real effort task and isolation set-up (i.e., H1) and 2) to address the question concerning inequality (i.e. H2). Hereafter, the key Baseline is Baseline 2 because we wish to compare the boost from own group social information with that which occurs when there are multiple group reference points in Treatments 1, 2, 3 and 4.

In Treatment 1 (Random), subjects are told in Task 2 that there are two groups, they are told the average performance in the past for both groups (171 = 'High' group and 108 ='Low' group) and that they have been randomly affiliated with one of the groups. We use the terms 'High' and 'Low' in this and the remaining Treatments and they always refer to the same group reference points for performance (171 and 108 respectively) that were used in Baseline 2. For convenience in the presentation of the results, we shall always refer to the 'High' and 'Low' groups, even though in Baseline 2, subjects only knew their own group performance level (and not the title 'High' or 'Low' that was associated with these performance levels in the subsequent Treatments).

It is through the comparison of the boost in performance between Task 1 and Task 2 in Baseline 2 and Treatment 1 that we address H3, H4 and H5 concerning the effects of multiple rather than single group reference points. This comparison is cleanest for this purpose because they share the same reference points and allocation procedure to groups and the only difference is the presence of a single or multiple reference points.

Treatment 2 (Free Choice) is the same as Treatment 1, except subjects choose which group to affiliate with. The comparison of the boost in Treatment 2 with that in Treatment 1 addresses H6.

Treatment 3 (Paid Choice) is the same as Treatment 2, except that there is a small cost of affiliating with both the 'High' and the 'Low' group. The cost is the same (to avoid both this being a source of difference between groups and the introduction of more than one change between treatments) and we test H7 by comparing the boost in performance through the social information in Treatments 2 and 3.

Treatment 4 (Delegation) is the same as Treatment 1 except that subjects choose someone who will select a group on their behalf. They nominate one of four previous participants, identified only by a number, to choose a group for them. They were informed that each of these four participants had taken part earlier and had made a choice of group in the past; and by nominating that person, his/her choice would be implemented for the subject.[10] The comparison between the boost in Treatment 4 and Treatment 1 tests H8.

Treatments 2 and 3 play a further role in the experimental design. It is possible that subjects respond to the information on group performance through higher performance without identifying with their own group. This is possible because they could integrate the information on the two performance levels into a single social information reference point (which they then respond to). In other words, they need not identify with their own group and yet they could still respond to the social information 'nudge'. However, if this was the case, then there would be no reason for subjects to select one group rather than another in Treatments 2 and 3. Their choices should as a result be random. On the other hand, if subjects do identify with their own group and make within-own-group comparisons and between-group comparisons, then the choice of group will matter. They will not be random and, under some conditions, they will be associative (see Falk and Knell, 2004). This is what we find (see electronic appendix Table B1). More generally, if own group identification did not vary across Treatments (because it was absent in the experiment all together), then there would be

no reason to expect that the boost to performance would vary across the Treatments because each would then have what is the same provision of two group reference points.

The group reference points are 'minimal' in the sense that the affiliation to groups occurs in the experiment. This was another design choice. We made it for two reasons. First, using pre-existing or natural groups complicates the interpretation of the results because they may depend on the specificities of the group. Second, minimal group affiliation enables us to study the possible influence of different mechanisms for group affiliation. While the minimal group paradigm is well-established in the experimental literature (see Chen and Li, 2009, and Hargreaves Heap and Zizzo, 2009) and has these advantages in this context, it has the possible disadvantage that the affiliations are very weak. This is, in part, why it is important to have the cross check for this possibility in Treatments 2 and 3 where there is group choice.

Our two group reference points are mutually exclusive. This was again a design choice to reflect some important types of multiple group references. It echoes the way, for example, that one is either an immigrant or one is not; one is Italian or one is not; one is in the top or the bottom set for maths, one cannot be in both.

EXPERIMENT DETAILS

A key feature of our design is the control for group reference points. We inform subjects about the performance level of the group to which they are affiliated in Baseline 2 and of the two groups in Treatments 1-4 and we want subjects to take these group reference points as given. Towards this end, we ran several sessions, each of which involved 28 to 34 subjects making individual choices in isolation. To ensure that subjects did not perceive any group affiliation with those in the laboratory, we recruited subjects to arrive at the experimental lab in 10-minute intervals. Once subjects arrived, they were individually seated at a computer and they could begin the experiment immediately; and when a subject finished the experiment, he/she could leave immediately.

Once seated, a subject read instructions at his/her computer. Subjects were informed that they would perform a set of independent tasks and that instructions for each task would be presented just prior to that task. The first task was the slider task of Gill and Prowse (2012) where the subject is presented a screen with 48 sliders that can take values from 0 to 100. The task is to position each slider at exactly 50 using the computer mouse. Subjects could work on the first task for a total of 16 minutes, with a new screen of 48 sliders being presented every two minutes. At any point, the subject could also decide to stop working on this task and to move on to the next task by clicking a button at the bottom of the screen. At all times, subjects could see the total number of sliders they had positioned correctly. Subjects accumulated earnings for the total number of sliders correctly positioned at 50 (before moving on to Task 2) at a piece rate of 5 points per slider. To ensure understanding, subjects answered control questions and participated in a non-incentivized two-minute practice round before beginning Task 1.

The second task was the same as Task 1: that is, 16 minutes of the slider task at the same piece rate. However, before Task 2, subjects were given information on the performance of people at an earlier date (except in Baseline 1). In Baseline 2 they were either told that members of their affiliated group had in the past either performed at a mean number of 171 or 108 sliders positioned at 50 in 16 minutes. In the Treatments 1-4, subjects were told that they would be affiliated with either a High performance or Low performance group, where again the mean number of sliders positioned at 50 in 16 minutes in these groups were 171 and 108 respectively. The mechanism for affiliation was then described and varied across Treatments 1-4 in the manner set out above (also see below). Thereafter Task 2 is identical to Task 1 with the exception that reference point information was displayed on the top of the screen throughout. So, group affiliation had no payoff implications. Before proceeding, subjects had to answer control questions, including questions that tested their understanding that group affiliation had *no* pay-off implications.

The only detail to add about the mechanisms for group affiliation is that in Paid Choice, subjects had to pay a fee of 100 points for the 'right' to choose a group.

Having completed the real effort task twice, subjects undertook three further tasks designed to complement the normal demographic information that is used as controls in individual level regressions. The third task was a series of incentivized lottery choices intended to measure subjects' risk aversion in the gain and loss domains, ambiguity aversion and conformism (Holt and Laury 2002). The fourth task was not incentivized and subjects received a fixed payment of 100 points for answering a hypothetical question intended to capture his/her degree of self-serving bias (see Babcock and Loewenstein, 1997). The fifth task was the Arad and Rubinstein (2012) incentivized 11-20 level-k task.[11]

All sessions were conducted at the University of East Anglia. The experiment was completely computerized and was programmed in z-Tree (Fischbacher 2007). A total of 244 subjects participated in our four main treatments (61 in each), 25 subjects in Baseline 1 and 58 subjects participated in Baseline 2. Experimental earnings were calculated in points and were converted to Pounds at the rate of 200 points to £1. Subjects were able to collect their earnings at their convenience after the end of the day's session. The average subject spent about 60 minutes in the lab and earned £13.45 including a £2 show-up fee.

4. Results

SINGLE REFERENCE POINT

Table 1 gives the mean performance on Task 1, Task 2 and the difference between them in the two Baselines and the four Treatments. The first two rows address H1 and H2. Wilcoxon Signrank tests show that the boost to performance is always significantly different from 0 in both Baseline 1 and Baseline 2 (and all the Treatments) and the boost is significantly higher in Baseline 2 than Baseline 1 (Wilcoxon ranksum $p = 0.0389$). Table 1 also provides the total number of subjects who drop out in each task and treatment. Every task and treatment has a

frequency of around 10% or lower with the exception of Task 1 in the Random treatment. These reduced frequencies are consistent with an unsystematic level of noise.

[TABLE 1]

Table 2 reports individual level panel random effects regressions on performance, where we control for possible learning/boredom and re-start effects through Task number dummies and screen number dummies. In addition, we control for subjects' demographic characteristics (age, gender and dummy for international student) and their responses in Tasks 3 through 5.[12] The first column examines performance in Baselines 1 and 2. None of the screen time trend dummies are significant. The Task 2 dummy is positive and the interaction between Task 2 and Baseline 2 is also weakly significant (p-value = 0.062), suggesting, albeit less convincingly than in the aggregate data, that the boost to performance in Task 2 is larger in Baseline 2.[13]

[TABLE 2]

> **Result 1 (in favor of H1).** *Relative to when no reference point is provided, there is clear evidence in the aggregate data and weaker but similar evidence in the individual data that the provision of a single reference point leads to a boost in performance.*

There is no apparent change in the variance of output performance in either Baseline between Task 1 and Task 2. The absence of an effect from the single reference point on inequality is reinforced when we consider whether those who initially performed below their reference point in Task 1 boost their performance by more in Task 2 than those who were performing initially above this reference norm. Table 3 disaggregates in this way for Baseline 2 between those who received the single High or Low reference point norm. The boost to performance is always significantly positive but Wilcoxon ranksum tests show that it is not significantly

higher for those who initially scored below the reference norm than those who were initially above the norm for the High norm (p=0.4350) and it is only marginally significantly higher for those initially performing below the Low norm (p=0.0743).

[TABLE 3]

Result 2 (against H2): *There is no evidence that the provision of a single reference point increases individual inequality in performance.*

MULTIPLE REFERENCE POINTS

Table 1 shows that the boost to performance in Task 2 is always significantly greater than zero for Treatments 1-4. Furthermore, the boost is always significantly greater than that found in Baseline 1, where there are no reference points (Wilcoxon ranksum test for Random (p = 0.0114), Free Choice (p = 0.0014), Paid Choice (p = 0.0752) and marginally for Delegation (p = 0.0952). These Treatment differences are also found in individual regressions (see electronic appendix Table B2). The size of the boost, however, appears similar to that when there is one reference point in Baseline 2 except in Free Choice where it is higher (35.41 cf. 24.14) and Delegation where it is lower (18.77 cf. 24.14).

We turn now to H3. In Table 1, the average boost for those in the Baseline 2 (24.14) is not significantly different from that in the Random Treatment (25.21): Wilcoxon test p-value = 0.5199. The right column of Table 2 reports the individual regression on the combined behavior in the Random treatment and Baseline 2. There is a Baseline × Task 2 interact dummy to distinguish whether behavior is different in Task 2 in the Baseline 2 as compared with Random. The coefficient on this is not significant, again suggesting no difference in the boost.

Result 3 (in favor of H3): *There is no evidence of a difference in the boost to average performance in Task 2 between Baseline 2 (single group reference) and the Random treatment (two group reference points with random allocation).*

The variance of performance on Task 2 in Random does fall (as compared with Task 1), unlike that in Baseline 2 (see Table 1). We disaggregate along the same lines as Table 3 between those who performed above or below their reference norm for the Random treatment in Table 4. The boost is now strongly significant and larger for those who initially performed below the reference norm in the Low group than those who initially performed above the norm (p=0.0003). Again there is no difference in boost between initial above and below the norm for those in the High reference group.

[TABLE 4]

Result 4 (against H4): *There is evidence that the provision of multiple group reference points reduces individual inequality as compared with the provision of a single reference point.*

This result is reinforced when we look at all the two reference point Treatments (1-4) in this way in Table 5. The boost is significantly higher for those initially below the norm (than those initially above) in the Low reference point groups (p=0.0060) but not in the High reference point groups (p=0.9762).

[TABLE 5]

As the boosts to performance appear to be strongest among those who initially performed below the reference norm in the Low group in the 2 reference group Treatments, we test for this formally with individual regressions in Table 6. We divide performance in Task 1 into four interval groups and introduce these as dummies interacted with task 2 into the individual

performance regression: Upper = 171 and above; Upper Middle = (171, 139]; Lower Middle = (138, 108]; and Lower = 107 and below.[14] The omitted dummy in this equation is Lower and so the coefficients on the other interval dummies capture the difference with respect to those in these intervals and the Lower interval group. We find that there is no difference among the coefficients on the Upper, Upper Middle and Lower Middle dummies interacted with Task 2 ($p > 0.10$ for all pairwise post-regression tests) and they are significant and negative in all Treatments except Delegation. This suggests that the boost in Task 2 for individuals in these intervals is significantly less than that of individuals in the Lower interval on Task 1 in all Treatments except Delegation.

[Table 6]

> **Result 5 (Robustness check of result against H4):** *The very lowest performers (i.e. those below 108) in Task 1 increase their performance on Task 2 by significantly more than other individuals when there are 2 group reference points, except in the Delegation Treatment.*

Result 5 is potentially consistent with both the preference based and convergence on norm bases explanations of the influence of social information. That there is no evidence, however, to support the inference that high initial performers reduce their performance towards the norm in Tables 5 and 6 counts against the norm convergence account.

> **Result 6 (against the norm convergence account):** *There is no evidence that those whose performance in Task 1 is above the norm in their group conform to this norm by reducing their performance in Task 2.*

Turning to the inter-group dimension of inequality, a Wilcoxon test shows no significant difference in the boost between High and Low groups in Baseline 2 ($p = 0.3386$) or in all treatments with multiple reference points combined ($p = 0.3956$). This result is also robust

across Treatments (Random: p = 0.6431; Free Choice: p = 0.3638; Paid Choice: p = 0.8356; Delegation: p = 0.4735). Individual task regressions yield the same conclusion: the group affiliation dummy is never significant (see electronic appendix Table B3).

> **Result 7 (in favor of H5):** *There is no evidence that the boost to performance in Task 2 is different for those affiliated to the High group as compared with those in the Low group.*

We turn now to the possible influence of the mechanism of group affiliation that is revealed by the comparison across the 2 reference points Treatments. The Wilcoxon test for aggregate differences suggests it is higher under Free Choice than Paid Choice and Delegation at 10% level (p = 0.072 and p = 0.090, respectively). We also test for differences through Treatment dummies interacted with Task 2 in the individual performance regressions in Table 7. The omitted Treatment dummy is for Random. Focusing on the All column, the coefficient on Free Choice interacted with Task 2 is positive and significantly different at the 10% level from the omitted Treatment dummy (Random) and it is higher than the other Treatment dummy interacts (which are not significantly different from zero). The difference between Free Choice × task 2 dummy and the other choice dummies × task 2 is significant in both cases (p = 0.0494 for Paid Choice and p = 0.0105 for Delegation).

[TABLE 7]

> **Result 8 (weakly in favor of H6 and more strongly in favor of H7 and H8):** *The boost to performance on Task 2 is higher under Free Choice than in the other Treatments, and most clearly when compared with Paid Choice and Delegation.*

4. Discussion

These results, taken at face value, are important and new.

They are new because, to our knowledge, multiple group social information 'nudges' have not been examined experimentally. The evidence on social information 'nudges' comes from single group reference points. This is a weakness because in many natural settings there are multiple groups. Although this means there are no direct cross checks with the other experiments in the literature, there are points of comparison. For instance, the boost to performance in our experiment with a single reference group (17.2%) is similar to that found in Azmat and Iriberri (2016) when there is a single reference group and a piece rate payment system (=17.3%) and Falk and Ichino (2006) when there is a single group and a flat payment (=16.3%). Neither of these comparator experiments allowed a statistical test of whether this performance boost was related to performance prior to the provision of a group reference point because both rely on between subject comparisons. Nevertheless, there is some suggestion in Falk and Ichino (2006) that, in the same way as we find, the biggest boosts occurred among low performers. This is consistent with what also appears to be the case in the Mas and Moretti (2009) experiment with respect to poor initial performers and it is a stronger result with respect to the good performers (= above average in their study). It is also consistent with the findings of some peer group studies (see Carrell et al*.,* 2009).

The results are important, not least for their policy implications. First, a social information policy 'nudge' appears generally attractive. It boosts individual performance similarly in multiple group settings as when there is a single reference group; and there is no adverse effect on inequality.[15] Indeed, individual inequality falls because the biggest boosts are among the low individual performers.

Second, notwithstanding this general encouragement, a policy maker has to be aware of how group reference points are generated because the effects are not the same. They are weakest when group affiliation occurs through a mechanism of 'delegated choice' and to a lesser extent 'paid choice'.[16] The former is consistent with the evidence from other experiments on

control aversion (e.g., see Falk and Kosfeld, 2006). The latter counts against the common intuition that people value something they pay for, but it is consistent with experimental evidence that money priming encourages individual self-sufficiency and so could weaken attachments to a group.

Third, the policy of promoting consumer choice in the provision of public services needs careful assessment. This policy is typically designed to encourage efficiency through greater competition and is supported by the provision of information on public service performance (as, for example, in school and hospital league tables in the UK). The evidence from this experiment suggests caution in associating any gains with competition *per se* because much of its effects may be coming from what is, in effect, the social information 'nudge' aspect of the policy.

In addition, the results are important because they give some insights into the mechanisms responsible for these social information effects. There is, for example, no evidence to support convergence to a group norm or the development of a counter-cultural norm of low performance in the Low group. A preference based explanation of the response to social information is better supported by the data (see Results 2 and 6). There is also evidence, consistent with other findings in the experimental literature, that free choice encourages identification with a group while small monetary fees and delegating decisions over membership detract from this identification.

5. Conclusion

There is evidence that the provision of social information, a group performance reference point, affects individual performance positively and so can act as policy 'nudge'. What is not known from the existing literature is whether the effects on individual behavior are similar when there are multiple, exclusive groups. This is an important gap because many group reference points naturally arise in a context where there is more than one group and whether

there are one or many group reference points is itself often a policy choice. It is also not clear from the existing experimental literature whether the provision of group reference points affects not only average performance but also the inequality of performance across the population. This is important because policy makers often care about both average performance and its individual variability across the population. Our experiment addresses both issues. In addition, it examines whether the mechanism generating group affiliation, itself often a policy choice, affects the influence of such social comparisons.

We find that a social information 'nudge' with multiple groups boosts average performance by a similar amount to a single group 'nudges'. We also find individual inequality usually falls and group inequality is unchanged. In particular, individuals who initially perform among the worst produce the biggest boost in performance after the provision of the group reference points. The exception to this conclusion is that the boost among the very weakest performers does not occur when group affiliation is generated through a form of Delegated choice. There is also some evidence that the boost to average performance is weakest in these circumstances too. In contrast, we find that the average boost is generally largest when group affiliation is freely chosen.

In short, a social information 'nudge', judged in these respects, appears rather attractive but its attraction does depend on how group affiliation is generated.

References

Abeler, J., A. Falk, L. Gotte and D. Huffman. 2011. Reference Points and Effort Provision. *American Economic Review* 101(2), 470-492.

Ahn, T.K., R. M. Isaac and T. C. Salmon. 2008. Endogenous Group Formation. *Journal of Public Economic Theory* 10, 171-194.

Ahn, T.K., R. M. Isaac and T. C. Salmon. 2009. Coming and Going: Experiments on Endogenous Group Sizes for Excludable Public Goods. *Journal of Public Economic* 93, 336-351.

Akerlof, G. A. and R. E. Kranton. 2000. Economics and Identity. *Quarterly Journal of Economics* 115(3), 715-53.

Arad, A. and A. Rubinstein. 2012. The 11-20 Money Request Game: A Level-k Reasoning Study. *American Economic Review* 102 (7), 3561-3573.

Araujo, F.A, E. Carbone, L. Conell-Price, M.W. Dunietz, A. Jaroszewicz, D. Lame, R. Landsman, L. Vesterlund, S. Wang and A.J. Wilson. 2015. The Effect of Incentives on Real Effort: Evidence from the Slider Task. *CESifo Working Paper No. 5372*

Asch, S.E. (1951). Effects of Group Pressure on the Modification and Distortion of Judgments. in H. Guetzkow (Ed.), *Groups, Leadership and Men*. Pittsburgh: Carnegie Press, 177–190.

Azmat, G. and N. Iriberri. 2010. The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students. *Journal of Public Economics* 94 (7-8), 435–452.

Azmat, G. and N. Iriberri. 2016. The Provision of Relative Performance Feedback Information: An Analysis of Performance and Satisfaction. *Journal of Economics and Management Strategy* 25(1), 77-110.

Babcock, L. and G. Loewenstein. 1997. Explaining Bargaining Impasses: The Role of Self-Serving Biases. *Journal of Economic Perspectives* 11(1), 109-126.

Ball, S., C. Eckel, P. Grossman and W. Zame. 2001. Status in Markets. *Quarterly Journal of Economics* 116(1), 161-188.

Bellemare, C., P. Lepage and B. Shearer. 2010. Peer Pressure, Incentives, and Gender: An Experimental Analysis of Motivation in the Workplace. *Labour Economics* 17 (1), 276-283.

Benndorf, V., H. Rau and C. Solch. 2014. Minimizing Learning Behavior in Experiments with Repeated Real-Effort Tasks. *Working paper*.

Beugnot, J., B. Fortin, G. Lacroix and M.C. Villeval. 2013. Social Networks and Peer Effects at Work. *IZA Discussion Papers* 7521.

Carrell, S., R. Fullerton and J. West. 2009. Does Your Cohort Matter? Estimating Peer Effects in College Achievement. *Journal of Labor Economics* 27(3), 439–464.

Charness, G., D. Masclet and M.C. Villeval. 2013. The Dark Side of Competition for Status. *Management Science* 60(1), 38-55.

Charness, G., R. Cobo-Reyes, J. Lacomba, F. Lagos and J.M. Perez. 2016. Social Comparisons in Wage Delegation: Experimental Evidence. *Experimental Economics* 19 (2), 433-459.

Chen, Y. and S. X. Li. 2009. Group Identity and Social Preferences. *American Economic Review* 99(1): 431-57.

Eriksson, T., A. Poulsen and M.C. Villeval. 2009. Feedback and Incentives: Experimental Evidence. *Labour Economics* 16(6), 679-688.

Falk, A. and A. Ichino. 2006. Clean Evidence on Peer Pressure. *Journal of Labor Economics* 24(1), 39-58.

Falk, A. and M. Knell. 2004. Choosing the Joneses: Endogenous Goals and Reference Standards. *Scandinavian Journal of Economics* 106(3), 417-435.

Falk, A., and M. Kosfeld. 2006. The Hidden Costs of Control. *American Economic Review* 96(5), 1611-1630.

Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10(2), 171-178.

Frank, R. 1985. *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.

Georganas, S., M. Tonin and M. Vlassopoulos. 2015 Peer Pressure and Productivity: The Role of Observing and Being Observed. *Journal of Economic Behavior and Organization* 117, 223-232.

Gill, D. and V. Prowse. 2012. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review* 102(1), 469-503.

Grossman, P. and M. Komai. 2013. The Economic Impact of Envy: Evidence from a Multi-Period Game with Attacks and Insurance. *Monash Economics Working Paper* 01-13.

Halpern, D. 2015. *Inside the Nudge Unit*. London: W.H.Allen.

Hargreaves Heap, S. and D. J. Zizzo. 2009. The Value of Groups. *American Economic Review* 99(1): 295-323.

Holt, C. and S. Laury. 2002. Risk Aversion and Incentive Effects. *American Economic Review* 92(5), 1644-55.

Mas, A. and E. Moretti. 2009. Peers at Work. *American Economic Review* 99(1), 112-45.

Page, T., L. Putterman and B. Unel. 2005. Voluntary Association in Public Good Experiments: Reciprocity, Mimicry and Efficiency. *Economic Journal* 115(506), 1032-1053.

Parco, J., A. Rapoport and W. Amaldoss. 2005. Two-Stage Contests with Budget Constraints: An Experimental Study. *Journal of Mathematical Psychology* 49(4), 320-338.

Phelps, E. 1972. The Statistical Theory of Racism and Sexism. *American Economic Review* 62, 659-661.

Shampanier, K., Mazar, N. and D. Ariely. 2007. Zero as a Special Price: The True Value of Free Products. *Marketing Science* 26 (6), 742 – 757.

Sheremeta, R. 2010. Experimental Comparison of Multi-stage and One-stage Contests. *Games and Economic Behavior* 68 (2), 731-747.

Spence, M. 1973. Job Market Signalling. *Quarterly Journal of Economics* 87 (3), 355-74.

Sunstein, C. and R. Thaler. 2008. *Nudge: Improving decisions about health, wealth and happiness*. Yale University Press.

Tajfel, H., and J. C. Turner. 1979. An integrative theory of intergroup conflict. in *The social psychology of intergroup relations*, ed. Stephen Worchel and William Austin, 33-47.

Vohs, K.D., N.L. Mead and M.R. Goode. 2006. The Psychological Consequences of Money. *Science* 314(5802), 1154-1156.

[1] There are also experiments where other types of social information (e.g., over other people's wages and rights) also affect performance (e.g., see Charness *et al,* 2016) and there are experiments where peer effects are small or non-existent (see Georganas *et al*, 2015, Bellemare *et al,* 2010, and Beugnot *et al,* 2013).

[2] There are many experiments with multiple groups that have focused on how group membership affects strategic behavior in interactions between members of different groups (see Ball *et al,* 2001). But to our knowledge, there seem to be virtually none that are concerned with how information about other groups' performance (as well as one's own) motivates behavior in what is in other respects a non-strategic setting. Grossman and Komai (2013) provide an unusual recent example where there are two such groups, but they study a slightly different question with respect to envy. There are a rich and a poor group, each has its own hierarchy of earnings and taken together they establish the overall (society) hierarchy of earnings. They are concerned with whether subjects envy those who are better off within their group and/or within the society as a whole. They find evidence of both but within group envy is the most powerful.

[3] And in others there is no effect at all (see Eriksson *et al*, 2009). A further reason for tentativeness in summarizing the literature is that studies like Falk and Ichino, 2006, and Azmat and Iriberri, 2016, are between subject designs and do not readily allow a test for whether some subject groups are more inclined to respond to group reference points than others. Our within subject design facilitates such tests.

[4] Stereotypes have a similar comparative character and temporal structure in the sense that they are based, albeit more or less loosely, if at all, on previous behaviors and they inherently relate a group's behavior to that of other groups. They, thus, potentially function as a form of informal social information 'nudge'.

[5] Azmat and Iriberri (2010) offer a different information-based account of the influence that relative information has on behavior by applying what they refer to as 'self-perception theory' in relation to individual ability. On this account, information about average performance boosts (lowers) performance of those above (below) the reference point and so stretches the distribution of performances. They find no evidence of this and neither do we.

[6] Also see Falk and Knell (2004) who develop a model where people select their reference point and argue that there are two similar considerations in such a choice: self enhancement (which comes in these terms from being a big fish in small pond) and self-improvement (which comes from being in a big pond).

[7] It is perhaps worth emphasizing that although we are interested in such disputes over the psychology of social comparison, we are not primarily interested in deciding between them here. Our primary concern is with how multiple group reference points affect behavior. It is enough for us that there are such debates so as to render this a plausible (and interesting) open theoretical question that can therefore be usefully examined empirically.

[8] The impact of different rules of how groups can be formed endogenously – affecting two groups (insiders and outsiders) has also been analyzed in the context of provision behavior in public good games (Ahn et al., 2008; 2009).

[9] Prior to running our main experiment, we ran two sessions where 62 subjects performed the slider task for up to 30 minutes. The mean number of sliders positioned correctly in the last 16 minutes by the top and bottom halves were, respectively, 171 and 108.

[10] Our aim was to generate an equal probability of being assigned to the high as the low group under 'Delegation' so that it was comparable in this respect to 'Random'. We asked randomly chosen participants from our preliminary sessions (where, recall, subjects performed only the task with no group affiliation for up to 30 minutes) a non-incentivized hypothetical question after the experiment concerning whether they would choose to affiliate someone else with a High or Low performance group. After 4 such decisions, we achieved a

set such that there was 50-50 probability of being High/Low and so these 4 became the set of delegated choosers in the 'Delegation' treatment.

[11] This was the only task for which earnings depended on the decisions of another subject. At the end of the day's session, subjects were anonymously and randomly paired with another subject from the same day. The two subjects' earnings for this task were determined based on their choices in the 11-20 task.

[12] For brevity, we do not present coefficient estimates of these control variables. These variables did not have a significant impact on performance.

[13] Unlike Gill and Prowse, 2012, Benndorf et al, 2014, Araujo et al., 2015, and Georganas et al., 2015, we do not find a time trend in the screen data, but the re-start effect is consistent with a general kind of learning.

[14] We chose 139 as it is the mid-point between 108 and 171.

[15] The boost from the social information nudge in this context is particularly interesting because this is a real effort task where there is evidence that material incentives have little effect on performance (see Araujo *et al*, 2015).

[16] This difference (and the others across Treatments reported in Result 8) are also important because they would not be expected (just as associative choice in Treatments 2 and 3 would not be expected) if our subjects were responding to the two group references as a single information nudge (i.e. they attached no special weight to their own group reference point).

TABLES

Table 1. Mean individual performance

	N	Task 1	Drop-outs Task 1	Task 2	Drop-outs Task 2	Difference
Baseline 1 - No ref	25	144.96 (47.25)	3	158.32 (47.74)	4	13.36*** (12.57)
Baseline 2 – One ref	58	140.19 (44.72)	4	164.33 (44.92)	5	24.14*** (40.69)
Random affiliation	61	137.34 (41.54)	12	162.56 (39.95)	6	25.21*** (29.24)
Free choice	61	121.05 (47.93)	6	156.46 (33.39)	2	35.41*** (37.88)
Paid choice	61	129.36 (51.25)	8	152.57 (49.47)	5	23.21*** (27.42)
Delegation	61	125.92 (38.02)	9	144.69 (41.36)	8	18.77*** (33.02)

Standard deviations in parentheses. The *** indicate that, in all treatments, the boost in performance in Task 2 is statistically significant at the 1% level using a Wilcoxon signed rank test.

Table 2. Individual performance – the effect of reference point(s)

	Baseline 1 (no ref) vs Baseline 2 (single ref)	Random (multiple ref) vs Baseline 2 (single ref)
Task 2 dummy	2.733***	2.953***
	(0.609)	(0.559)
Screen # in Task 1	0.149	-0.101
	(0.098)	(0.110)
Screen # in Task 2	-0.106	-0.051
	(0.104)	(0.078)
Baseline 2 dummy	-0.457	1.515
	(1.994)	(1.196)
Baseline 2 × Task 2	1.402*	-0.191
	(0.750)	(0.826)
Constant	21.14***	25.45***
	(5.201)	(3.795)
Controls?	Yes	Yes
Observations	1328	1904

Dep. variable: Individual Performance on each screen. Standard errors clustered on subject in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Only Baseline 2 – two separate reference points

	All		Below the norm after Task 1			Above the norm after Task 1		
	Mean Task 1	**Mean Task 2**	**Obs**	**Mean Boost**	**z [p]**	**Obs**	**Mean Boost**	**z [p]**
Baseline 2 High ref	153.66 (38.13)	175.45 (40.61)	19	22.42 (49.08)	2.536 [0.0112]	10	20.6 (10.54)	2.807 [0.0050]
Baseline 2 Low ref	126.72 (47.35)	153.21 (46.92)	7	57.43 (72.48)	1.690 [0.0910]	22	16.64 (21.05)	3.280 [0.0010]

Column All includes aggregate mean performance in Task 1 and Task 2. The other columns present the mean of the boost in individual performance (Task 2 – Task 1). Figures in parentheses are standard deviations. The last column presents test-statistics [p-values] for Wilcoxon sign rank tests for Boost = 0.

Table 4: Only Random – two separate reference points

	All		Below the norm after Task 1			Above the norm after Task 1		
	Mean Task 1	**Mean Task 2**	**Obs**	**Mean Boost**	**z [p]**	**Obs**	**Mean Boost**	**z [p]**
Random High ref	145.18 (35.99)	165.54 (42.01)	24	20.25 (25.90)	3.230 [0.0012]	4	21 (18.31)	1.826 [0.0679]
Random Low ref	130.69 (45.19)	160.03 (38.58)	10	60.5 (42.26)	2.701 [0.0069]	23	15.78 (12.87)	3.712 [0.0002]

Column All includes aggregate mean performance in Task 1 and Task 2. The other columns present the mean of the boost to individual performance (Task 2 – Task 1). Figures in parentheses are standard deviations. The last column presents test-statistics [p-values] for Wilcoxon sign rank tests for Boost = 0.

Table 5. Means and Tests for zero performance boost by initial performance relative to norm in Treatments 1-4

	All		Below the norm after Task 1			Above the norm after Task 1		
	Mean Task 1	**Mean Task 2**	**Obs**	**Mean Boost**	**z [p]**	**Obs**	**Mean Boost**	**z [p]**
All groups	128.42 (45.10)	154.07 (41.69)	138	29.66 (41.25)	8.328 [0.0000]	106	20.43 (15.61)	8.358 [0.0000]
High groups	132.98 (46.66)	156.76 (42.69)	104	24.74 (38.35)	7.276 [0.0000]	25	19.76 (12.81)	4.253 [0.0000]
Low groups	123.29 (42.91)	151.05 (40.51)	34	44.71 (46.52)	4.335 [0.0000]	81	20.64 (16.45)	7.247 [0.0000]

Column All includes aggregate mean performance in Task 1 and Task 2. The other columns present the mean of the difference in individual performance (Task 2 – Task 1). Figures in parentheses are standard deviations. The last column presents test-statistics [p-values] for Wilcoxon sign rank tests for Difference = 0.

Table 6. Individual performance by interval groups on Task 1 (Random effects model)

	(1) All	(2) Random	(3) Free choice	(4) Paid choice	(5) Delegation
Task 2	49.68***	49.07***	63.05***	43.75***	38.08***
	(6.230)	(12.84)	(13.46)	(11.11)	(13.84)
Upper	117.4***	112.1***	118.1***	141.7***	109.2***
	(5.173)	(7.445)	(10.80)	(11.80)	(12.00)
Upper-middle	85.68***	76.01***	84.93***	99.33***	87.39***
	(4.740)	(6.481)	(9.618)	(10.32)	(12.03)
Lower-middle	57.71***	57.56***	57.45***	66.13***	56.14***
	(4.727)	(6.552)	(9.340)	(10.99)	(11.48)
Upper × Task 2	-30.68***	-35.37**	-42.30***	-20.25*	-22.83
	(6.726)	(14.04)	(13.73)	(11.96)	(16.22)
Upper-middle × Task 2	-32.94***	-29.07**	-43.41***	-30.07***	-24.74
	(6.512)	(13.05)	(13.97)	(11.40)	(15.30)
Lower-middle × Task 2	-32.17***	-31.07**	-37.00***	-31.32**	-24.65*
	(6.903)	(14.78)	(14.10)	(13.63)	(14.97)
Constant	82.84***	134.8***	59.14**	57.70	27.81
	(16.45)	(30.82)	(23.33)	(35.76)	(37.83)
Controls?	Yes	Yes	Yes	Yes	Yes
Observations	488	122	122	122	122

Dep. variable: Individual Performance on each task, leading to 2 observations per individual. Standard errors clustered on subject in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7. Individual performance with Treatment dummies

	(1) All Groups	(2) High Groups	(3) Low Groups
Task 2	25.21***	20.36***	29.33***
	(3.784)	(4.747)	(5.789)
Free Choice	-13.46	-0.233	-19.85*
	(7.960)	(11.81)	(10.23)
Paid Choice	-3.214	-4.521	-5.094
	(8.269)	(9.276)	(13.75)
Delegation	-7.819	-17.41*	1.915
	(7.216)	(9.352)	(10.07)
Free Choice × Task 2	10.20*	13.31	7.461
	(6.192)	(9.085)	(8.715)
Paid Choice × Task 2	-2.000	3.399	-7.233
	(5.367)	(7.143)	(7.312)
Delegation × Task 2	-6.443	-1.751	-10.37
	(5.707)	(8.501)	(7.367)
Constant	169.4***	192.7***	129.6***
	(26.95)	(38.65)	(37.98)
Controls?	Yes	Yes	Yes
Observations	488	258	230

Dep. variable: Individual Performance on each task. Standard errors clustered on subject in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. After regression (2), no treatment differences. After regression (3), Free vs. Paid, p = 0.0628; Free vs. Delegation, p = 0.0249.

Electronic Appendix 1: Experimental Instructions

Thanks for taking part in this experiment.

You are the only person booked to start at this time.

You will be paid £2 for turning up plus whatever you earn in the experiment. The experiment requires you to perform 5 independent tasks. Further instructions will be given before each task. You can earn points in each task. You will be told how many points you obtained in each task at the end of the experiment. These points will then be converted into cash at a rate of £1 per 200 points. More instructions about how to collect your money will be provided at the end of the experiment.

To begin task 1, press OK.

TASK 1

You will see a series of screens. Every 2 minutes a new screen will appear. In each screen, you will see 48 sliders. Each slider has values from 0 to 100. Each slider will appear at position 0. The slider can be adjusted and readjusted an unlimited number of times and the current position is displayed to the right of each slider.

Your task is to position <u>each</u> slider at exactly 50 with your mouse. You get 5 points per each slider you have positioned at 50.Suppose you position 5 sliders at exactly 50 then you will earn 25 points.

After 16 minutes, you will automatically be taken to the next task. You can also choose to move on to the next task at any time before the 16 minutes. This is done by clicking on the `next task´ button at the bottom of the screen.

Before you start this task, there will be a 2 minute practice screen.

Quiz before Task 1 (after practice screen):

1. Choose the correct one:

- The slider can be adjusted and readjusted an unlimited number of times and I will earn 5 points for every slider I move
- The slider can be adjusted and readjusted an unlimited number of times and I will earn 5 points for each slider that it is positioned at 50 before the time runs out
- I will not be paid for this task
- The slider can be adjusted and readjusted an unlimited number of times and I will earn 10 points for each slider that it is positioned at 50 before the time runs out

2. I can move to the next task:

- at the end of 16 minutes
- when I press the `next task´ button
- either of the above
- after 2 minutes

TASK 2

Your task is, again, to position sliders at exactly 50 with your mouse.

Random Treatment: Before you carry out the task, you will be affiliated randomly either with a `high performance´ or a `low performance´ group.

Free Choice Treatment: Before you carry out the task, you must choose a group affiliation. You can be affiliated with either a `high performance´ or a `low performance´ group.

Paid Choice Treatment: Before you can carry out the task, you must choose a group affiliation. But there is a fee of 100 points for the affiliation. You can be affiliated with either a `high performance´ or a `low performance´ group.

Delegation Treatment: Before you carry out the task, you must select a person to determine your group affiliation. There are four people, numbered 1-4, who have chosen an affiliation for someone else. The choice of the person you select will determine your affiliation. You can be affiliated with either a `high performance´ or a `low performance´ group.

This affiliation will NOT affect the points per slider that you get by positioning the slider at exactly 50: that is you get 5 points per slider. On average, members of the `high performance´ group position 171 sliders at exactly 50 and members of the `low performance´ group position 108 sliders at exactly 50.

Remember that you will have a maximum of 16 minutes to carry out your task this time. Every 2 minutes a new screen with sliders will appear. In each screen, you will see 48 sliders as before. Each slider has values from 0 to 100. Each slider will appear at position 0. The slider can be adjusted and readjusted an unlimited number of times and the current position is displayed to the right of each slider; and you get 5 points per each slider you have positioned at 50.

After 16 minutes, you will automatically be taken to the next task. You can also choose to move on to the next task at any time before the 16 minutes. This is done by clicking on the `next task´ button at the bottom of the screen.

Quiz before Task2:

1. Choose the correct one:

- The slider can be adjusted and readjusted an unlimited number of times and I will earn 5 points for every slider I move
- The slider can be adjusted and readjusted an unlimited number of times and I will earn more money for each correctly position slider if I am in the `high performance´ group

- The slider can be adjusted and readjusted an unlimited number of times and I will earn 5 points for each slider that it is positioned at 50 before the time runs out
- I will not be paid for this task
- The slider can be adjusted and readjusted an unlimited number of times and I will earn 10 points for each slider that it is positioned at 50 before the time runs out

2. I can move to the next task:

- at the end of 16 minutes
- when I press the `next task´ button
- either of the above
- after 2 minutes

3. Choose the correct one:

Random Treatment:

- I will be assigned to a group depending on my previous performance
- I can choose any group
- I will be assigned randomly to a group

Free Choice Treatment:

- I should choose the group exclusively depending on my previous performance
- I will be assigned randomly to a group
- I can choose any group

Paid Choice Treatment:

- I should choose the group exclusively depending on my previous performance
- I have to pay a fee and I will be assigned randomly to a group
- I can choose any group but I have to pay a fee

Delegation Treatment:

- I can choose the groups from 1 to 4

- I will be assigned randomly to a group by choosing a number from 1 to 4
- I must select a person from 1 to 4 and that person will determine the group

Group choice/assignment before Task 2:

There are two groups `high performance´ (average: 171) and `low performance´ (average: 108).

Random Treatment: You have been randomly assigned to group `high performance´ *or* `low performance´

Free Choice Treatment: Which group do you choose?

- high performance
- low performance

Paid Choice Treatment: This choice will cost you 100 points. Which group do you choose?

- high performance
- low performance

Delegation Treatment: You must select a person to determine your group affiliation. Which person do you select to determine your group affiliation?

1 2 3 4

Person <person choice> had selected `high performance´ <or `low performance´>.

Information displayed on the top of the screen in all treatments:

Your group is **high performance <***or* **low performance>**. In 16 minutes, people correctly positioned, on average, 171 sliders in the high performance group and 108 sliders in the low performance group.

Currently, the number of correctly positioned sliders is:

TASK 3

You now need to make 10 decisions for each of four successive computer screens. Each decision is a paired choice between two options (for example, "Option A" and "Option B"). You will make ten decisions and record these in the final column, but only one of them from each computer screen will be used in the end to determine your earnings. You will only know which one at the end of the experiment.

Before you start making your ten choices, please let us explain how these decisions will affect your earnings for this part of the experiment. After you have made all of your decisions, the computer will randomly select which of the ten decisions will be used to determine your earnings. In relation to this decision, the computer will then randomly select the outcome based on the probabilities assigned to the option you chose.

As an example, assume that, for the randomly selected decision, the option to the left pays 100 points with a 10% chance and 50 points with a 90% chance, while the option to the right pays 80 points with a 20% chance and 45 points with a 80% chance. Assume that you chose the option to the left for this decision; then there is a 10% chance that will earn 100 points and a 90% chance that you will earn 50 points.

Please raise your hand if you have any questions.

Information displayed on the top of the screen in all treatments:

Risk Aversion screen: No information displayed.

Ambiguity Aversion screen: ?% chance means that you do not know how likely each outcome is. Two unknown probabilities add up to 100%.

Loss Aversion screen: You will lose points in this period.

Conformism screen: It would be nice if some of you were to choose **Option __** .

TASK 4

In this task, you will be given 100 points for answering a hypothetical question.

Imagine that you have been given a job. The same job has been given to another person. Both of you have the same skills and qualifications for the job. You completed the job in 10 hours. The other person completed the job in 7 hours and was paid £42.

In your opinion, what would be the fairest amount for you to be paid (between £0 and £100)?

Submit your suggestion: _____

TASK 5

There will be many people participating in today´s experiment. You will be randomly matched with one of them. Neither of you will be told who the other participant is. Both of you receive the same instructions:

Each of you must choose a number between 11 and 20 (decimals are not allowed). Each of you will receive an amount of points 10 times the number you have chosen.

A participant will receive an additional 200 points if he or she chooses a number that is exactly one less than the number that the other participant has chosen.

Quiz before Task 5:

Choose the correct one:

- The amount of points I earn is the number I choose times 10 and if I choose exactly one number lower than the other person, I will be paid 200 extra points
- The amount of points I earn is the number I choose times 10
- If I do not choose a lower number than the other person, I will not earn anything

Choose your number: _____

Electronic Appendix 2: Additional Analyses

Table B1. Group choice and performance on task 1

	Free group choice			**Paid group choice**		
Performance Class in Task 1	**Low**	**High**	**Associative choosing**	**Low**	**High**	**Associative choosing**
Upper	0	8	0.000***	3	9	0.047**
Upper-Middle	8	6	0.605	2	17	0.001***
Lower-middle	13	7	0.057*	9	5	0.090*
Lower	13	6	0.032*	6	10	0.773
Total	**34**	**27**		**20**	**41**	

Note: The column Associative choosing indicate the level of significance of a binomial test for is the hypothesis of a probability equal to 0.5 versus the alternative of associative choosing (i.e. Upper classes favor high group and Lower classes favor low group) where *** is at 1%, ** at 5% and * at the 10% level.

Table B2. Individual performance – the effect of multiple reference points vs no reference point

	(3) Random	(4) Free Choice	(5) Paid Choice	(6) Delegation
Task 2 dummy	1.307**	2.349***	2.972***	3.165***
	(0.631)	(0.626)	(0.511)	(0.596)
Screen # in Task 1	-0.255*	0.0382	0.134	0.0597
	(0.134)	(0.109)	(0.0932)	(0.0978)
Screen # in Task 2	-0.193*	-0.132	-0.175*	-0.292**
	(0.101)	(0.0836)	(0.0995)	(0.106)
Treatment dummy	-0.372	-1.533	-0.166	-1.277
	(1.343)	(1.352)	(1.473)	(1.305)
Treatment × Task 2	1.593***	2.782***	1.302**	0.731
	(0.590)	(0.700)	(0.579)	(0.625)
Constant	28.33***	6.466	22.72***	17.64**
	(4.843)	(5.302)	(6.875)	(5.679)
Controls?	Yes	Yes	Yes	Yes
Observations	1376	1376	1376	1376

Dep. variable: Individual Performance on each screen. Standard errors clustered on subject in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B3. Individual performance – the effect of the group

	(1) All	(2) Random	(3) Free choice	(4) Paid choice	(5) Delegation
Task 2 dummy	27.76***	29.33***	36.79***	22.10***	18.96***
	(2.885)	(5.905)	(6.646)	(4.556)	(4.647)
High group	9.463*	11.39	20.82*	3.245	-8.583
	(5.627)	(10.47)	(12.39)	(14.19)	(11.43)
Task 2 × High	-3.981	-8.976	-3.127	1.656	-0.358
	(4.239)	(7.651)	(10.35)	(7.118)	(8.592)
Constant	161.1***	221.7***	54.72	229.1***	124.3*
	(26.38)	(48.77)	(56.90)	(69.93)	(64.05)
Controls?	Yes	Yes	Yes	Yes	Yes
Observations	488	122	122	122	122

Dep. variable: Individual Performance on each task. Standard errors clustered on subject in parenthesis.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$