3-17-2017

Social Information "Nudges": an Experiment with Multiple Group References

Shaun P. Hargreaves Heap
King's College London

Abhijit Ramalingam
University of East Anglia

David Rojo Arjona
Chapman University, rojoarjo@chapman.edu

Recommended Citation
Hargreaves Heap, S., Ramalingam, A. & Arjona, D. R. (2017). Social information "nudges": An experiment with multiple group references. *Southern Economic Journal, 84*(1), 348–365. doi: 10.1002/soej.12210

Social Information "Nudges": an Experiment with Multiple Group References

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

Shaun HARGREAVES HEAP (CORRESPONDING AUTHOR)
Department of Political Economy, King's College London, Strand, London WC2R 2LS, United Kingdom; s.hargreavesheap@kcl.ac.uk; Tel: +44-20-7848 1689

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

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.

Keywords: social comparison, 'nudge', effort provision, reference points

JEL classifications: C91, D03, D60

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). The provision of such information has, as a result, become a popular candidate public policy 'nudge' (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. 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.[1] 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 generalise 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. For instance, there is not only the difference between an individual and his or her group's behaviour ('within-group'); 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

[1] There are many experiments with multiple groups that have focused on how group membership affects strategic behaviour 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 behaviour 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.

and possibly other ways, then it may, in turn, alter how the provision of such group reference points affects behaviour.

This would not matter if there were only ever single reference groups but many social information policy 'nudges' naturally occur in multiple group settings. A popular public policy in education, for instance, is to provide league table information on individual school performance in standardised tests. This supplies the students (and teachers) in each school with both an own group reference point and also a reference point with respect to how other groups perform. The insight of Azmat and Iriberri (2010) suggests that own group information in isolation boosts educational performance, but does the same occur with the provision of league table information when this own group information is coupled with information about the performance of other groups? This is, in effect, our first question.

The answer to it also has a potentially wider significance for policy evaluation: for example the assessment of policies that involve introducing consumer choice into the provision of public services like education and health. This is because, while these policies are typically introduced because such choice creates a market-like pressure to improve the service, they are also accompanied by the provision of information on different providers' performance (i.e. a social information 'nudge'). However, if multiple group social information 'nudges' *per se* boost performance, then this needs to be taken into account when evaluating such consumer choice policies. It is possible, for instance, that much of the gain could arise from the social information 'nudge' and not the pressure of the market.

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 behaviour 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).[2] Both responses are understandable depending on whether the provision of own group information triggers competitive or norm-conforming preferences. Since the presence of two

[2] And in others there is no effect at all (see Eriksson *et al*, 2009). A further reason for tentativeness in summarising the literature is that studies like Falk and Ichino, 2006, and Azmat and Iriberri, 2012, 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.

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.

Groups are also often hierarchically differentiated by their performance (as in school league tables or differences between natural groups) and policy makers can be interested in how interventions affect group differences as much as individual ones. This introduces a further dimension to our first two questions: that is, are the effects of social comparison on behaviour the same for members of the 'high' performing group as for those in the 'low' performing one?

Finally when there is more than one group, the issue arises as to how individuals become affiliated with one rather than another. Our third question follows from this. It is important for public policy reasons both for anticipating the effects of policy and in guiding policy, where this is possible, over how group affiliation occurs. For example, to return to the policy of providing school league table information, do the effects of this public policy depend on whether the allocation of individuals to schools is random or 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 in more detail in the next section. Section 3 explains the experimental design. 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 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

2.1 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 behaviour (and hence behaviour 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.[3] Aggregate performance therefore increases, but there is no clear effect on the variance of individual performances within the group.

Second, there are norm-based arguments for social comparison that predict behaviour 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.

The evidence from single group reference point experiments on which type of motive best explains the 'within-group' effect on behaviour is mixed (e.g. see Azmat and Iriberri, 2010,

[3] Azmat and Iriberri (2010) offer a different information-based account of the influence that relative information has on behaviour 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.

Abeler et al., 2011, Falk and Ichino, 2006, and Mas and Moretti, 2009). The provision of multiple group reference points could also play into this debate over the 'within-group' effect in conflicting ways. 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 behaviour because any deviation from a group norm may now be associated with membership of another group. 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.

In addition, the presence of multiple group references introduces the possibility of another type of social comparison effect on behaviour: a 'between-group' comparison effect (see Frank, 1985).[4] Again, there are conflicting intuitions over such an effect. For example, 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 an array of conflicting arguments and intuitions of this kind and, since it matters whether a policy of providing group reference points either boosts performance and/or reduces inequality, this is an important matter for empirical determination.[5] This is the motivation behind our first two questions.

[4] 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).

[5] It is perhaps worth emphasising 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 behaviour. 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.

2.2 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 identify more closely with their group when they share more things in common with other group members. They may also identify more closely for this and other reasons when their group affiliation has been chosen by themselves (as compared with being settled by someone else or through the throw of a die). While this is a powerful intuition in liberal societies where freedom of choice is highly valued (and see Page *et al*, 2005, for experimental support), it is not the only one.[6] 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. This motivates our third question: it is an open question because there is no direct experimental evidence and, again, there are conflicting intuitions.

3. Experimental design

3.1 Design overview

Individuals in the experiment engage in a real effort task twice, first as isolated (as explained below) individuals with no social information (Task 1) and then as isolated individuals but with group reference point(s) (Task 2). At the beginning of Task 2, individuals receive one or two group reference point(s): they always know the performance of the group to which they are affiliated and, when there are two groups, they also know the other group's performance level. The comparison of performance between Task 1 and Task 2 (after controlling for learning/boredom) is, therefore, designed to reveal the influence, if any, of group reference point(s) on behaviour.

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

We then study whether this group reference point influence depends a) on whether there are one or two groups (Question I), and when there are two, b) whether inequality between individuals and groups changes as a result of this influence (Question II) and c) whether this influence depends on the mechanism generating group affiliation (Question III). We do this by having five different Treatments. Individuals in the Baseline Treatment have a single group reference point in Task 2: their own. It may be a High or a Low reference point. This is determined randomly, but there is always only one.[7] In Treatments 1-4, there are two group reference points in task 2: the individual's own group and that of the other group. One of these is a High reference point and the other is a Low reference point. Treatments 1-4 are distinguished by the procedure for generating group affiliation in Task 2 to either the High or the Low group.

Treatment 1 (Random) randomly allocates individuals to the two groups. This is designed to capture group affiliations that arise naturally. It also provides the appropriate comparison for behaviour in the Baseline (one group reference point) to address Question I. In Treatments 2-4, subjects select their groups (i.e. whether it is High or Low). In Treatment 2 (Free choice), subjects choose their group. They do the same in Treatment 3 (Paid Choice) but in addition they have to pay a fee (the same for both High and Low groups) to affiliate with a group. We modify Free Choice in this way because it is often argued that the introduction of a small cash nexus is enough to alter the relationship people have to each other (see Vohs et al. 2006). While subjects directly choose their group in Treatments 2 and 3, in Treatment 4 (Delegation) subjects indirectly choose their affiliation. They choose someone who will select a group on their behalf. This kind of delegated choice is of course common when there is a social process of representative decision making (e.g. as in representative democracies) but it might also be plausibly thought to capture the social process governing workplace decisions in many market settings (e.g. a worker chooses an employer and the employer then has the managerial power to control/order and so makes decisions for the worker).

Our two group reference points are mutually exclusive. This was a design choice that 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.

[7] We have another baseline control treatment where there are no groups. This provides a check on whether any increase in performance in task 2 in these group treatments is due to a between task learning/boredom effect (rather than a continuous boredom/learning effect) and/or a simple desire to improve individual performance (rather than one that is triggered by social comparison).

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.

3.2 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 the Baseline 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.[8] 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. We selected the Gill and Prowse task because there is no prior evidence of significant time trends in performance on this task (due to learning or boredom) that might otherwise render the interpretation of any change in performance between task 1 and 2 difficult. 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-incentivised two-minute practice round before beginning Task 1.

[8] Independent of each other and for the first four tasks, they were independent of any other participant.

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. In the Baseline 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.[9] 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 the Treatments 1-4 in the manner set out above (also see below). Thereafter Task 2 is identical to Task 1. 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 their choice. This fee had to be paid regardless of which group they chose, so that the size of the fee itself would not influence the choice. The only difference from Free Choice is the introduction of a cash nexus. In the Delegation treatment subjects nominated one of four previous participants, identified only by a number, to choose a group for them. Subjects were only informed that each of these four participants had taken part earlier and had made a choice of group in the past.[10] Whichever previous participant was nominated by the subject, his/her choice would be implemented for the subject.

The third, fourth and fifth tasks were designed to complement the normal demographic information that is used as controls in individual level regressions. The third was a series of incentivised lottery choices intended to measure subjects' risk aversion in the gain and loss

[9] Prior to running our main experimental treatments, we ran two sessions where 62 subjects performed just the slider task for up to 30 minutes in place of Task 1 and Task 2. The rest of the tasks they performed were the same as the ones described above. The mean number of sliders positioned correctly in the last 16 minutes by the top and bottom halves were, respectively, 171 and 108. There was no evidence of significant time trends in the performance of this task.

[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-incentivised hypothetical question after the experiment concerning whether they would choose to affiliate 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.

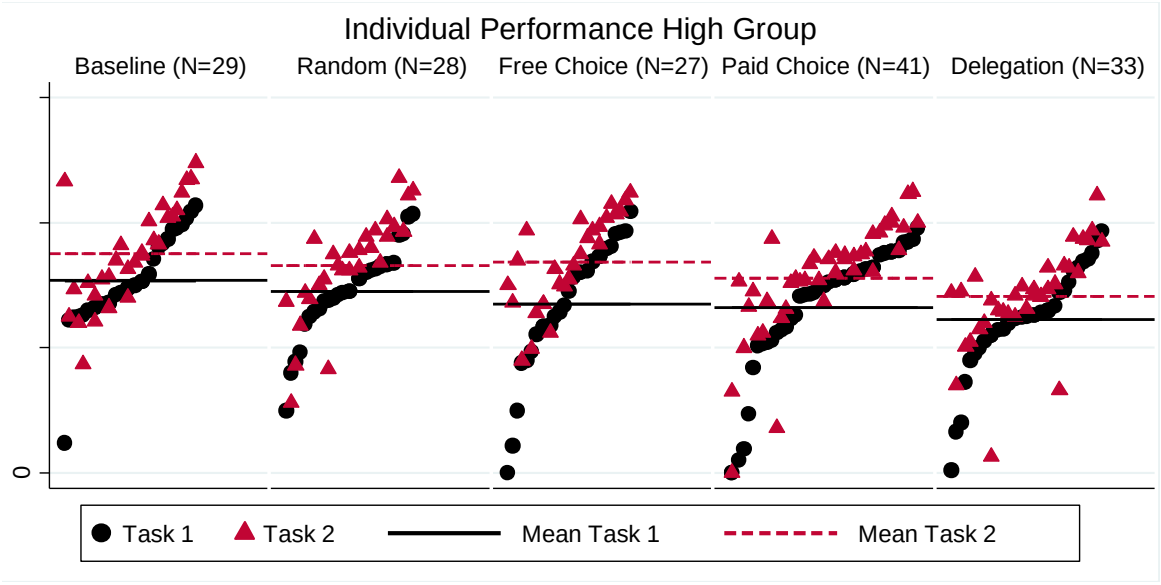
domains, ambiguity aversion and conformism (Holt and Laury 2002). The fourth task was not incentivized and subjects received a fixed payment of 100 points for answering a hypothetical question intended to capture his/her degree of self-serving bias (see Babcock and Loewenstein, 1997). The fifth task was the Arad and Rubinstein (2012) incentivised level-k task.

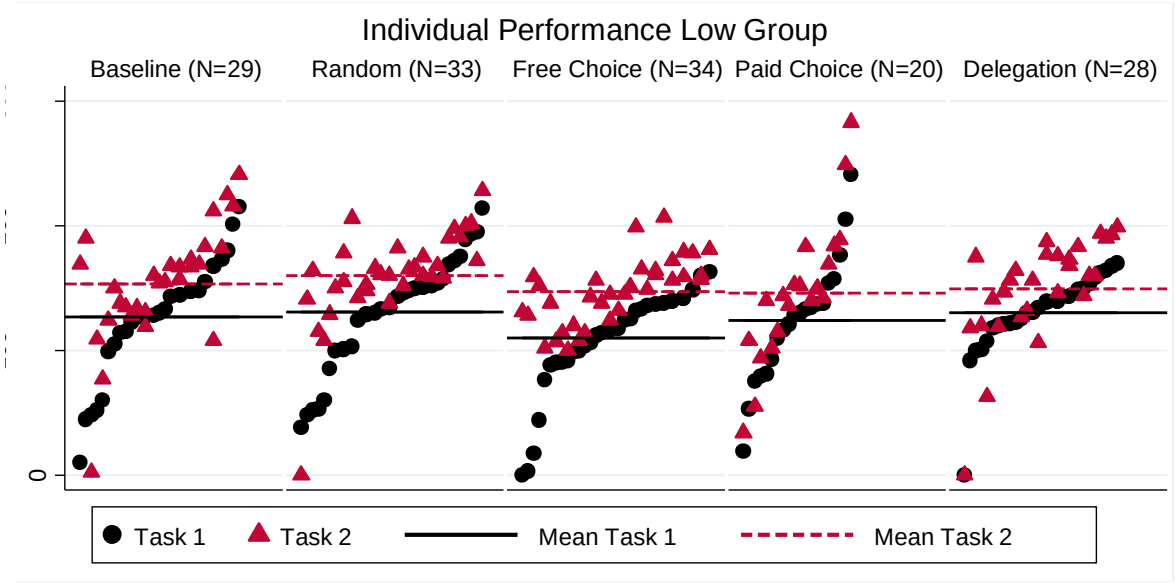
All sessions were conducted at the University of East Anglia. The experiment was completely computerised and was programmed in z-Tree (Fischbacher 2007). 244 subjects participated in our four main treatments (61 in each) and 58 subjects participated in the Baseline. Experimental earnings were calculated 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 on the day in which they participated. The average subject spent about 60 minutes in the lab and earned £13.45 including a £2 show-up fee.

4. Results

Figure 1 gives the performance on each task for each treatment for each individual, arranged in ascending order based on performance in task 1. It also disaggregates performance by the High and Low groups in each treatment.

Figure 1. Individual performance on each task by reference group and treatment





Five features stand out in Figure 1 in relation to our questions.

1) Almost everyone's performance improves in Task 2.
2) The boost is particularly marked among the very lowest performers on Task 1 (that is the lowest 10 performing subjects in each Treatment).
3) There is no marked difference with respect to 1) and 2) between High and Low groups.
4) There is no marked difference with respect to 1) and 2) across Treatments with the possible exception of Delegation in relation to 2) where there are also some very low performers on task 1 whose performance on task 2 actually worsens.
5) The boost is greatest in the Free Choice treatment

1) is supported by Wilcoxon signed rank tests on the aggregate performance on Task 2 as compared with Task 1: it is significantly different from 0 ($p < 0.01$) for each group in each Treatment. 1) is also supported by individual level regressions on performance in each task in Table 1. The task 2 dummy is significant and positive in all cases. This regression also disaggregates by screen number on each task to test whether boredom or experience is generating a time trend in performance that could complicate the interpretation of the apparent boost to individual performance. The only significant time trends in task 1 are in the Random treatment, where it is negative, and in the High Baseline, where it is positive; and, in task 2 under Delegation, where it is negative. The negative time trend in task 1 for Random reinforces the conclusion that there is a jump in task 2 because in so far as the time taken on

task 1 had been increased and apparent boredom had continued to reduce performance, then the jump on task 2 would have been even bigger. The time trend in Delegation works in the opposite direction. If boredom had continued to affect performance then a longer period on both tasks would have reduced the boost to performance in task 2 in this treatment. The positive time trend in task 1 for the High Baseline would for the same reason suggest that the boost in task 2 would have appeared smaller had learning continued over a longer time period in task 1.[11]

Table 1. Individual performance on each screen (Random effects model)

	High Baseline	Low Baseline	Random	Free Choice	Paid Choice	Delegation
Task 2	4.119***	3.447***	1.981***	4.638***	4.037***	3.738***
	(0.804)	(1.012)	(0.622)	(0.715)	(0.586)	(0.634)
Screen # in Task 1	0.366***	0.018	-0.379**	0.034	0.168	0.064
	(0.125)	(0.209)	(0.175)	(0.139)	(0.114)	(0.122)
Screen # in Task 2	0.040	-0.009	-0.113	-0.027	-0.087	-0.252**
	(0.154)	(0.167)	(0.110)	(0.073)	(0.106)	(0.120)
Constant	30.913***	16.867***	29.65***	4.801	27.68***	13.79**
	(5.626)	(5.772)	(5.548)	(7.038)	(8.238)	(7.004)
Controls?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	464	464	976	976	976	976

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

> **Result 1:** *After controlling for possible learning/boredom, individual performance on Task 2 is greater than Task 1 in all Treatments. Had the time spent on each task been increased, the significant positive/negative time trend in Task1/Task 2 in the Baseline(High)/Delegation treatments could qualify this result for these treatments.*

To examine whether the boost to performance in Task 2 differs when there are two group reference points rather than one, we compare the boost for those in the Baseline High group

[11] We also controlled for the possibility that learning might occur discretely between task 1 and task 2 rather than on screen-by-screen basis (or that individuals simply wanted to do better on task 2 than task 1) by running a further baseline control where there were no group affiliations. The results are reported in the electronic appendix. Performance on task 2 was significantly higher than on task 1, but the boost was significantly lower (between one half and one third) than in all our Treatments experiments and than in the Low group version of the Baseline.

with that achieved by the High group in the Random Treatment; and analogously the boost in Baseline Low group with the boost in the Low group in the Random Treatment. The corresponding Mann-Whitney tests reveal no significant difference for either the High group or the Low group affiliation ($p = 0.837$ and $p = 0.462$, respectively).

> **Result 2:** *There is no difference in the boost to average performance in Task 2 between the Baseline Treatment (single group reference) and the Random treatment (two group reference points with random allocation).*

There is no significant difference in the boost between High and Low groups in a Wilcoxon test ($p = 0.1882$). This result is 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.

> **Result 3:** *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 next consider whether individual behaviour converges to the own group norm. If the effect of social information is to encourage higher performance by all individuals, then those who were performing below the group norm will appear to converge towards the group norm for this reason. So for these individuals behaviour motivated by convergence to the norm is indistinguishable from behaviour aimed at improving performance. The two types of behaviour pull in opposite directions for those who were originally performing above their group average (on Task 1) and so they are the individuals to consider for this purpose. Table 2 disaggregates between those who were initially above/below their respective group norms on Task 1 and performs a Wilcoxon signed-rank test to see in which direction the performance boost is significant for each type of subject. Although the performance boost is typically lower for those who performed above the norm on Task 1, the change in Task 2 is still very clearly positive (and of similar magnitude for those in the Low as those in the High groups and this is what we also find in the analogous individual regressions).

Table 2. Means and Tests for zero performance boost by initial performance relative to norm

	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	128.418	154.069	138	29.659	8.328	106	20.434	8.358
	(45.103)	(41.690)		(41.251)	[0.0000]		(15.611)	[0.0000]
High groups	132.985	156.759	104	24.74	7.276	25	19.76	4.253
	(46.664)	(42.688)		(38.352)	[0.0000]		(12.81)	[0.0000]
Low groups	123.296	151.052	34	44.706	4.335	81	20.642	7.247
	(42.909)	(40.515)		(46.516)	[0.0000]		(16.446)	[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.

> **Result 4:** *There is no evidence that those whose performance in Task 1 is above the norm in their group conform to this norm either by <u>not</u> increasing or by reducing their performance in Task 2.*

Nevertheless, Table 2 suggests that the variance in performance falls in Task 2 as compared with that in Task 1 (standard deviation across all groups falls from 45.1 to 41.69) and this occurs because low performers on Task 1 improve by more than the high ones in both groups. The marked improvement by the particularly poor performers on task 1 was also evident in Figure 1. We test for this formally in the regression that is reported in Table 3. 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. 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 3. 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 In Task 2	-30.68***	-35.37**	-42.30***	-20.25*	-22.83
	(6.726)	(14.04)	(13.73)	(11.96)	(16.22)
Upper-middle In Task 2	-32.94***	-29.07**	-43.41***	-30.07***	-24.74
	(6.512)	(13.05)	(13.97)	(11.40)	(15.30)
Lower-middle In 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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dep. variable: Individual Performance on each task, leading to two observations per individual. Standard errors clustered on subject in parentheses.

> **Result 5:** *The very lowest performers (i.e. those below 108) in Task 1 increase their performance on Task 2 by significantly more than other individuals, except in the Delegation Treatment.*

Finally we examine whether the boost varies across 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). Figure 1 points to considerable individual variation in these boosts and so we also test for differences through Treatment dummies interacted with Task 2 in the individual performance regressions in Table 4. 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 with the other Treatment interact dummies is significant in both cases (p = 0.0494 for Paid Choice and p = 0.0105 for Delegation).

Table 4. 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 in Task 2	10.20*	13.31	7.461
	(6.192)	(9.085)	(8.715)
Paid Choice in Task 2	-2.000	3.399	-7.233
	(5.367)	(7.143)	(7.312)
Delegation in 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

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

After regression (2), no treatment differences.
After regression (3), Free vs. Paid, p = 0.0628; Free vs. Delegation, p = 0.0249.

Result 6: *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. They have, for instance, 3 implications for policy. 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.[12] 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'. Third, the policy of promoting consumer choice in the provision of public services needs careful assessment when accompanied by social information because much of its effects may be coming from the social information 'nudge'.

They are also new because multiple group social information has typically not been examined experimentally. The evidence on social information 'nudges' comes from single group reference points, yet 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.[13] 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 (2012) 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

[12] 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 De Araujo *et al*, 2015).

[13] Although we have not reported it here as part of the results, there is evidence in our experiment of the Falk and Knell(2004) prediction that, in effect, people will choose their groups associatively: that is high/low performers choose the High/Low group (see electronic appendix for details). More generally, this suggests group affiliation matters for our subjects in this experiment (and so plausibly motivates them even though the groups are minimal, otherwise there would be no reason to expect the choice of group to be other than random.

performers (= above average in their study). It is also consistent with the findings of some peer group studies (see Carrell et al., 2009).

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 behaviour 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.

Acknowledgements

The authors thank Pablo Brañas-Garza, David Cooper, Phil Grossman, Tim Salmon, Roberto Weber 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.

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.

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. 2012. “The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness.” *Centre for Economic Performance Discussion paper 1116*.

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.

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.

Clark, A. and A. Oswald. 1996. “Satisfaction and Comparison Income.” *Journal of Public Economics*, 61(3), 359-81.

De 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.”, *Working Paper.*

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.

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

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

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.

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." The 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.

Sheremeta, R. 2010. "Experimental comparison of multi-stage and one-stage contests." *Games and Economic Behaviour,* 68 (2), 731-747.

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

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

Appendix A: 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 each 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: _____

Appendix B: Additional Analysis

Table B1. Mean performance

	N	Task 1	Task 2	Difference
Random affiliation	61	137.344	162.557	25.21312***
Free choice	61	121.049	156.459	35.40984***
Paid choice	61	129.361	152.574	23.21312***
Delegation	61	125.918	144.689	18.77049***
Only high group	29	153.655	175.448	21.7931***
Only low group	29	126.724	153.207	26.48276***
No Reference	25	144.96	158.32	13.36***

The *** indicate that, in all treatments, the boost in performance in Task 2 is statistically significant at the 1% level.

p-values for Mann-Whitney tests comparing the difference (increase in effort in Task 2) between the No ref treatment and the other treatments:

1. Random: p = 0.0114
2. Free Choice: p = 0.0014
3. Paid Choice: p = 0.0752
4. Delegation: p = 0.0912
5. Only High reference: p = 0.1990
6. Only Low reference: p = 0.0200

Table B2. 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.