Overcoming the Tragedy of Personnel Evaluation?

Momme von Sydow[1,2] (momme.von-sydow@urz.uni-münchen.de)
Niels Braus[2] (n.braus@stud.uni-heidelberg.de)
Ulrike Hahn[1,3] (u.hahn@bbk.ac.uk)
[1]University of Munich (LMU), Munich Center for Mathematical Philosophy (MCMP),
Ludwigstr. 31, D-80539 München, Germany
[2]University of Heidelberg, Department of Psychology, Hauptstr. 47-51, D-69117 Heidelberg, Germany
[3]Birkbeck College, University of London, Malet Street, London WC1E 7HX, U.K.

Abstract

Human beings are essentially – by nature or second nature - members of groups. They contribute to these groups not just as isolated individuals but also through their interaction with others. Consequently, personnel evaluation in companies and organizations requires assessing not only evaluating individual performance but also the overall direct and indirect effect one has on a team. Others' work may be improved or hampered by the presence of a particular employee. We investigate Two-level Personnel-Evaluation Tasks (T-PETs) with information on individual and group earnings, where an individual focus may lead to evaluate the overall best employee as being the worst. We have previously found a Tragedy of Personnel Evaluation where focus on direct individual impact did have such systematic effect. In two experiments, one on team size, the other on kinds of information provided, we explore the boundary conditions of this effect and suggest how it may be overcome.

Keywords: Tragedy of Personnel Evaluation; Rationality of Personnel Decisions; Inner-Individual Dilemma; Social Psychology; Personnel Evaluation; Personnel Selection; Bounded Decision Making; Causal Induction

Introduction

The success of teams in organizations or companies not only relies on the direct performance of individuals, but often also on interactions between team members (Mathieu, Maynard, Rapp, & Gilson, 2008; Memmert, Plessner, Hüttermann, Froese, Peterhänsel, & Unkelbach, 2015). Individuals may, for instance, help or hinder each other. The vital role of prosocial or altruistic behaviours for teams in organisations and companies (George & Bettenhausen, 1990; Li, Kirkman, & Porter, 2014; Nielsen, Hrivnak, & Shaw, 2009; Organ, 1997; Podsakoff, Whiting, Podsakoff, & Mishra, 2010) and for functioning societies on the whole is being increasingly acknowledged (Engel, 2011; Hendrich et al., 2005; Gollwitzer, Rothmund, Pfeiffer, & Ensenbach, 2009; Post, 2005; cf. Melis et al., 2016). In Organizational Psychology several types of contextual performance (Organ, 1997; van Scotter & Motowidlo, 1996) and prosocial behavior (Brief & Motowidlo, 1986; Li, Kirkman, & Porter, 2014) have been distinguished. Researchers have also pointed out that not only is prosocial behavior crucial for the success of organizations, but that people are actually sometimes rewarded for it (Organ, 1997; Scotter, Cross, & Motowidlo, 2000; Grant & Patil, 2012, 562).

In previous work, we began investigating participants' behaviour as hypothetical human-resource managers evaluating employees working in different configurations each shift (von Sydow & Braus, 2016). We employed Two-level Personnel-Evaluation Tasks (T-PETs) that, across several rounds ('shifts'), provide information on how both individuals and teams contribute to a store's earnings. Crucially, the individual and team information suggest opposite rankings of the employees' contributions. By design, the presence of a so-called 'altruist', someone who positively affects the performance of the others, was most positively correlated with the overall team performance (r = .99), even though the altruist individually performed the worst. We focused on the example of *one* employee strongly affecting the *whole* group, as this case is influential in biological models of altruism that assume an unconditional advantage to all group members in the presence of an altruist (Sober & Wilson, 1999; Wilson & Wilson, 2007; but Nowak & Sigmund, 2005). The participants' task involved evaluating employees (Personnel Evaluation) and selecting the best team (Personnel Selection). Participants saw only *one* group (one shop) and the teams were assembled by selecting 4 out of 5 employees (thus 5 team configurations were possible).

Results from von Sydow & Braus (2016) suggest what they called a "Tragedy of Personnel Selection". After 40 shifts, repeated measurement, and no time-constraint for analyzing the data of a shift, participants systematically judged the overall best employee to be the worst (on the different hidden-profile problem, Mojzisch, Grouneva, & Schulz-Hardt, 2010). Recently we also explored negative interactions (egoist detection) in T-PETs (subm.). This tragedy is reminiscent of the well-known "Tragedy of the Commons", a notion for the often tragic outcomes of social dilemma situations (such as public-good games). Note however, that the 'T-PETs' do not strictly involve social dilemma, as the participant manager has the explicit goal of choosing the best team for the company. It is only what has been called an *inner-individual* dilemma (von Sydow, 2015) between two levels of goal descriptions, since it is irrational to optimize more specific goals at the expense of lowering overall utility. Since positive (and negative) interactions with other employees are ubiquitous, and number-based evaluations are important in HR-management, these results suggest such tragedy may well be found in everyday life.

Here we present two new experiments exploring the generalizability or boundary-conditions of the tragedy. Experiment 1 varies team sizes and begins exploring the idea of multiple groups in parallel. Experiment 2 investigates longer learning periods and whether with forced focus on the group level people are able to detect the altruist.

Experiment 1

Design

In the T-PETs we provided information on both individual and overall group level earnings. Experiment 1 investigates the extent to which the Tragedy of Personnel Evaluation depends on group size (number of workers: 3, 4, 5 versus 7) and the number of groups (one versus two) (Figure 1, Table 1). From Conditions 1 to 4, group size increases. In a shift, all workers apart from one are working. Condition 5 has the same group size as Condition 4 but is characterized by a group-comparison scenario, where the 6 employees are split in two groups with three employees each.

Table 1: Numbers of workers and groups, and their mean earnings (normal workers, NW; altruist, A) in the five conditions.

	C1	C2	C3	C4	C5
			Number		
Groups	1	1	1	1	2
All Workers	3	4	5	7	7
Shown Workers	2	3	4	6	3/3
			Mean Earnings (€)		
NW with A	3000	3000	3000	3000	3000
NW without A	2000	2000	2000	2000	2000
Altruist (A)	1500	1500	1500	1500	1500
Group with A	4500	7500	10500	16500	7500/ 6000
Group without A	4000	6000	8000	10000	6000/ 6000

Method

Participants 221 Participants from the US began the experiment via MTURK. 158 participants finished it, passing all selection criteria (time spent on first page and correct answer out of four, rephrasing the instructions). The participants obtained a reward of 2$. 46% were male, the mean age was 35 years; 53% mentioned having a Bachelor's or Master's degree, and 39% a high school degree as highest level of education. Participants were randomly assigned to one of the five conditions.

Procedure and material The payoff structure for the individual employees remains constant over the five conditions, whereas the differing group sizes led to differing earnings of the groups (Table 1). We adopted the repeated measurement design with 18 shifts for each of the 4 rounds, and at the end of a round a rating followed by a selection task. The total number of shifts was 72. (This number is higher than the 40 trials investigated in previous studies.) In C3, for instance, the selection task was to select a team of 4 from 5 available workers that would be best for the company (Figure 1). In C5 people could select 6 workers from 7 for both groups. In the last round, we additionally asked for the employee with the greatest and lowest utility, and assigned a Need-For-Cognition-Task (Cacioppo, Petty, & Kao, 1984), a working memory task, a Commentary, and demographics.

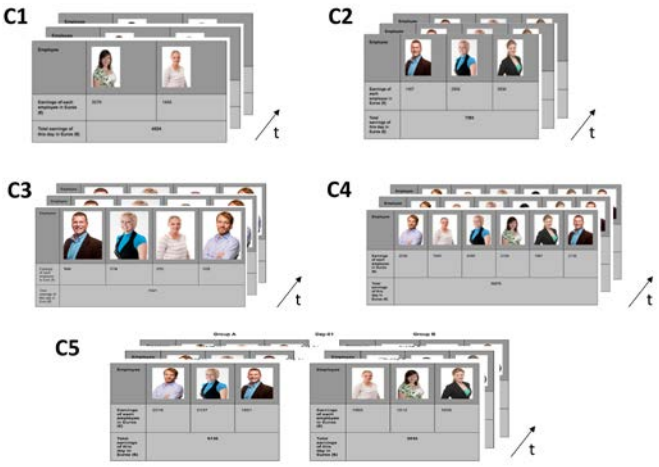

Figure 1: Illustration of the materials. In the T-PETs the overview information for each round contains, in the first row the photos of the employee (in random order); in the second row their individual earnings; and in the third row the overall group/team earnings.

Results

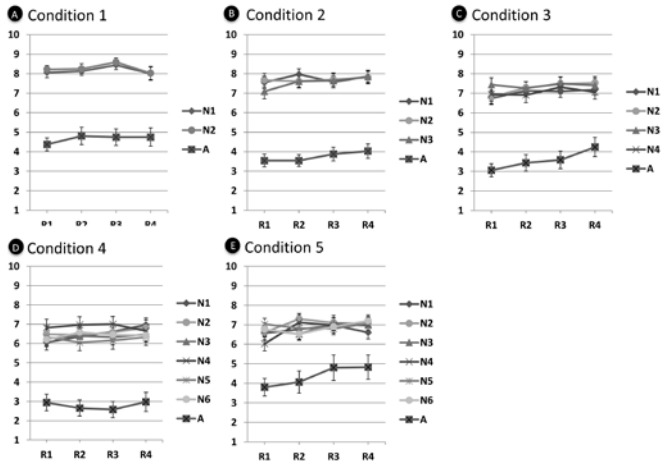

Figure 2: Average ratings (with SE) for the normal workers (*N1-N6)* and the altruist worker (*A*) of Conditions 1, 2, 3, 4 and 5 (Panels A to E).

Figure 2 shows that the rating for the altruist in all conditions remained clearly below the other ratings. In an ANOVA of the altruist ratings only, Condition is a significant between-subject factor ($F(4,153) = 3.57$, $p < .01$) and Phase (the four test phases) a significant within-subject factor (Pillai-Spur-Test, PST, $F(3, 151) = 3.92$, $p < .05$), with no further significant interaction effects ($p = .32$). This outcome seems in line with the prediction that one obtains the best results for the condition with the fewest workers

(C1), despite the altruist increased overall effect on the mean group earnings in larger groups, and for the group condition (C5). Bonferroni-corrected post-hoc comparisons show significant differences between C1 and C4 ($p < .01$) and between C4 and C5 ($p < .05$). However, the ratings remain predominantly based on individual comparisons in all conditions.

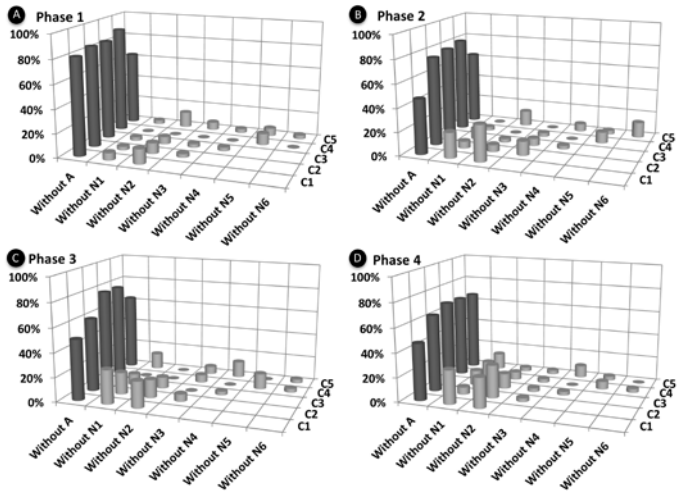

Figure 3: The proportion of 'managers' choosing a team and excluding the altruist (*A*) or a normal worker (*N*) in the test phases (Panel A to D) of the personnel selection task. The dark shading represent selections based on individual earnings, the light ones correspond to overall earnings.

Figure 3 shows that in all conditions the largest proportion of participants tends to expulse the overall most useful worker, the altruist, from the team. This is clearly the case in Phase 1 (always, $p < .01$). The proportion of other choices overall increased over time ($\chi^2(1, 316) = 13.15$, $p < .001$). Nonetheless altruist expulsion remained statistically above chance in Phase 4 (apart from C1, $p = .10$; all other $p < .01$). For the variable which worker is deemed to have the least utility for the company (not presented here), in Round 4, the choice of the altruist even remained dominant for all conditions ($p < .01$).

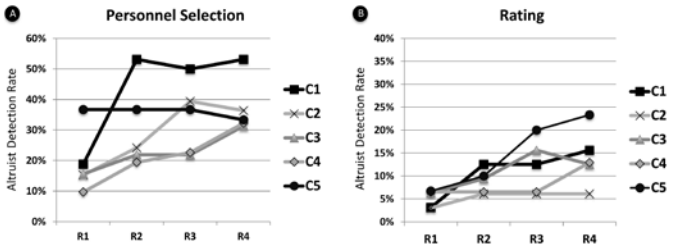

Figure 4: Average altruist-detection rate for Conditions 1, 2, 3, 4, 5 in the personnel selection task (Panel A) and the rating task (proportion of altruist rated larger than *all* other workers; Panel B).

Figure 4A shows the increase of group-level answers particularly for only a few employees (C1). The group condition C5 has a relatively high start but does not increase. However, Figure 4B suggests that there is also an increase for C5 if one considers the stricter criterion of rating the altruist to be higher than *all* other workers.

Figure 5A shows at least descriptively that in the highest utility task the altruist is positively singled out relatively frequently in C1 and C5 (but note the different numbers of workers). Additionally, participants' comments were deemed 'insightful' if they detected possible differences between an individual's direct and overall earnings (Figure 5B). Note that this measure is not directly affected by the number of answer-options. There was a reliably higher number of insightful comments in C1 than C2, C3 and C4, but not higher than in C5.

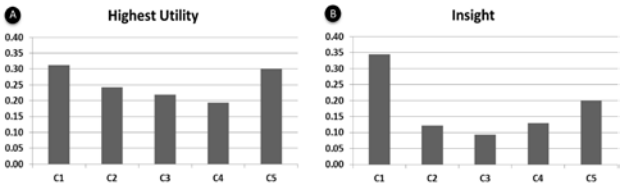

Figure 5: A) Proportion of Participants selecting the altruist to be of highest utility. B) Insight rate shown in Comments.

Overall, Experiment 1 shows that the tragedy is quite stable over group size. However, it also suggests that although the altruist's summative effect increased with the group size, the best participant performance was in the condition with the lowest number of employees (C1). The results also only showed subtle advantages of introducing different groups on the altruist detection rate (C4 vs C5).

Experiment 2

Experiment 2 investigates conditions where people are forced to focus on the group level only, to see whether all participants realize that the altruist performs better on this level. We thus investigate whether the tragedy of personnel evaluation is due to an inability to see complex effects on a group level (despite information concerning this level). Additionally, and in contrast to Experiment 1, we distinguish different mean individual earnings of normal workers to check the extent to which people distinguish even slighter performance differences on the individual level.

Design

The experiment had a mixed 2 (information: global-only versus local-and-global) × 2 (earnings of normal workers: homogeneous versus heterogeneous) between-subjects design, with a within-subjects factor of four test phases (Table 2). In each test phase both evaluation and personnel selection tasks were assigned. Additionally, in the last round, highest and lowest utility tasks and other tests were completed, as well as a Need-For-Cognition (NFC) test.

In the local-and-global conditions, participants were provided with almost the same overview information as in Condition 3 of Experiment 1 (only the Altruist individually contributed 1600 instead of 1600). In each round,

information was given on both the direct earnings of the four workers on a shift and the overall earnings of the shift. The overall earnings involved not only the direct effects of individuals but also their indirect effects. In the global-only conditions, only the *overall* payoffs of a group (shift) were presented, without showing individual contributions.

Table 2: The four conditions, also showing the overall versus direct impact of a worker on group-earnings

Condition	C1	C2	C3	C4
Information	Local and global	Global only	Local and global	Global only
Earnings NW	Homogeneous		Heterogeneous	
Overall impact	A >> NW1 = NW2 = NW3 = NW4		A >> NW1 > NW2 > NW3 > NW4	
Direct impact	NW1 = NW2 = NW3 = NW4 > A		NW1 > NW2 > NW3 > NW4 > A	

Note: NW = normal worker; A = altruist.

The homogeneous and heterogeneous conditions correspond to either identical or different individual impact of the normal workers (see Table 2). The group earnings remained identical in both kinds of conditions. The 'altruist' (*A*) always has the most positive impact on the overall earnings. NW earnings (€) without *A* were 2000 (homogeneous); 1400, 1800, 2200, 2600 (heterogenous); with *A*, 3000 (homogeneous); 2400, 2800, 3200, 3600 (heterogeneous), but the altruist had the lowest direct (individual) impact, 1600.

Method

Participants As in Experiment 1, relatively strict selection criteria for participants were used to ensure high data quality. After passing a first criterion (time spent on the first page), 150 people properly started the task and 7 people failed the second criterion (correct rephrasing of the task; four options). Of the remaining 143 volunteers, 122 finished the experiment, and only their data was analysed. Participants were recruited from MTURK: 57% were male, 42% female; mean age was 33, and 68% had a Bachelor's or Master's degree (with 32% a high school degree). They received $2 for participation.

Procedure and material We used almost the identical materials and procedure as in Experiment 1, C3. The experiment had 80 rounds, with four test phases administered after Rounds 20, 40, 60, and 80. In all four test phases, participants completed both a personnel-evaluation task and a personnel-selection task. In the final test phase, we additionally administered a highest-/lowest-utility task, a ranking task, a Kimchi-Palmer item, an attention-test item, and an 18-item Need-For-Cognition Test (Cacioppo, Petty, & Kao, 1984).

In the global-and-local conditions, the overview information presented in each round corresponds to C3 in Figure 1. In the global-only conditions, the second line of this panel (presenting the individual earnings of each employee) was omitted.

Results

Figure 10 shows the mean ratings for the workers' contributions to company earnings. An overall ANOVA with Workers (5 workers) and Phases (4 phases) as within-subjects factors, and Conditions as between-subject factor, yielded a highly significant effect of Conditions × Workers (Pillai-Spur Test, PST, $F(12, 306) = 22.5$, $p < .001$). This corresponds to the predicted change of rank of the altruist's ratings in the global-only versus global-and-local conditions. Additionally, the factors Workers, Phase × Condition as well as Phase × Worker approached significance (PST, $F(4, 100) = 17.3$, $p < .001$; PST, $F(9, 309) = 1.69$, $p = .09$; PST, $F(12, 92) = 1.51$, $p = .13$). Changes over the phases were not significant.

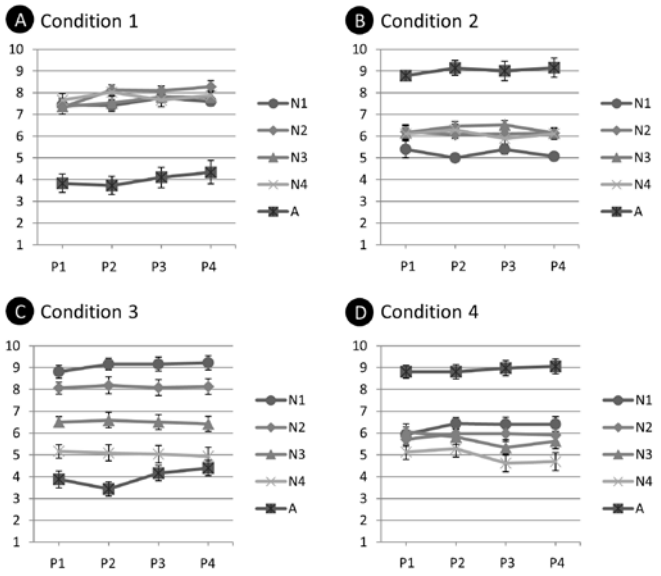

Figure 6. Average ratings (with SE) in Experiment 4 for the four normal (*N*) and altruist (*A*) workers in test phases P1 to P4 of Conditions 1, 2, 3, and 4 (Panels A to D).

In the homogeneous local-and-global condition (C1, Panel A), the altruist was again evaluated as the worst, despite being most strongly correlated with high overall earnings. In an ANOVA for Phase 4, the within-subject factor Workers was clearly significant (PST, $F(4, 23) = 10.1$, $p < .001$), and contrasts confirmed that all normal workers were rated higher than the altruist (all, $p < .05$). In the heterogeneous Condition 3, participants were well able to differentiate between normal workers with different individual performance. A corresponding ANOVA showed a general effect of Workers (PST, $F(4, 24) = 38.8$, $p < .001$) and significant contrasts between the workers in the predicted order, N1 > N2 > N3 > N4 > A (each, $p < .001$).

In the global-only conditions (C2 and C4), in which people were to base their ratings of a worker's utility on the teams' overall earnings only, they clearly detected that, of all workers, the altruist correlated most demonstrably with high overall team earnings. Participants grasped this surprisingly early. An ANOVA for C2 (test phase 4) shows

significant results for the factor Workers (PST, $F(4, 22) = 23.04$, $p < .001$), and pairwise contrasts show that the altruist is rated higher than all normal workers (always with $p < .001$). In Condition 4, the order of the average ratings of the altruist and the normal workers was likewise reversed (relative to Condition 3). In an ANOVA a significant effect of the factor Workers was found (PST, $F(4, 26) = 15.5$, $p < .001$); and contrasts show that the altruist was rated significantly higher than even the normal worker, who was rated highest ($p < .001$). One ANOVA without the altruist reached significance (PST, $F(4, 27) = 3.62$, $p < .05$), but only one Bonferroni-corrected post hoc comparison between normal workers (the one expected to differ most: NW1-NW4) led to significant results ($p < .05$). In sum, despite clearly detecting that the altruist has a larger effect on overall output in the global-only conditions, participants show a reduced ability to distinguish between the normal workers.

Figure 6 shows the proportion of 'managers' choosing a worker to have the "highest" (Panel A) or "lowest" (Panel B) "total utility for the company" in the final test phase.

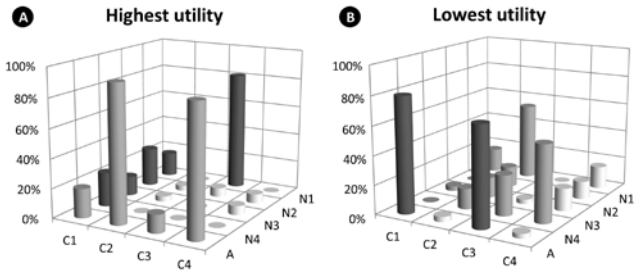

Figure 7. Percentage of participants choosing a normal worker (*N*) or the altruist (*A*) as of highest (Panel A) or lowest (Panel B) overall utility for the company in Phase 4 in Conditions C1, C2, C3, and C4. The choices corresponding to individual earnings are marked in black; those corresponding to overall earnings in dark gray.

Figure 7 presents the team selections in the personnel selection task. In Condition 1, we replicated a strong tendency to select a team *without* the overall best member, the altruist (from five possible configurations). Even in Phase 4, after 80 rounds, 70% of the participants selected this team, $\chi^2(1, N = 30) = 46.88$, $p < .001$. Its reduction was not reliable, $\chi^2(1, N = 60) = .80$, $p = .37$. By contrast, Condition 2 shows that participants provided with global-information-only were highly capable of quickly detecting that the altruist should be part of the team (Phase 1, $\chi^2(1, N = 28) = 7.00$, $p < .001$). Also the contrast between Conditions 1 and 2 was highly significant (Phase 4), $\chi^2(1, N = 58) = 27.15$, $p < .001$. In the heterogeneous global-and-local Condition 3, selections began with a high proportion of no-altruist team-choices in Phase 1 (Figure 12), $\chi^2(1, N = 32) = 67.57$, $p < .001$. In Phase 4, these individual-related selections, which exclude *A*, are likewise found to be above chance (56%), $\chi^2(1, N = 32) = 31.01$, $p < .001$; but now the group-related selections are above chance as well (excluding N4, with 34%); $\chi^2(1, N = 32) = 4.13$, $p < .05$. By contrast, in Condition 4 (a global-only condition), even in Phase 1 the optimal team-related selection (with-altruist team excluding N4) was the most frequently selected (43 %), $\chi^2(1, N = 32) = 11.28$, $p < .001$; and the no-altruist team, conversely, was selected below chance (3 %), $\chi^2(1, N = 32) = 5.70$, $p < .05$. In Phase 4, the selection of the no-altruist team was still selected with low relative frequency (3 %), and the optimal team by 59 % of participants.

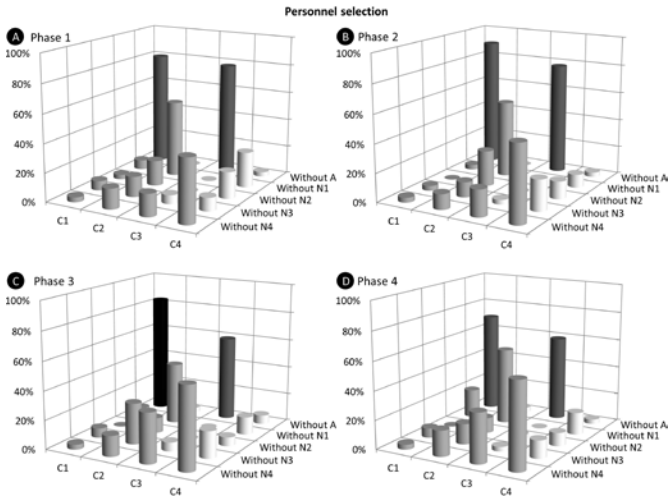

Figure 8. The results of the personnel selection task in the four test phases show the proportion of 'managers' choosing a particular team, thus excluding either worker *N1*, *N2*, *N3*, *N4*, or the altruist *A*. In the global-and-local conditions, C1 and C3, the black columns correspond to the predicted selections based on individual performance only. In the global-only conditions, C2 and C4, no individual-level information was available. In all conditions, the dark gray columns represent the optimal selection(s) based on overall performance of teams.

In the local-and-global conditions we coded comments as insightful that showed understanding of the differences between an individual and a group level. After 80 learning rounds, at least 38% of the participants in these conditions were classified as providing comments with insight (33% in C1 and 43% in C3). Of these participants, 87% selected the altruist personnel selection task (in Phase 4), whereas from the participants not demonstrating insight only 3% made this selection. Finally, from the additional tests, only the Need-For-Cognition Scale (2.6 vs. 10.5), $t(60) = 1.93$, $p = .03$ (one-tailed) correlated with insightful comments.

General Discussion

Experiment 1 shows that the tragedy of personnel selection is very stable across different team sizes. Even in the smallest team, most participants in the role of 'managers' evaluated the most useful worker for the group to be the worst. However, Experiment 1 suggests that small team sizes mitigates this problem, and a minority in this condition saw the difference between individual and overall contribution of an employee. Experiment 2 shows that

people are in principle well able to detect the strong correlation between presence of the altruist and high team performance (with r = .99) very early on from the group-level information. However, in other conditions most make no use of this ability and seem to ignore the overall payoff, focusing only on workers' direct individual contribution.

More generally, the findings may be due to people's problems dealing with decisions involving a Simpson's Paradox (Fiedler et al., 2003; Sydow et al., 2016; Waldmann & Hagmayer, 2001). If people do not merely optimize in a standard decision-theoretic way (here by simply choosing the team with the highest past performance), and instead, as we suggest, aim for a deeper understanding by identifying clear causal or logical patterns between events (e.g., Funke, 2001; Hagmayer & Meder, 2013; Osman, 2010; Sloman & Hagmayer, 2006; von Sydow, 2016; Waldmann & Hagmayer, 2001), this may yield the disadvantage of overlooking small correlations, pathways, exogeneities or interactions (Novick & Cheng, 2004), even if they may add up, tragically, to be the predominant effect of a scenario.

Acknowledgments

This work was partly supported by the grant Sy111/2-1 from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program *New Frameworks of Rationality* (SPP 1516). We are grateful to Maria Waltmann and Martha Cunningham for help and corrections. The first author is grateful to Daniel Holt, Klaus Fielder and Michael Waldmann for discussing related issues. Furthermore, we would like to thank anonymous reviewers for helpful comments and suggestions.

References

Brief, A. P. & Motowidlo, St. J. (1986), Prosocial Organizational Behavior. *Academy of Management Review, 11*(4), 710–725.

Fiedler, K., Walther, E., Freytag, P., & Nickel, S. (2003). Inductive reasoning and judgment interference: Experiments on Simpson's paradox. *Personality and Social Psychology Bulletin, 29*, 14–27.

Grant, A. M. & Patil, S. V. (2012). Challenging the norm of self-interest. Minority influence and transitions to helping norms in work units. *Academy of Management Review, 37(4)*, 547–588.

Engel, C. (2011). Dictator Games: A Meta Study. *Experimental Economics, 14*, 583–610.

Funke, J. (2001). Dynamic systems as tools for analyzing human judgment. *Thinking and Reasoning, 7*, 69–89.

Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality, 43(6),* 999–1005.

Hagmayer, Y., & Meder, B. (2013). Repeated causal decision making. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 39*, 33-50. doi: 10.1037/a0028643

Hardin, G. (1968). The Tragedy of the Commons. *Science, 162* (3859), 1243–1248. *doi:10.1126/science.162.3859.1243*

Li, N., Kirkman, B. L., & Porter, C. O. L. H. (2014). Toward a Model of Work Team Altruism. *Academy of Management Review, 39(4)*,541–565. http://dx.doi.org/10.5465/amr.2011.0160

Mathieu, J., Maynard, M. T., Rapp, T., Gilson, L. (2008). Team effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse Into the Future. *Journal of Management, 34*(3), 410-476. DOI: 10.1177/0149206308316061

Melis, A. P., Hare, B., Tomasello, M. (2006). Chimpanzees Recruit the Best Collaborators. *Science, 311*, 1297–1300.

Memmert, D., Plessner, H., Hüttermann, S., Froese, G., Peterhänsel, C., & Unkelbach, C. (2015). Collective fit increases team performances: Extending regulatory fit from individuals to dyadic teams. *Journal of Applied Social Psychology, 45*, 274–281. doi: 10.1111/jasp.12294

Mojzisch, A., Grouneva, L., & Schulz-Hardt, S. (2010). Biased evaluation of information during discussion: Disentangling the effects of preference consistency, social validation, and ownership of information. *European Journal of Social Psychology, 40*(6), 946-956. doi:10.1002/ejsp.660

Nielsen, T. M., Hrivnak, G. A., & Shaw, M. (2009). Organizational Citizenship Behaviour and Performance. A Meta-Analysis of Group-Level Research. *Small Group Research, 40*(5), 555-577.10.1177/1046496409339630

Novick, L. R., & Cheng, P. W. (2004). Assessing Interactive Causal Influence. *Psychological Review, 111*, 455–485.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437/27*, 1291–1296.

Organ, D. W. (1997). Organizational Citizenship Behaviour: It's Construct Clean-Up Time. *Human Performance, 10*(2), 85–97.

Osman, M. (2010) Controlling Uncertainty: A Review of Human Behavior in Complex Dynamic Environments. *Psychological Bulletin, 136*(1), 65-86.

Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Mishra, P. (2010). Effects of organizational citizenship behaviors on selection decisions in employment interviews. *Journal of Applied Psychology, 96 (2)*, 310–326.

Schulz-Hardt, S., & Mojzisch, A. (2012). How to achieve synergy in group decision making: Lessons to be learned from the hidden profile paradigm. *European Review of Social Psychology, 23*(1), 305-343. doi:10.1080/10463283.2012.744440.

Sloman, St. & Hagmayer, Y. (2006). The Psycho-Logic of Choice. *Trends in Cognitive Science, 10*(9), 407-411.

Sober, E., & Wilson, D. (1999). *Unto Others: The Evolution of Unselfish Behavior*. Harvard University Press.

Van Scotter, J. R., Motowidlo, S. J., & Cross, T. C. (2000). Effects of task performance and contextual performance on systemic rewards. *Journal of Applied Psychology, 85*(4), 526–535.

von Sydow, M. (2015). The Tragedy of Inner-Individual Dilemmas. In D. Noelle, et al. (Eds.), *Proceedings of the Thirty-Seventh Annual Conference of the Cognitive Science Society* (pp. 2517-2522). Austin, TX: Cognitive Science Society.

von Sydow, M. (2016). Towards a Pattern-Based Logic of Probability Judgements and Logical Inclusion "Fallacies". *Thinking & Reasoning, 22(3)*, 297-335. doi:10.1080/13546783.

von Sydow, M., & Braus, N. (2016). On the Tragedy of Personnel Evaluation. In A. Papafragou, et a. (Eds.), *Proceedings of the Thirty-Eighth Annual Conference of the Cognitive Science Society* (pp. 105-110). Austin, TX: Cognitive Science Society

von Sydow, M., Hagmayer, Y., & Meder, B. (2016). Transitive reasoning distorts induction in causal chains. *Memory & Cognition, 44*(3), 469–487. doi:10.3758/s13421-015-0568-5

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*, 27–58. doi: 10.1016/S0010-0277(01)00141-X

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *Quarterly Review of Biology, 82*(4), 2007, 327–348. doi: 10.1086/522809