Do professions curb free-riding? An experiment

Michał Krawczyk[1] · Krzysztof Szczygielski[1]

Abstract

The question of ethical conduct is key for professionals, such as lawyers, doctors, or experts of different kinds. We run a laboratory experiment aimed at investigating whether acting within a profession leads to more (or less) ethical, prosocial behaviour compared to acting outside of it. We also investigate how professionals react to others' misbehaviour. We invite subjects studying or having studied economics, law or medicine and either match them in mixed groups or in homogeneous groups (telling them that we did so). We then let them play public goods games with punishment. Overall, there is little difference in cooperation levels and patterns of punishment between the homogeneous and heterogeneous groups. If anything, our subjects free ride more when matched with their peers than in a mixed group.

1 Introduction

Many services important for all modern societies, notably legal services and health care are provided by "professions". While the term is often imprecisely defined, some key characteristics of an "ideal type" profession are easily identifiable.

Professions typically provide services (rather than physical goods); these are often sophisticated and entailing a high degree of information advantage over the consumer. A formal organization often exists, which oversees the process of joining the profession, i.a. verifying the competence of a candidate acquired in the course of the education process. It also represents the professionals vis-à-vis the government and other actors in the society. It may also take some of the judiciary duties with respect to the members of the profession, particularly in case they violate a codified

✉ Krzysztof Szczygielski
  kszczygielski@wne.uw.edu.pl

[1] University of Warsaw, Warsaw, Poland

or tacit code of ethical conduct. In short, professional organizations frequently perform some regulatory functions (cf. Ogus 1995, 2000; van den Bergh 2004).

Because information is asymmetric and formal sanctions limited (and often subject to the decision of the self-regulatory organization), professionals are hoped to observe high ethical standards. In this study we investigate one aspect of willingness to adhere to professional ethics. We want to verify whether professionals are more likely to behave ethically, and whether they are more likely to punish others' misbehavior, when confronted with fellow professionals, compared to facing a mixed group. These two effects would contribute to sustainability of high ethical standards within a profession. Then again, it is also conceivable that professionals are *less* willing to punish their peers, which could lead to the opposite effect. The mechanisms we are considering are logically independent of two other important factors: whether ethical individuals tend to self-select to some professions (Handy and Katz 1998; Brekke and Nyborg 2010) and how professional training affects ethical decision making (Frank et al. 1993); the latter issues received considerably more attention in experimental literature, at least as applied to the economist profession.

The paper is structured as follows. In Sect. 2 we briefly review the literature on professions and on the relevant economic experiments. Then, in Sect. 3, we present the experiment and in Sect. 4 we discuss its outcomes. Section 5 concludes.

2 Background

The main reason professional markets are regulated is that there is a strong information asymmetry between the buyer and the seller. Indeed, professional services are, as a rule, either experience goods, i.e. the consumer cannot assess their quality prior to purchase, or, even more frequently, credence goods, implying that the consumer is unable to judge the quality (or even the necessity) of the service, even ex post.[1] This means that professionals may be tempted to provide services that are of insufficient quality, not really needed, and/or to overcharge their customers (moral hazard). Furthermore, some of the individuals with higher ethical standards may be discouraged from working in the profession, while some of those inclined to opportunism may join it (adverse selection). As a result, demand for such services may dwindle and their quality may be suboptimal, both with negative externalities for the economy at large (Van den Bergh 2004).

Given the high level of complexity of professional knowledge, governments often delegate regulatory functions to the professions themselves in the hope of achieving a competent, flexible and cheap (for the government) legal process.[2] There is a long history of self-regulation of professions such as lawyers, doctors, and other medical

---

[1] See Nelson (1970) for the discussion of experience goods, and Dulleck and Kerschbamer (2006) for a review of works on credence goods.

[2] The self-regulatory organizations of professionals are also regarded as a way the professions try to handle the moral hazard and adverse selection problems mentioned earlier in this section. See the models by Leland (1979), Gehrig and Jost (1995), or Baron (2010).

occupations (cf. Van den Bergh 1999, 2004; Philipsen and Faure 2002; Abbott 1983, 2014; Larson 2013). All the same, self-regulation poses the risk of abuse, two obvious example being that the professionals may seek to restrict the competition in the marketplace, and that they might not always be willing to take disciplinary actions against their misbehaving colleagues.

In this study we focus on the latter problem. On one hand, the professionals should be motivated to fight the misconduct, as it increases the social trust in the profession, reinforces the case for professional autonomy and prevents the government from stepping in. Several models in industrial organization demonstrated that the prospect of possible government regulation, or the risk of entering a political fight, can influence the self-regulatory organizations in their decisions to set and execute the quality standards, cf. DeMarzo et al. (2005), Heyes (2005), Grajzl and Baniak (2009), Maxwell et al. (2000), and Baron (2011).[3] The importance of group reputation was also studied in theoretical biology, where Masuda (2012) showed that strategy involving a stereotypic assessment of other individuals (based on their group membership) was evolutionary stable in a cooperation game.

On the other hand, sociological studies of professions revealed that the likelihood of formal prosecution of unethical behavior is increasing in the public visibility of offense (Abbott 1983). This finding is quite important in the context of this study where we observe subjects in mixed and in homogenous groups. If visibility is important, we can expect that our subjects are more concerned about ethical behavior in mixed groups (where the "offense" i.e. the lack of altruism is more visible) than in homogenous groups.

Because of the specific design used, our project is also related to literature on experimental public goods games (PGG). This literature is much too wide to be properly reviewed here, the reader is referred to Ledyard (1995), Zelmer (2003), and, especially, Chaudhuri (2011). It is a typical finding that most subjects reciprocate others' behaviour, Fischbacher et al. (2001), although to what extent they follow the best or the worst example or the average may depend on the specific design choices, Bigoni and Suetens (2012). In any case, partly because some group members consistently contribute nothing and others tend to follow suit at some point, cooperation levels typically decline over time (Neugebauer et al. 2009a, b; Fischbacher and Gächter 2010). One often-tested manipulation that helps prevent it involves peer punishment (Fehr and Gächter 2000). As this feature corresponds to a situation in which appropriate behaviours may only be enforced by informal, communal sanctions, not by the central authorities (for example due to imperfect monitoring/asymmetric information), it is a natural design choice for our purposes.

Given our interest in the role played by the visibility of offense, we run our experiments in groups that differ in composition. Hence, our project is also related to PGG experiments with *sorting*. One important reason why sorting can be expected

---

[3] There have been several instances of professions losing their self-regulatory powers. One example is the re-regulation of the American bar in the wake of the Clark report (Johnson 2005; Zacharias 2009). Another is the reorganization of the British medical profession following the scandals of the 1990s and early 2000s (Dixon-Woods et al. 2011).

to play a major role is that first impressions matter a lot: what other group members do in the first period determines contributions in subsequent periods (Engel et al. 2014). The bulk of studies on PGG with sorting investigated endogenous matching into groups based on previous behaviour, the participants being aware of the mechanism when making subsequent decisions. For example Gächter and Thöni (2005) first let their participants play one-period PGG and then match them into groups based on their decisions (high contributors with high contributors, low with low) or randomly. The typical finding is that matching based on previous behaviour triggers more cooperation (see Sect. 4.2.2. of Chaudhuri 2011).

Perhaps the most closely related to our study are those in which the groups are made more vs. less homogeneous in terms of some characteristics that are external to the game. Peters et al. (2004) and Molina et al. (2016) found that both adults and children contributed more to the public good when matched with their family members than when matched with strangers. Chakravarty and Fonseca (2014) investigated how group composition affected contributions to the public good in the context of minimal group paradigm—the subjects were divided into Klee-lovers and Kandinsky-lovers. They found that the contributions to the public good were highest for low—but non-zero—levels of diversity.

We are not aware of public good games that would compare contributions depending on the composition of the group in terms of field of study. As a matter of fact, it is not even equivocal if contributions differ by educational background. While some papers starting with the influential study by Marwell and Ames (1981) reported that economists contribute less in the PGG, the systematic meta-analysis of Zelmer (2003) does not confirm this finding. Moreover, none of the afore-mentioned studies comparing homogeneous and heterogeneous group allowed for punishment of free-riders. Whereas in the basic PGG it is natural to expect higher contribution in a homogenous group, it is an interesting empirical question whether the same will be the case in a PGG with punishment. Indeed, subjects may expect (and actually experience) little punishment in a homogeneous group, so that free riding becomes relatively attractive. We thus hypothesize that contributions will not be higher in homogeneous groups compared to the heterogeneous groups in our experiment.

3 Design, procedures, sample

The experiment involved a linear public goods game with punishment played in groups of four, with fixed matching, see "Appendix 1" for the translated instructions. In each of 10 periods participants were given 20 points and asked to distribute them between the Private Account and the Public Account. The points collected in the Public Account were multiplied by 1.5 and distributed equally among all group members, implying the Marginal per Capita Return of .375.[4] Once made, decisions about individual contributions to the Public Account were disclosed to other group members, who could assign up to three punishment points in total (per period).

---

[4] The MPCR was .5 in the first, pilot session.

Each punishment point meant deducting one point from the punisher and three points from the punished. The total number of punishment points received (but not the identity of the punishers) was revealed and the next period followed. At the end of the tenth period the sum total of points was calculated and paid out in cash at the rate of 1 point = .20 PLN (ca .05EUR).[5] Thus, for example, if all group members behaved selfishly (no contributions to the public account, no punishment), they would accumulate 10×20 points and earn 40PLN (ca. 10EUR) each.

Only individuals registered in the subject pool as representing one of three educational backgrounds: economics/management, medical sciences, law were invited to participate in this experiment. In the Homogenous Treatment (Homo) subjects were assigned into groups as homogeneous as possible and told that this was the case. In the Heterogeneous Treatment (Hetero) the matching was random and the participants were not told who the others could be.

In practice, economists, which are most easily recruited to the experiments, represented 60% of the 136 participants; a few individuals did not report belonging to any of the three groups invited and were labelled as "other". On average, the participants were 24.4 years old (median 22.5). About 70% of them were students. Some 56% were female. Nearly 22% had no siblings, 49% had one, 18% had two and 10% more than two.

In view of the focus of the paper it should be noted that even the students in our sample have already learned about professional ethics. All students have to complete a course on copyright and ethics of authorship at the beginning of their curriculum; students of medicine learn medical ethics in one of the first 3 years and students of law may choose a course in legal ethics (which they typically do in their second or third year). Certain basic tenets of professional ethics are mentioned in several other courses.

4 Results

4.1 Investment in the public account

While participants were allowed to distribute less than 20 points between the two accounts, they, predictably, hardly ever did that. There is thus no need to investigate investments in the Private Account and the Public Account separately; henceforth we will only talk about the latter and call it simply "investment". Figure 1 shows the distribution of individual investments by period, jointly for both treatments. As it is typical in PGG (and many other games), prominent (round) numbers (0, 10, 20 but also 5 and 15) are often picked. It is also clear that intermediate choices (5,10, 15) tend to drift towards 20 over time, except for the ultimate (and perhaps penultimate) period, in which some participants switch to investing nothing or almost nothing.

---

[5] To be precise, the experiment discussed in this paper was followed by another, unrelated experiment. Participants were ex ante informed that one of these would be selected to determine actual payments.
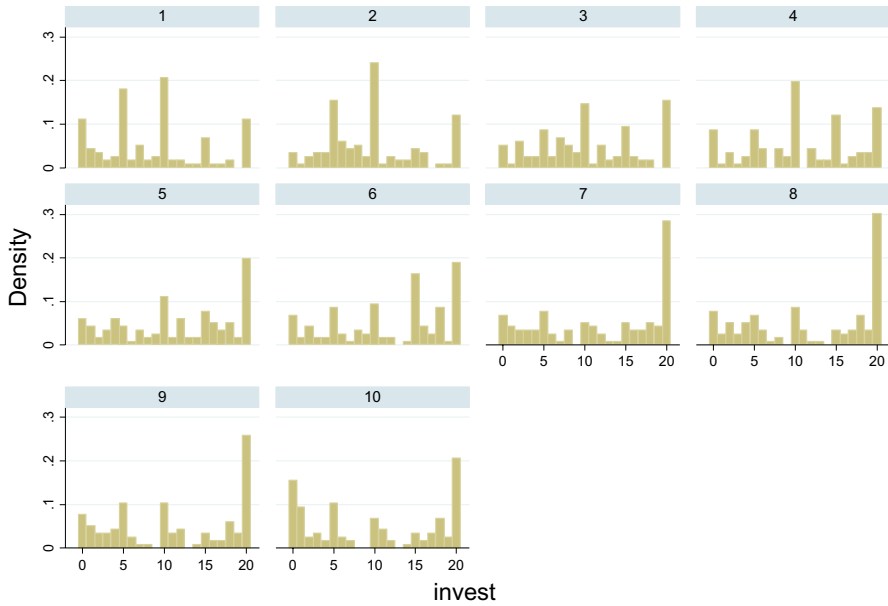
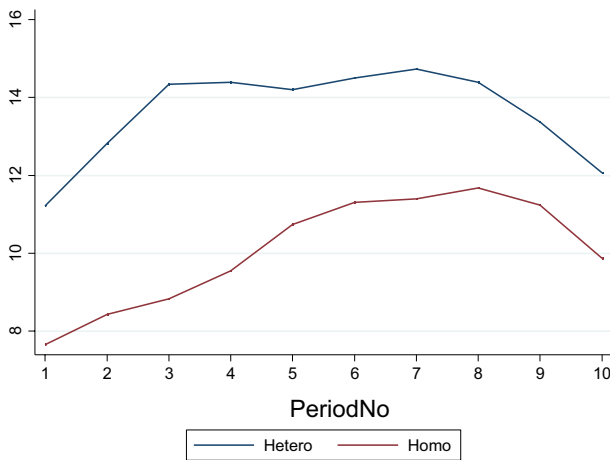**Fig. 1** Distribution of individual investments (0–20), by period

Fig. 2 Mean investment over time, by treatment

Figure 2 shows the trajectories of mean investment over time, by treatment (see "Appendix 2" for group-specific trajectories). It confirms the general tendency to increase investment over time, except for the last two periods. Comparing the two treatments, investment in the public account is lower when the participants know that they are matched with people of the same background.

	o1	*o2*	*o3*
homogeneous	− 0.591*	− 0.626*	− 0.667*
age		− 0.003	0.014
siblings		0.213	0.232
law			− 0.198
medical			0.043
other			− 0.473
graduate			− 0.233

Table 1 First-period investments: ordered logit

Economics as base category. See "Appendix 3" for definitions of variables

*Significant at 10%

When testing if this difference is statistically significant, one has to remember that individual investment decisions within a group are not independent. Two approaches are therefore possible. First, we can focus on the first period only—these decisions cannot be influenced by others, so each subject's choice is independent. Because the distribution is not normal as seen in Fig. 1, we use the non-parametric Mann–Whitney test and observe borderline significant ($p = .075$). A similar picture emerges when we control for other characteristics in an ordered logit regression, see Table 1.

Again, the treatment variable is marginally significant (p values close to .06). The only other variable that might affect investment is the number of siblings: predictably, those with more brothers and sisters invest a bit more in the public account (van Lange et al. 1997, although e.g. Knight and Kagan 1982 find no effect in PGG). However, this effect is not significant at conventional levels.

The second approach uses all the data. Because others' contributions typically (and also in our case) have a significant impact on own contribution, we can only safely assume that group means (not individuals means) are independent observations once we average across all the periods. Of course, this drastically reduces the number of observations and thus statistical power and we cannot reject the null hypothesis of identical mean investment under Homo and Hetero ($p = .34$).

One could of course envisage a panel regression model, hoping that the influence of others' decisions can be accurately captured by lagged variables clustering standard errors on group level. Running it on investment levels, however, would be inappropriate, because this variable is clearly non-stationary; for example, in more than 44% of the cases, participants simply invest the same amount as in the previous period. We thus run a model on changes from the previous period, which is not helpful in assessing central tendency by treatment and will therefore be reported later. Based on the tests for individual decisions made in the first period and group means across all periods, we can formulate:

Result 1 Homogenous groups invest at most as much, perhaps less, than heterogeneous groups in the Public Account.

Table 2 Punishment levels: Poisson regression

	p1	p2	p3	p4
homogeneous	− 0.309	− 0.237	− 0.148	− 0.019
law	0.284	0.301	0.116	0.143
medical	0.232	0.375	0.182	0.176
other	0.085	0.182	0.148	0.113
graduate	0.194	0.223	0.155	0.183
siblings	0.015	− 0.023	0.017	0.013
male	0.158	0.094	0.004	0.017
period	− 0.044	− 0.024	− 0.031	− 0.032
inv_min_diff		0.093***	0.093***	0.106***
inv_med_diff		− 0.001	− 0.012	− 0.024
inv_max_diff		− 0.024	− 0.024	− 0.022
lag_pun_given			0.239	0.230
lag_pun_received			0.067	0.063
lag_pun_received*homo			0.036	0.033
inv_min_diff*homo				− 0.035
inv_med_diff*homo				0.031
inv_max_diff*homo				− 0.008

Economics as base category. See "Appendix 3" for definitions of variables

***Significant at 1%

This tendency may be associated with limited trust in own profession. For all the three curricula we consider ambition and willingness to compete are required to thrive. This may explain why participants did not expect their colleagues to cooperate much and adjusted their own behaviour. Another possible reason for the difference, as speculated before, is that unethical, selfish behaviour was more broadly visible (visible to "others") in the heterogeneous condition.

4.2 Punishment behaviour

We now turn to the punishment behaviour. This time, already Period 1 behaviour is affected by other group members' investments. Again, we can calculate mean levels of punishment across all periods, aggregated at group level, and test if they differ across treatments. We cannot reject the null hypothesis of no difference ($p = 0.34$).

As tests indicate that this variable is stationary, we can also run a panel Poisson regression on the number of punishment points assigned by each individual, in a given period, clustering errors for groups, see Table 2.

Generally speaking, the variables considered fail to explain much of the punishment behaviour. The only (strongly) significant impact is that of the difference between own contribution and the lowest of contributions made in the relevant period by other group members (*inv_min_diff*): predictably, participants punished more if at least one other group member was relatively selfish. The differences

Table 3 Determinants of investment changes: regression analysis

	i1	i2	i3
homogeneous	0.262	0.104	− 0.024
lag_diff	− 0.390***	− 0.398***	− 0.395***
lag_invest	0.032*	0.030*	0.018
lag_pun_deserved	0.064***	0.062**	0.060*
period	− 0.258***	− 0.212***	− 0.206***
lag_pun_received		− 0.235	− 0.233
lag_pun*homo		− 0.067	− 0.095
lag_pun_des.*homo		0.047	0.053
last_period		− 0.677	− 0.692
law			− 0.537
medical			− 0.546
other			− 0.547
graduate			0.179
siblings			0.113
male			− 0.279

Economics as base category. See "Appendix 3" for definitions of variables

***Significant at 1%, **Significant at 5%, *Significant at 10%

between own investment and maximum or median of others do not seem to play a role. Likewise, there is little counter-punishment (punishment received in the previous period, *lag_pun_received*, has no impact). Importantly, there is no effect of treatment, neither directly nor in interaction with the differences in contributions.

The same picture emerges if we analyse each of the three punishment decisions made by each player in each period separately—again, participants generally punish those group members who contributed less than they did and treatment plays no role (we supress this model in view of space; note also that it must be treated with caution, because there was inevitably strong dependence between each player's three punishment decisions in a given period). Thus we establish:

Result 2 Patterns of punishment do not differ between the Homogeneous and Heterogeneous treatments.

4.3 Changes in investment levels

We now turn to investigating how past behaviour—investment and punishment—affected changes in investment levels. In view of findings reported in existing literature we allow for the possibility that participants try to align their behaviour with that of others. And that is what we indeed observe (see Table 3): the coefficient for *inv_diff* indicates that on average they try to close about 40% of the gap between own and others' lagged contribution. We also investigate the impact of punishment points received. It is plausible that punishment only encourages contributing more

if it is "deserved". We thus include the variable *lag_pun_deserved*, which is defined as the product of the number of punishment points received and the deviation of last period's investment from its global mean.[6] Thus, *lag_pun_deserved* is positive only if the participant has just been punished for indeed contributing less than others normally contribute. As can be found in Table 3, this variable has a significant impact in the predicted direction. Again, the treatment manipulation does not seem to make a difference, neither directly, nor in interaction with other variables.

5 Discussion and conclusions

Professions play a key role in the provision of several types of socially important services. Because these are typically complex, implying a high degree of information asymmetry, a serious risk of abuse of trust arises, calling for high ethical standards taught and enforced by the profession. In this study we employed experimental methodology to investigate one aspect of ethical behaviour of professionals, namely, whether, controlling for educational choices and the causal impact of the training itself, the sense of being in a community of professionals makes unethical choices less attractive or more punishable. In doing so we attempted to contribute to two literatures: on professional groups and on experimental public good games.

While a number of our findings are natural and consistent with prior studies (e.g. punishment induces a more ethical behaviour), our main result seems unintuitive: in general, we do not find that acting within a professional group contains free-riding; if anything, our subjects free ride more when matched with their peers than in a mixed group. Albeit this is somewhat speculative, we may offer a few potential explanations. One pertains to the afore-mentioned notion of 'visibility of offence' as the factor improving professions' ethical standards. Another possibility is that subjects had pessimistic expectations about their kind and acted accordingly themselves. They could have also expected less punishment within a homogeneous group. As numerous studies show that peer punishment is an effective mechanism against drop of cooperation in PGGs, any factor that reduces its actual or anticipated level may be detrimental to investments in the public account. This explanation is consistent with all our basic findings: because less punishment may have been expected in homo, first-period investments were a bit lower in this condition. However, as in practice there was just as much punishment under Homo as under Hetero, investments caught up and were not significantly different aggregating over all periods.

This unexpected finding may also open up an interesting line of research within the literature on public good games. While most studies so far found that more homogeneous groups contributed more, it turns out this finding is perhaps not robust to the introduction of punishment. Put differently, the possibility of punishment may be much more important in sustaining cooperation in heterogeneous than in homogenous groups.

---

[6] Considering the deviation from this particular group- and period-specific mean leads to analogous results.

Generally speaking, our results suggest that professional ethos per se does not harness selfishness. This finding, if confirmed in future empirical work, could be potentially important in the discussion about the legal regime of professions, where it has been suggested that systems such as co-regulation or competitive self-regulation would be favourable to traditional self-regulation (cf. Grabosky and Braithwaite 1986; Van den Bergh 2004).

Of course, extrapolating findings on students of any specific field to professionals may only be done with caution. Still, using such "surrogates" is an attractive and popular approach, because student subjects, unlike professionals, are readily and cheaply available, although they often have some work experience already and may join the profession soon. As a result, studies typically find small differences between behaviour of students and that of professionals in related fields (see Frechette 2011; Liyanarachchi 2007; Mortensen et al. 2012), providing justification for this approach.

Appendix 1: Instructions

In this experiment you will be making decisions, which will affect your and others people's earnings. Therefore please make these decisions carefully. Note, however, that there are no "correct" and "incorrect" answers here; we are interested in your own preference.

The participants of the experiment will be matched into groups of four. You will not know who exactly is in your group. You will stay in the same group throughout the experiment.

During the experiment you will earn "points". At the end of the experiment each point will be worth 20 grosze [.2PLN].

The experiment will consist of ten periods. Each period will consist of two phases: the Allocation Phase and the Punishment Phase.

The allocation phase

At the beginning of the Distribution Phase each of the four group members receives 20 points and must decide, independently of others, how to distribute them between the "Individual Account" and the "Public Account". As the name suggests, the points allocated to the Individual Account will simply be kept by the participants in question. The sum total of points allocated to the Public Account by all group members will be multiplied by 1.5 and split equally among all group members. That means that whenever you transfer 1PLN to the Public Account, you get 37.5grosze out of it (while others get 112.5 grosze). We will further explain it using some examples (beware: these examples should not be treated as a hint as to what could be a

reasonable or appropriate decision; they are only meant to help us explain the rules; in particular, we mostly use round numbers only to make them simpler).

Example 1 All the participants allocate all the points to their Individual Accounts. Thus each of them has 20 points (the Public Account is empty).

Example 2 Two participants allocate all the points to the Public Account and the other two allocate all the points to their Individual Accounts. There is thus $2 \times 0 + 2 \times 20 = 40$ points in the Public Account, or 60 points after we multiply by 1.5. Each participant gets $60/4 = 15$ points from the Public Account. Those participants who allocated everything to the Public Account thus end up with 15 points each. Those who allocated everything to their Individual Accounts additionally have 20 points they kept, thus they have 35 points each.

Example 3 Participant 1 (P1) allocates 5 points to the Public Account, U2: 12, U3: 7, U4: 0. The remaining point are allocated to the Individual Accounts. In total there is thus $5 + 12 + 7 + 0 = 24$ points in the Public Account, or 36 after the multiplication. Thus each participant gets 36/4-9 from the Public Account. In total, U1 has $15 + 9 = 24$ points, U2: $8 + 9 = 17$, U3: $13 + 9 = 22$, U4: $20 + 9 = 29$.

Remark 1 You are allowed to allocate less than 20 points if you really want to, e.g. allocate 7 points to the Individual Account and 8 points to the Public Account (thus leave 5 points unallocated). This, however does not seem a reasonable thing to do.

When all group members are done with their allocation decision, they proceed to the Punishment Phase.

The punishment phase

Each participant finds out how much the other group members have just allocated to the Public Account. Remember, everyone remains anonymous. Each participant then decides as to the allocation of PUNISHMENT Points to other group members. In total you cannot allocate more than three. Each Punishment Point you allocate costs you one point. Each Punishment Point that you *receive* costs you three points.

Example 3a P1 finds out that while he allocated 5 points to the Public Account, others allocated U2: 12, U3: 7, U4: 0. He decides not to allocate any Punishment Points. At the same time other group members are asked to make analogous decisions. They also do not use any Punishment Points. The final earnings for this period remain thus as given in Example 3, that is U1: 24, U2: 17, U3: 22, U4: 29.

Example 3b Let us now assume that the Allocation Phase followed as before, but then P2 allocated one Punishment Point to P4, P3 allocated two Punishment Points to P4 and others did not use any. Thus the earnings of P1 remains unchanged (24

points), One point is deducted from the earnings of P2 (corresponding to one Punishment Point she allocated), thus she has 16 points in the end. P3's earnings is reduced by two because of the two Punishment Points she allocated, making 20. Finally, P4 received three Punishment points (two from P3 and one from P2), which reduce his earnings by $3 \times 3 = 9$ points, leaving him with 20 points.

Example 3c Let us now assume that the Allocation Phase followed as described before and then P2 allocated one Punishment Point to each of the three groupmates. U2 loses 3×1 points, others lose 1×3 points, so the final earnings are U1: 21, U2: 14, U3: 19, U4: 26.

Remark 2 As you can see from these examples, you do not have to use up all Punishment Points. You are free to give none, for example.

At the end of each period you will find out how many Punishment Points you received in total. You will NOT know who exactly allocated how many of them. For example, you may find out that you got 4 Punishment Points and not that P1 gave one and P2 gave three. Neither will you know how other participants punished yet other participants. The computer will also calculate you total number of points for this period. You will see how it follows from the value of the Public Account, your Individual Account and the number of Punishment Points given and received, in accordance with the rules explained before. When all the group members learn that, the subsequent period will follow.

The same participant that is called P1 in one period will also be called P1 in any other period and likewise for other participants.

At the end of period 10 the computer will sum up all your points from the 10 periods. Again, each point will be worth 20 grosze. For example, if you collect 150 points over the ten periods, will make 30 PLN.

Remark 3 On all decisions screens you will see a clock counting down (starting at 2 min). It will be indicative only: when this time is gone, you will still be able to make a decision. Do not worry too much about it therefore. In particular in the early periods making a careful decision might take a few moments. However, both the experimenters and the participants would like the experiment to proceed smoothly. Thus, to the extent possible, do not take an excessive amount of much time to decide. In particular, if it is systematically you that all the group members wait for (you will realize that if you never see the notice "Please wait for others' decisions…" on your screen)—please try to speed up a bit.

By contrast, the result screens, on which you do not have to make any decisions, will disappear by themselves after 1 min (but we encourage that you click OK as soon as you are ready to follow).

Please re-read the relevant passage or raise your hand if you have any doubts. If everything is clear, please wait for the experiment to begin.
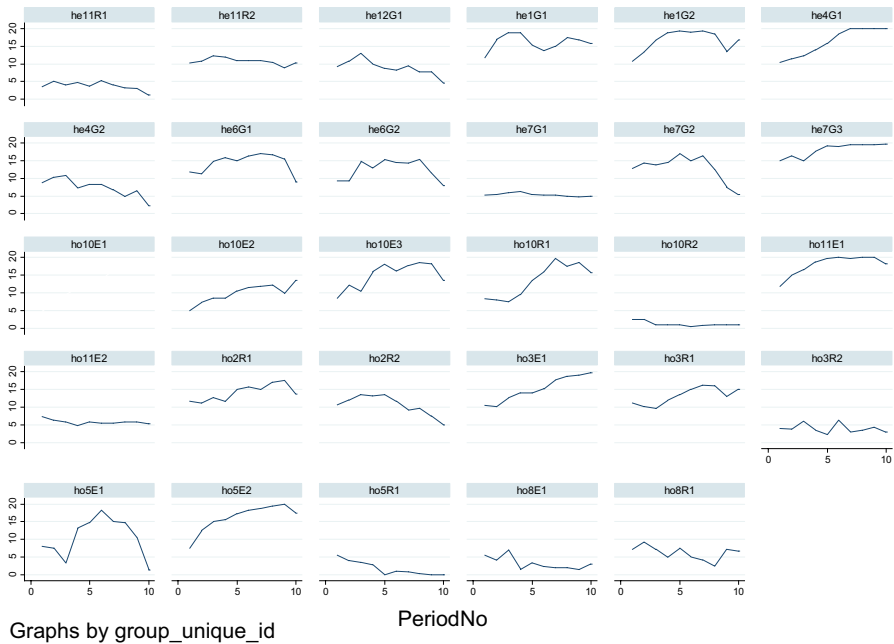
Appendix 2: Group-specific investments

See Fig. 3.

Fig. 3 Trajectories of mean investment, by group

Appendix 3: Description of variables

homogenous	homogenous group
X*homo	takes the value of X in homogeneous groups and 0 otherwise
lag_diff	the difference between one's own and others' mean average investments in the previous period
lag_invest	one's own investment in the previous period
lag_pun_received	punishment received in the previous period
lag_pun_deserved	the product *lag_pun_received* and the difference between one's own investment and mean investment in all rounds and groups
period	period number
last_period	dummy variable indicating the last period
law, medical, other	academic majors
graduate	dummy variable indicating having completed at least bachelor's program
siblings	number of siblings

male	dummy for males
inv_min_diff	the difference between one's own contribution and the lowest of contributions made in the same period by other group members
inv_med_diff	the difference between one's own contribution and the median of contributions made in the same period by other group members
inv_max_diff	the difference between one's own contribution and the highest of contributions made in the same period by other group members
lag_pun_given	the number of punishment points given in the previous period

References

Abbott, A. (1983). Professional ethics. *American Journal of Sociology*, *88*(5), 855–885.

Abbott, A. (2014). *The system of professions: An essay on the division of expert labor*. Chicago: University of Chicago Press.

Baron, D. P. (2010). Morally motivated self-regulation. *American Economic Review, 100*(4), 1299–1329.

Baron, D. P. (2011). Credence attributes, voluntary organizations, and social pressure. *Journal of Public Economics, 95*(11–12), 1331–1338.

Bigoni, M., & Suetens, S. (2012). Feedback and dynamics in public good experiments. *Journal of Economic Behavior & Organization, 82*(1), 86–95.

Brekke, K. A., & Nyborg, K. (2010). Selfish bakers, caring nurses? A model of work motivation. *Journal of Economic Behavior & Organization, 75*(3), 377–394.

Chakravarty, S., & Fonseca, M. A. (2014). The effect of social fragmentation on public good provision: An experimental study. *Journal of Behavioral and Experimental Economics, 53*, 1–9.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics, 14*(1), 47–83.

DeMarzo, P. M., Fishman, M. J., & Hagerty, K. M. (2005). Self-regulation and government oversight. *Review of Economic Studies, 72*(3), 687–706.

Dixon-Woods, M., Yeung, K., & Bosk, C. L. (2011). Why is UK medicine no longer a self-regulating profession? The role of scandals involving "bad apple" doctors. *Social Science and Medicine, 73*(10), 1452–1459.

Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature, XLVI*, 5–42.

Engel, Ch., Beckenkamp, M., Glöckner, A., Irlenbusch, B., Hennig-Schmidt, H., Kube, S., et al. (2014). First impressions are more important than early intervention: Qualifying broken windows theory in the lab. *International Review of Law and Economics, 37*, 126–136.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review, 100*(1), 541–556.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397–404.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives, 7*(2), 159–171.

Fréchette, G. R. (2011). Laboratory experiments: Professionals versus students. Available at SSRN 1939219.

Gächter, S., & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, *3*(2–3), 303–314.

Gehrig, T., & Jost, P. J. (1995). Quacks, lemons, and self regulation: A welfare analysis. *Journal of Regulatory Economics, 7*(3), 309–325.

Grabosky, P. N., & Braithwaite, J. (1986). *Of mannners gentle: Enforcement strategies of Australian business regulatory agencies*. Oxford: Oxford University Press.

Grajzl, P., & Baniak, A. (2009). Industry self-regulation, subversion of public institutions, and social control of torts. *International Review of Law and Economics*, *29*(4), 360–374.

Handy, F., & Katz, E. (1998). The wage differential between nonprofit institutions and corporations: Getting more by paying less? *Journal of Comparative Economics, 26*(2), 246–261.

Heyes, A. G. (2005). A signaling motive for self-regulation in the shadow of coercion. *Journal of Economics and Business, 57*(3), 238–246.

Johnson, V. (2005). Justice Tom C. Clark's legacy in the field of legal ethics. *Journal of the Legal Profession, 29,* 2005.

Knight, G. P., & Kagan, S. (1982). Siblings, birth order, and cooperative-competitive social behavior: A comparison of Anglo-American and Mexican-American children. *Journal of Cross-Cultural Psychology, 13*(2), 239–249.

Larson, M. S. (2013). *The rise of professionalism: Monopolies of competence and sheltered markets* (2nd ed.). Piscataway: Transaction Publishers.

Ledyard, J. (1995). Public goods: A survey of experimental research. In J. Kagel & A. Roth (Eds.), *Handbook of experimental economics*. Princeton: University Press.

Leland, H. E. (1979). Quacks, lemons, and licensing: A theory of minimum quality standards. *Journal of Political Economy, 87*(6), 1328–1346.

Liyanarachchi, G. A. (2007). Feasibility of using student subjects in accounting experiments: A review. *Pacific Accounting Review, 19*(1), 47–67.

Marwell, G., & Ames, R. E. (1981). Economists free ride, does anyone else?: Experiments on the provision of public goods. *Journal of Public Economics, 15*(3), 295–310.

Masuda, N. (2012). Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation. *Journal of Theoretical Biology, 311,* 8–18.

Maxwell, J. W., Lyon, T. P., & Hackett, S. C. (2000). Self-regulation and social welfare: The political economy of corporate environmentalism. *The Journal of Law and Economics, 43*(2), 583–618.

Molina, J. A., Ferrer, A., Giménez, J. I., Gracia-Lazaro, C., Moreno, Y., & Sanchez, A. (2016). *The effect of kinship on intergenerational cooperation: A lab experiment with three generations*. Working paper.

Mortensen, T., Fisher, R., & Wines, G. (2012). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum*, *36*(4), 251–265.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy, 78*(2), 311–329.

Neugebauer, T., Perote, J., Schmidt, U., & Loos, M. (2009). Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology, 30*(1), 52–60.

Ogus, A. (1995). Rethinking self-regulation. *Oxford Journal of Legal Studies, 15*(1), 97–108.

Ogus, A. (2000). Self-regulation. *Encyclopedia of Law and Economics, 5*, 587–602.

Peters, H. E., Ünür, A. S., Clark, J., & Schulze, W. D. (2004). Free-riding and the provision of public goods in the family: A laboratory experiment. *International Economic Review, 45*(1), 283–299.

Philipsen, N. J., & Faure, M. G. (2002). The regulation of pharmacists in Belgium and the Netherlands: In the public or private interest? *Journal of Consumer Policy, 25*(2), 155–201.

Van den Bergh, R. (1999). Self-regulation of the medical and legal professions: Remaining barriers to competition and EC law. In B. Bortolotti & G. Fiorentini (Eds.), *Organized interests and self-regulation: An economic approach* (pp. 89–130). Oxford: Oxford University Press.

Van den Bergh, R. (2004). Towards efficient self-regulation in markets for professional services. *European Competition Law Annual, 2006,* 155.

Van Lange, P. A., De Bruin, E., Otten, W., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*(4), 733.

Zacharias, F. C. (2009). The myth of self-regulation. *Minnesota Law Review, 93,* 1147–1190.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics, 6*(3), 299–310.