Two strikes and you are out!
An experiment on exclusion

Jonathan Levy[*]

October 5, 2019

Abstract

In many cases poor performers must receive a warning in the form of a strike prior to exclusion. This study provides a comparison between a two-strike exclusion policy and a zero-tolerance exclusion policy as a means for fostering cooperation in groups. The results from our experiment suggest that group members tend to cooperate more after receiving a strike. However, requiring group members to issue strikes to one another prior to exclusion seemed to be less effective than allowing for exclusion without prior receipt of strikes.

Keywords: Exclusion, ostracism, public good, teamwork, experiment.

JEL Classification: C92, H41, D23.

[*]University of Technology Sydney, Economics Discipline Group, PO Box 123, Broadway NSW 2007 Sydney, Australia. jonathan.levy@uts.edu.au

1 Introduction

The literature on exclusion has only examined forms of exclusion without the need for issuing prior notice. Such forms of exclusion are comparable to the zero-tolerance approach utilized by US companies to dismiss employees for poor performance. However, there are cases where poor performers are notified about their performance prior to exclusion. For example, in many workplaces if an employee is accused of performing poorly or in a negligible manner, they may be subjected to a Performance Improvement Plan.[1] While the employee's performance is being evaluated, they must adhere to the necessary improvements, otherwise they will receive a series of strikes. In this example, employers have grounds to dismiss an employee if they have received the necessary number of strikes while on the plan.

One could argue that the n-strike exclusion policy may be more efficient than the zero-tolerance exclusion policy. We focus on two-strikes without loss of generality and test in the lab. The exclusion of potentially valuable group members will inevitably have a negative impact on the welfare of the group as the productive capacity of the group will no longer be as large. A two-strike exclusion policy addresses this inefficiency by requiring group members to issue strikes to one another prior to excluding them. Thus, group members are now able to disclose their disapproval to one another which might provide sufficient incentives for cooperation without necessitating the exclusion of any members of the group.[2]

On the other hand, in a setting where we require group members to issue strikes to one another prior to exclusion the initial level of cooperation might be relatively low as there is no immediate threat of exclusion prior to receiving a strike. After group members receive strikes from one another the level of cooperation may rise to a significantly higher level. Whether cooperation levels rise enough to justify the use of a two-strike exclusion policy instead of a zero-tolerance exclusion policy is difficult to establish. Our experiment allows us to determine which of the two exclusion policies is most effective in fostering cooperation in groups.

This paper adds to the emerging literature on exclusion in groups. Neuhofer and Kittel (2015) and Solda and Villeval (2019) found strong evidence to suggest that lengthening the duration of exclusion has a positive effect on the level of cooperation within groups. Cinyabuguma et al. (2005) and Maier-Rigaud et al. (2010) used a majority voting rule to determine exclusion in groups. Whereas, Croson et al. (2015) simply excluded the lowest contributor in each group after each period. More recently Kopányi-Peuker et al. (2018) varied the periods for which one could be excluded. In all instances the threat of exclusion appeared to have a prevailing positive impact on cooperation regardless of the length of exclusion, or how the individual facing exclusion was determined. To our knowledge, this is the first experimental study evaluating a two-strike exclusion policy.

---

[1]Performance Improvement Plans are mostly used in offices around the UK, Europe and Australia.

[2]Masclet et al. (2003), Andreoni and Petrie (2004) and Rege and Telle (2004) have shown that people tend to behave more cooperatively when faced with the threat of disapproval. Moreover, Sheremeta et al. (2011) and Charness and Yang (2014) provide evidence to suggest that individuals will cooperate more after being given a second chance to participate in a group.

2 Experimental design and procedures

102 students from the University of Technology Sydney (UTS) participated in a study which contained six experimental sessions of a computerized experiment, programmed in z-Tree; see Fischbacher (2007). The experiment was conducted in the UTS Behavioural Lab. Participants were recruited through ORSEE; see Greiner (2015). This experiment was a between subject design. The following table outlines the overall structure of the experiment.

Table 1: Treatment summary

Treatment	Number of groups	Games played
No exclusion	6	10 period game + 10 period game
Exclusion without strikes	6	10 period game + 10 period game
Exclusion with strikes	5	10 period game + 10 period game

In each treatment subjects participated in a repeated linear public good game. Participants were randomly assigned to groups of 6 and stayed in the same groups throughout each 10-period game. In each period, participants first had to simultaneously choose how much to contribute to the Group account out of an endowment of 10 Francs. 1 Franc was equal to 0.025 AUD. The payoff for participant i was calculated according to the following function:

$$\pi_i = 10 - g_i + 0.4 \sum_{j=1}^{n} g_j$$

where g is the contribution to the Group account. The choice of these parameter values ensures that the social optimum is for everyone to fully invest in the Group account, while the individuals' strictly dominant strategy is to invest nothing.

After the contribution stage each group member was informed about the contribution levels of the other members of their group. However, unlike in Maier-Rigaud et al. (2010), contributions made by other group members were shown in a fixed order on the screen in all treatments. Thus, it was possible for each group member to track other group members' contributions over time.[3] This measure was taken to control for the reputation effect across all treatments. In the exclusion without strikes treatment, after examining the contributions made by others, each group member could then assign disapproval points to exclude other group members or to refrain from assigning disapproval points. Each group member was able to allocate any whole number of disapproval points between 0 and 10 to as many members of the group as they wished to.[4]

In both exclusion treatments group members were only informed about the total number of disapproval points they received upon exclusion. Participants needed to receive more than 50% of the maximum number of disapproval points that could possibly be assigned to be excluded, i.e. at least 26 out of 50 if the group consisted of six members. Excluded members continued to receive 10 Francs for all remaining

---

[3]Further to this, participants were provided record sheets to fill out to enhance their ability to track contributions made by other group members over time.

[4]This voting format has been implemented in other studies such as Masclet et al. (2003) and Sheremeta et al. (2011). It allows for richer empirical analyses as the dependent variable is subject to a larger degree of variation.

periods and did not participate in future play of the game. The direct effect of excluding a group member was the decrease in group size. Hence, the decision to exclude a group member always reduced the potential maximum contribution to the Group account.

The exclusion with strikes treatment was very similar to the exclusion without strikes treatment described above, however, there was one crucial difference between the two treatments. The difference was that in the exclusion with strikes treatment participants had to receive a strike prior to being excluded. If a group member received more than half the maximum number of disapproval points possible but **had not** previously received a strike, then that individual would receive a strike; if a group member **had** previously received a strike, then that individual would be excluded in the same way as in the treatment without strikes. Note, after a group member had received a strike the group was made aware at every feedback stage that this individual had previously received a strike. This allowed each group member to understand the consequences of their disapproval assignment.

Our experimental design yields the following testable hypotheses regarding the contributions made towards the Group account by participants.

Hypothesis 1: Individuals will contribute more after receiving a strike than prior to receiving a strike.

Hypothesis 2: Individual contributions will be highest in the exclusion without strikes treatment, lower in the exclusion with strikes treatment, and lowest in the no exclusion treatment.

Hypothesis 3: Contributions at the group level will be lowest in the no exclusion treatment. It is unclear which of the two exclusion treatments will yield higher group level contributions.

For a detailed derivation of the above hypotheses see appendix A.

3 Results

3.1 Individual level contributions

The following figure illustrates how individuals responded to receiving strikes as well as the average individual contribution level across treatments.[5]
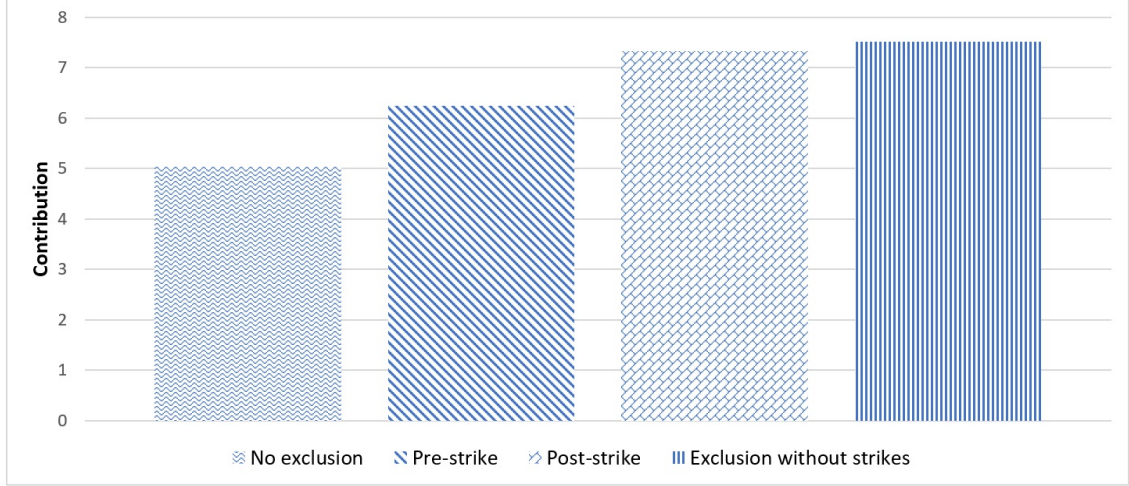

Figure 1: Individual contribution levels across treatments

The pre-strike contribution level for an individual represents the average of all contributions made towards the Group account by an individual prior to receiving a strike. The post-strike contribution level for an individual represents the average of all contributions made towards the Group account after receiving a strike.[6] Figure 1 indicates that individuals tended to contribute more after receiving a strike (6.25 vs. 7.33). We estimated a random-effect model of individual contribution on treatment dummy-variables and on the dummy-variable for whether an individual previously received a strike, while clustering errors at the group level. Our estimates indicate that the response to strikes was significant (p-value < 0.05 for periods 1-10 and periods 11-20). Refer to Table B.1 in appendix B for more details.

Result 1: *Individual contributions increased in response to receiving a strike.*

Result 1 provides strong support for the first hypothesis that after receiving a strike, group members suddenly faced a genuine threat of exclusion. Hence, stricken group members made contributions towards the Group account as if they were operating with the fear of exclusion.

Our estimates indicate that individual contributions were higher in both exclusion treatments than in the no exclusion treatment (p-value < 0.05 for periods 1-10 and periods 11-20 for both treatment dummies). The coefficient estimates for the treatment dummies also indicate that individual contributions were higher in the

---

[5]Note, we did not include observations relating to excluded individuals in our average contribution calculations in the Figure 1.

[6]Note, we did not include observations where individuals received their first strike after the 8th period in the game because strikes received after the 8th period should have no effect on contribution behaviour. We also did not include contributions that were made in the final period in our calculation of the average contribution due to end-game effects.

exclusion without strikes treatment than in the exclusion with strikes treatment. The post-strike contribution level was not significantly different to the individual level contribution in the exclusion without strikes treatment.[7] Refer to Table B.1 in appendix B for more details.

Result 2: *On average individuals contributed the most towards the Group account in the exclusion without strikes treatment, less in the exclusion with strikes treatment and the least in the no exclusion treatment.*

Result 2 provides support for the second hypothesis. Specifically, participants behaved much more cooperatively when faced with a genuine threat of exclusion. In the exclusion without strikes treatment participants only faced a threat of exclusion after they received a strike. This in turn explains why contributions made towards the Group account were higher in the exclusion without strikes treatment than in the exclusion with strikes treatment.

3.2 Group level contributions

Figure 2 provides a graphical representation of contributions made at the group level.
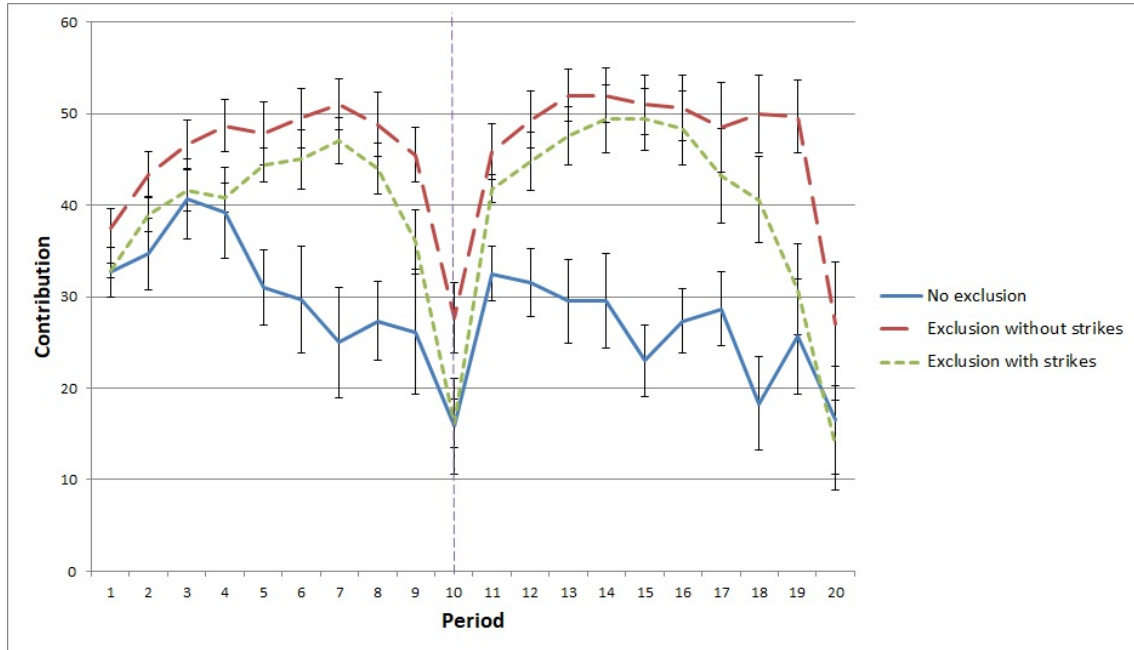

Figure 2: Contributions at the group level

We estimated a random-effect model of group level contribution on treatment dummy-variables, while clustering errors at the session level. Our estimates indicate that contributions made towards the Group account at the group level were higher in both exclusion treatments than in the no exclusion treatment (p-value < 0.05 for periods 1-10 and periods 11-20 for both treatment dummies). The coefficient estimates for the treatment dummies also indicate that contributions made towards the

---

[7]The difference in the estimated coefficients between the two exclusion treatments was statistically significant for periods 1-10 but not for periods 11-20.

Group account at the group level were higher in the exclusion without strikes treatment than the exclusion with strikes treatment.[8] Refer to Table B.2 in appendix B for more details.

Result 3: *On average contributions made at the group level towards the Group account were highest in the exclusion without strikes treatment, lower in the exclusion with strikes treatment and the lowest in the no exclusion treatment.*

Result 3 provides support for hypothesis 3. This result highlights that efficiency was highest in the exclusion without strikes treatment and lowest in the no exclusion treatment. Hence, the positive effect that the threat of exclusion had on cooperation seemed to overshadow the negative impact on overall efficiency derived from the inability to signal disapproval in the zero-tolerance setting.

We found no significant difference in the prevalence of exclusion between the two exclusion treatments. The average prevailing group size in the exclusion with and without strikes treatments was 4.60 and 5.08 respectively.

4 Conclusion

Prior to conducting this research, the impact that a two-strike exclusion policy would have in fostering cooperation in groups was unclear. We found that individuals tend to cooperate more after receiving a strike. This finding highlights the possibility for undesirable group members to rehabilitate themselves without the need for suspension or expulsion. However, requiring group members to issue strikes to one another prior to exclusion seemed to be less effective than allowing for exclusion without prior receipt of strikes. As a result, the zero-tolerance approach commonly utilized by US companies to dismiss employees for poor performance may indeed be the most effective and efficient form of exclusion for fostering cooperation in groups.

Acknowledgment

---

[8]The difference in the estimated coefficients between the two exclusion treatments was statistically significant for periods 1-10 but not for periods 11-20.

List of references

Andreoni, J. and Petrie, R., 2004. "Public goods experiments without confidentiality: a glimpse into fund-raising." *Journal of Public Economics*, 88(7-8), pp. 1605-1623.

Charness, G. and Yang, C.L., 2014. "Starting small toward voluntary formation of efficient large groups in public goods provision." *Journal of Economic Behavior and Organization*, 102, pp. 119-132.

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

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

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

Greiner, B., 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1), pp. 114-125.

Kopányi-Peuker, A., Offerman, T. and Sloof, R., 2018. "Team production benefits from a permanent fear of exclusion." *European Economic Review*, 103, pp. 125-149.

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

Masclet, D., Noussair, C., Tucker, S. and Villeval, M.C., 2003. "Monetary and non-monetary punishment in the voluntary contributions mechanism." *American Economic Review*, 93(1), pp. 366-380.

Neuhofer, S. and Kittel, B., 2015. "Long-and short-term exclusion in the public goods game: An experiment on ostracism." Department of Economic Sociology, University of Vienna.

Rege, M. and Telle, K., 2004. "The impact of social approval and framing on cooperation in public good situations." *Journal of Public Economics*, 88(7), pp. 1625-1644.

Sheremeta, R.M., Tucker, S.J. and Zhang, J., 2011. "Creating Self-Sustained Social Norms through Communication and Ostracism."

Solda, A. and Villeval, M.C., 2019. "Exclusion and Reintegration in a Social Dilemma." *Economic Inquiry*.

A   Explanation for hypotheses

Subgame perfect equilibrium

As highlighted in Maier-Rigaud et al. (2010), employing standard game theoretic techniques will not allow for derivation of any meaningful predictions. To illustrate this point, consider a setting where it is not necessary to issue strikes prior to exclusion. It is evident that in the last period all players remaining in the game will contribute zero towards the Group account, so that the payoff is 10 for all players (including the excluded ones). By backwards induction, cooperation unravels from the end until the beginning of the game and the only requirement for subgame perfection is for each member of the group to contribute zero in every period. Any configuration of disapproval assignment and group sizes can be part of an equilibrium, that is, there exist a multiplicity of equilibria with different disapproval assignment levels and corresponding different group sizes throughout the game. A similar rationale can be used to stipulate the same behaviour should be expected when an individual can only be excluded after they have received a strike. Hence, subgame perfection predicts no difference in the level of cooperation across the three treatments.

Explanation for hypothesis 1

Suppose the average contribution made towards the Group account when group members **do not** face a threat of exclusion is g_{NE}, and the average contribution made towards the Group account when group members **do** face a threat of exclusion is g_E.[9] Essentially, the threat of exclusion raises the cost of free riding. As a result, we expect group members to contribute less towards the Group account when they do not face a threat of exclusion than when they are faced with a threat of exclusion. Research conducted by Cinyabuguma et al. (2005), Maier-Rigaud et al. (2010), Croson et al. (2015) provides strong evidence to support this claim. In light of this, we assume that $g_E > g_{NE}$. Consider the treatment where we only allow for a group member to be excluded after they have received a strike. Before a group member receives a strike, they face no threat of exclusion. However, after a group member receives a strike that group member suddenly faces a genuine threat of exclusion. Essentially, this implies that after receiving a strike the representative group member switches from contributing g_{NE} to g_E. Given the fact that we have assumed $g_E > g_{NE}$, we can claim that individuals will contribute more after receiving a strike than prior to receiving a strike.

Explanation for hypothesis 2

Suppose the public good game is repeated finitely for $T + 1$ periods and the representative group member receives a strike in period $t \in [1, 2, 3, ..., T + 1]$. Note, in the final period of the game there can be no threat of exclusion as the game will cease to continue after this period. Since there is never a threat of exclusion in the final period of the game, we expect contributions to be the same in this period across all three treatments. As a result, we will focus on analysing average contributions over T periods. For t periods the representative group member will contribute g_{NE}

---

[9]Note, we do not include excluded individuals in the calculation of average individual contribution.

as they face no threat of exclusion. For the remaining $T - t$ periods the representative group member faces a genuine threat of exclusion, hence, they contribute g_E. Therefore, the average contribution made by the representative group member in the exclusion with strikes treatment, g_{ES}, over T periods can be expressed by the following:

$$g_{ES} = \frac{tg_{NE} + (T - t)g_E}{T}.$$

Now that we have derived an expression for the average contribution level in the exclusion with strikes treatment it is possible to make comparisons between average contribution levels in each of the three treatments. First, we must evaluate the difference in average contributions between the exclusion without strikes and the exclusion with strikes treatments. We have that

$$\begin{aligned} g_E - g_{ES} &= g_E - \frac{tg_{NE} + (T - t)g_E}{T} \\ &= \frac{t(g_E - g_{NE})}{T}. \end{aligned}$$

Recall, $g_E > g_{NE}$ by assumption. We also know by construction that group members will never receive strikes prior to the first period, i.e. $t > 0$. Therefore, it must follow that $g_E > g_{ES}$.

Now we must compare average contribution levels between the exclusion with strikes treatment and the no exclusion treatment. We have that

$$\begin{aligned} g_{ES} - g_{NE} &= \frac{tg_{NE} + (T - t)g_E}{T} - g_{NE} \\ &= \frac{(T - t)(g_E - g_{NE})}{T}. \end{aligned}$$

If a group member receives their first strike in the second last period, T, then they cannot be excluded in the final period of the game. Therefore, it is futile giving another member of your group a strike in any period after the third last period of the game as it will most likely not alter their behaviour. For this reason, we believe group members will typically receive a strike at some point before the penultimate period, i.e. $t < T$. Combining this with the assumption that $g_E > g_{NE}$ leads to the proposition that $g_{ES} > g_{NE}$. As a result, we expect $g_E > g_{ES} > g_{NE}$.

Explanation for hypothesis 3

We established in hypothesis 1 that individual level contributions towards the provision of the public good should be higher in the two exclusion treatments than in the no exclusion treatment. In turn, we also expect this to be true at the group level.

In the discussion that follows we highlight the trade-offs between the exclusion with and without strikes treatments. First, consider a setting where we allow for exclusion without prior receipt of strikes. In such settings group members might behave extremely cooperatively throughout the course of the game as they will constantly be wary of the threat of exclusion. In contrast, in a setting where we require group members to issue strikes to one another prior to exclusion the initial level of cooperation might be relatively low as there is no immediate threat of exclusion prior to

receiving a strike. After group members receive strikes from one another the level of cooperation may rise to a significantly higher level. However, this increase in cooperation may be too little too late. Assuming that the group size is fairly stable across treatments the fact that group members might behave more cooperatively in the exclusion without strikes treatment implies that contributions made at the group level will also be higher in this treatment. Note, this reasoning is in line with hypothesis 2. Hypothesis 2 states that contributions at the individual level will be higher in the exclusion without strikes treatment than in the exclusion with strikes treatment. Consequently, one might expect to generate this result at the group level as well.

It is important to highlight that the exclusion of potentially valuable group members will inevitably reduce the productive capacity of the group. In light of this, group members operating in the exclusion without strikes treatment may be reluctant to exclude others as they may be wary of the negative repercussions. To illustrate the inefficiencies that may arise in the exclusion without strikes treatment consider the following example. Suppose we have a group of six individuals and five of them consistently contribute 90% of their endowment while the sixth group member consistently contributes 50% of their endowment towards the provision of the public good. In a setting where group members are not able to issue strikes to one another prior to exclusion it is likely that the sixth group member's relatively small contribution will be tolerated, as the group would prefer to keep this relatively small contributor in the group than to exclude them from the group. A two-strike exclusion policy may address this inefficiency. By requiring group members to issue strikes to one another prior to excluding them, group members are now able to disclose their disapproval to one another. Masclet et al. (2003), Andreoni and Petrie (2004) and Rege and Telle (2004) have shown that people tend to behave more cooperatively when faced with the threat of disapproval. Moreover, Sheremeta et al. (2011) and Charness and Yang (2014) provide evidence to suggest that individuals will cooperate more after being given a second chance to participate in a group. Hence, the two-strike exclusion policy may provide sufficient incentives for cooperation without necessitating the exclusion of any members of the group. As a result, requiring group members to issue strikes to one another prior to exclusion may in fact be more efficient than not having this requirement.

B   Estimation tables

In Table B.1 we estimate a random-effects regression of individual contribution on treatment dummy-variables and a dummy-variable for whether an individual previously received a strike, while clustering errors at the group level. Note, the baseline treatment is the no exclusion treatment.

Table B.1: Treatment comparison – Individual level contributions

Variable	Periods 1-10		Periods 11-20	
Exclusion without strikes	2.552***	2.552***	3.870***	3.870***
	(0.657)	(0.657)	(0.751)	(0.752)
Exclusion with strikes	1.591**	1.404**	2.786***	2.365***
	(0.265)	(0.634)	(0.747)	(0.793)
Received strike		1.323***		1.786***
		(0.213)		(0.642)
Constant	5.036***	5.036***	4.375***	4.375***
	(0.573)	(0.574)	(0.521)	(0.522)
No. of observations	991	991	980	980
Prob $> \chi^2$	0.000	0.000	0.000	0.000

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In Table B.2 we estimate a random-effects regression of group level contribution on treatment dummy-variables, while clustering errors at the session level. Note, the baseline treatment is the no exclusion treatment.

Table B.2: Treatment comparison – Group level contributions

Variable	Periods 1-10	Periods 11-20
Exclusion without strikes	13.007***	18.064***
	(2.674)	(3.855)
Exclusion with strikes	8.687***	15.266***
	(1.968)	(4.102)
Constant	29.896***	25.283***
	(1.725)	(3.730)
No. of observations	170	170
Prob $> \chi^2$	0.000	0.000

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$