

Thinking Locally or Globally? – Trying to Overcome the Tragedy of Personnel Evaluation with Stories or Selective Information Presentation

Momme von Sydow^{1,2} (momme.von-sydow@urz.uni-muenchen.de)

Niels Braus² (n.braus@stud.uni-heidelberg.de)

Ulrike Hahn^{1,3} (u.hahn@bbk.ac.uk)

¹University of Munich (LMU), MCMP, Ludwigstr. 31, D-80539 München, Germany

²University of Heidelberg, Department of Psychology, Hauptstr. 47-51, D-69117 Heidelberg, Germany

³Birkbeck College London, Department of Psychological Science, Malet Street, London WC1E 7HX, U.K.

Social dilemmas conceptually suggest distinguishing direct individual and group-level effects (also involving indirect effects on others). Furthermore, the success of organizations appears to rely on identifying not only individual excellence but positive impact on others as well. In ‘Two-Level Personnel Evaluation Tasks’ (T-PETs) participants as human resource managers evaluate employees when individual and group contributions are dissociated. Von Sydow, Braus, & Hahn (2018) have suggested a potential ‘Tragedy of Personnel Evaluation’: A group-serving employee with the smallest individual contribution but by far the greatest positive effect on the group’s overall earnings was often rated the most negatively. Here we investigate, in two experiments with conflicting information, whether emphasizing the group can avert the ‘tragic’ outcome. Our results suggest that the tragedy is not as complete as suggested, and that contextual information can mitigate the tragedy. Nonetheless, the results also corroborate the stability of underestimating the impact of team players.

Keywords: Co-variation Detection; Inner-Individual Dilemma; Co-operation; Multi-Level Approach Simpsons Paradox

Individual versus Group Utility

Co-operation between individuals – over and above direct *individual* benefit – is essential for the common good of organisations, companies, and society on the whole. Successful co-operation often involves setting personal interests aside and devoting oneself, at least partly, to the success of an organization or team. Thus when evaluating behaviour one must distinguish direct individual utility from group utility.

In evolutionary biology, multi-level selection-approaches stress the differences between behaviour benefiting either the individual or the group; and these models allow for the rise of altruism (Sober & Wilson, 1999; Nowak & Sigmund, 2005; Wilson & Wilson, 2007).

Likewise, social-dilemma situations can be interpreted as conflicts of individual and group interests; and it has been argued that purely self-interested, economically ‘rational’ behaviour may inevitably lead to the depletion of public and natural resources. This has become associated with the label ‘Tragedy of the Commons’ (Hardin, 1968). Research in game-theory, behavioural economics and psychology (e.g., on ultimatum games, dictator games and public-good games) has revealed strategies for solving social-dilemma situations, and that many people do *not* act in a purely self-interested manner, but rather demonstrate at least some preference for distributing goods justly or behaving in a group-serving manner (Engel, 2011; Hendrich et al., 2005,

Fehr & Gächter, 2002; Gollwitzer, Rothmund, Pfeiffer, & Ensenbach, 2009; Melis, Hare, & Tomasello, 2016).

In Organisational and Social Psychology, the importance of the team-level has been increasingly acknowledged, emphasizing that teams may be greater than the sum of their parts (Haslam, Steffens, Peters, Boyce, Mallett, & Fransen, 2017; Mathieu, Maynard, Rapp, & Gilson, 2008; Memmert, Plessner, Hüttermann, Froese, Peterhänsel, & Unkelbach, 2015). Likewise, the role of pro-social or altruistic role or extra-role behaviours in teams has been identified as central to the success of companies and organizations (Brief & Motowidlo, 1986; Li, Kirkman, & Porter, 2014; Nielsen, Hrivnak, & Shaw, 2009; Organ, 1997; Podsakoff, Whiting, Podsakoff, & Mishra, 2010).

A crucial question, however, is the extent to which people recognize those who clearly serve the overall good at the team level. This question should be particularly pressing for human-resource managers who must evaluate or select employees and must often only base their judgment on abstract performance data (sales numbers, etc.).

Tragedy of Personnel Evaluation

For human-resource management, it seems crucial to address the potential dissociation of employees’ individual and collective impact on team performance. While underlining this, recent research has also provided some first evidence that such behaviour is sometimes rewarded – particularly when managers have direct acquaintance with the processes and persons involved (Organ, 1997; Scotter, Cross, & Motowidlo, 2000; cf. Grant & Patil, 2012, 562). However, personnel managers must increasingly evaluate without first-hand experience, often based on abstract performance numbers (Brandl, 2002).

We have begun to study evaluation situations from an experimental perspective as well, using well specified Two-level Personnel-Evaluation Tasks (T-PETs; von Sydow & Braus, 2016, 2017; von Sydow, Braus, & Hahn, 2017, 2018). In these T-PETs, participants obtained information about employees’ earnings on individual as well as overall group levels. The T-PET used involves strongly conflicting information at both levels. The group-serving person A is characterized by lowest individual earnings yet has a consistent, strongly positive impact on the overall team earnings by substantially increasing the earnings of the other employees. This group-serving person is here called ‘altruist’. Although (behavioural) altruism in biology and economics seem to be associated with this kind of indi-

vidual and group impact, it should be noted that such patterns do not imply *motivational* altruism (only ‘prosocial’ behaviour). For simplicity, however, we call the group-serving person, team-player or positive interactor simply ‘altruist’ (A).

In previous work on T-PETs, participants evaluated the ‘altruist’ who was best for the team and company as worst, and they tended to ostracize the altruist in selection tasks. These results led to the suggestion of a potential ‘tragedy of personnel selection’: Personnel managers may neglect or underestimate group impact with substantial damage to personnel and ultimately companies and organisations (von Sydow & Braus, 2016; von Sydow et al., 2018). This occurred despite the strong correlations between group membership and team performance (von Sydow et al., 2018). Further studies showed that, although negative group-interaction or ‘egoist’ detection (egoism here again defined behaviourally only) differed slightly from ‘altruist’ detection, they both demonstrated a broadly similar tragedy of ignorance regarding overall group-level effects (von Sydow & Braus, 2017). Another study already has investigated the role of group size in T-PETs. Holding the effects of A on *single* other individually constant, the small group demonstrated no considerable advantage (von Sydow, Braus, & Hahn, 2017).

The current personnel-evaluation experiments again involve similar scenarios with conflicting information at the individual and group levels (T-PETs). However, they investigate whether a shift in the known importance or salience of the group level, by either varying cover-stories (Exp. 1) or selective information presented (Exp. 2), yields improvement. Experiment 2 additionally examines ratings distinguishing explicitly direct impact, impact on others, and overall impact on a team.

Experiment 1: Story-Induced Focus

We used a straightforward manipulation, providing participants with texts stressing either the role of the individual or the team as central for personnel management. The experiment had four conditions, with stories focusing on different levels (C1 individual; C2 global; C3 individual & global; C4 control, no focus and no additional text). Additionally, the order of the dependent variables in all four learning phases was counterbalanced (evaluation → selection vs. selection → evaluation).

Method

Participants The experiment was conducted via MTURK with participants from the US. 121 participants passed the two selection-criteria (time spent on the first page, and the correct choice of a rephrasing of the instructions) and finished the experiment. The participants obtained a compensation of \$1.50. 49% were male; the mean age was

Employee				
Earnings of each employee in Euros (€)	1611	3250	2895	2281
Total earnings of this day in Euros (€)	10037			

Figure 1: Example of shown earnings at the individual and group levels on a particular day.

36; 55% had a Bachelor’s or Master’s degree and 30% a high school degree as their highest level of education.

Procedure and material The computer experiment resembled previous T-PETs (von Sydow et al., 2018) and was implemented using SociSurvey.

Participants first obtained general instructions that their role as personnel manager was to evaluate the employees of a particular shop. Daily they would obtain information about individual and total earnings of the team working that day. Overall there were five staff members working in the shop, in day-shifts of four people.

On the next slide, participants read that the retired founder of the company had delivered a talk, mentioning the essential role of the personnel management to a company’s success. As space precludes exhaustive citation here, we present only the first and fourth of five paragraphs of C1 and C2:

C1: “What is a company? A company is composed of **individual employees working on their tasks**, and it rises and falls with their performance. Thus a company needs to incentivize the performance of **each individual employee**. [...] This alone will do justice to those individuals **who do a good job over those who do a bad job**. In particular, you need to detect those **employees who individually perform best and worst.**”

C2: “What is a company? A company is **more than just the sum of its employees**; it is a **whole**, a finely attuned **organism**. It is made up of **teams** in which employees need to **interact** in a positive way. Thus, a company needs to incentivize **team performance**. [...] This alone will do justice to those **teams with positive interaction** and to **good over bad team players**. In particular, you need to detect the **members of the group who support and those who exploit the group**” (bold print added).

C3 analogously emphasized the importance of monitoring both individual AND group effects of employees.

In the main part, participants sequentially obtained for each day transparent overview information about the individual earnings of each of the four employees (presented by a picture) working on the shift that day, as well as information on overall earnings (see Figure 1). The structure of the earnings is shown in Table 1 (we added some noise to each value; a normal distribution with SD = 600 €). On the level of individual earnings there were relatively small mean differences (400 €) between the four non-interacting normal workers N_x and the altruist A: $N1 > N2 > N3 > N4 > A$.

Apart from the lowest individual earnings of the ‘altruist’, his/her presence had by far the most positive impact on the group earnings – when *A* is present, the earnings of all co-workers increase by 1000 € leading to an overall average increase of 2500 € (normally exceeding the salient mark of 10,000 €).

Table 1: Mean earnings of normal workers (*Nx*) and altruist; and mean overall earnings.

	With Altruist	Without Altruist
N1	3,600 €	2,600 €
N2	3,200 €	2,200 €
N3	2,800 €	1,800 €
N4	2,400 €	1,400 €
Altruist	1,500 €	-
Total	10,500 €	8,000 €

The presence of this worker correlated with $r = .99$ with the overall outcome – a correlation easily detectable by machine-learning algorithms.

Overall, 80 panels (Table 1) were sequentially shown and participants could view the overview panels for each day as long as required, with a minimum of four seconds. The role of the altruist (team player) was randomly assigned to one of the five persons of which four are working in a particular shift. He or she appeared randomly 50% of the days (shifts) in the overview panels (Table 1); the four normal workers appeared randomly. We further counterbalanced the presentation-order of the four employees working in each shift.

In four test phases, after the 20, 40, 60 and 80 rounds, the ‘personnel managers’ evaluated all employees in an evaluation task and a selection task. The order of these tasks was counterbalanced. In the evaluation task, participants rated the contribution of all employees to the overall earnings of the company, on a scale of one to ten. In the personnel selection task they had to answer which four of the five workers they chose to work another day “to optimise the overall profit for the company”. At the end of the fourth test phase, we added a total utility task: “Which person is of the greatest/lowest total utility for your business?” and assessed participants’ preference for narrow self-interest and pro-social behaviour, using the social-value orientation scale (Murphy, Ackermann, & Handgraaf, 2011). Finally, they provided comments and demographic data.

Results

Figure 2 shows the average evaluation ratings by person in the four test rounds for the four conditions. First, Figure 2

reveals that the average ratings, in all conditions, mainly reflects the average *individual* earnings of the employee, with the ratings of the altruist, *A*, always being lower than all (or at least most) normal workers, *N*. A repeated-measures ANOVA with factors Person (*N1* to *N4*, *A*) and Phase (R1 to

R4) as within-subject factors, and the condition Story (C1 to C4) as between-subject factors, revealed a significant main effect of Person only, $F(2,257) = 216,20$, $p < .001$, $\eta^2 = .65$, but no significant main effects of Condition or Phase.

Second, the ratings reveal that the order of the mean ratings of the normal workers is always (even in Phase 1) in line with the actual (small) differences observed, resulting overall in order ($N1 > N2 > N3 > N4 > A$), with significant post-hoc tests for all four comparisons (each $p < .001$).

Figure 3 shows the percentage of group-based answers in the selection task (selections of *A*), the rating task (all $N \geq A$) through time, and the final comments (coding some insight into the difference between individual and group performance). Selection, comments, and in the beginning rating as well, seem to reveal a similar pattern: a slight advantage for the conditions with global stories (C2 and C3). Comparing these two conditions with those without global stories (C1 and C4), this predicted effect in Phase 4 reaches significance for selection: $\chi^2(df=1, N=121) = 4.30$, $p < .05$. For ratings there seems only an effect for the global condition, not the individual and global condition. This difference between selection and rating perhaps becomes understandable if one bears in mind that correct rating here involved a stricter test criterion than selection (not only judging *A* higher than *one* other worker, but *all* other workers).

Discussion Experiment 1

First, the results of Experiment 1 overall corroborate the postulated stability of the Tragedy of Personnel Evaluation with no strong impact of the story. The average ratings of the altruist, who actually very consistently causes strong improvements in group performance, are lowest; and at least 50% of the participants seem not to detect the effects of the presence of the ‘altruist’. However, the results also suggest

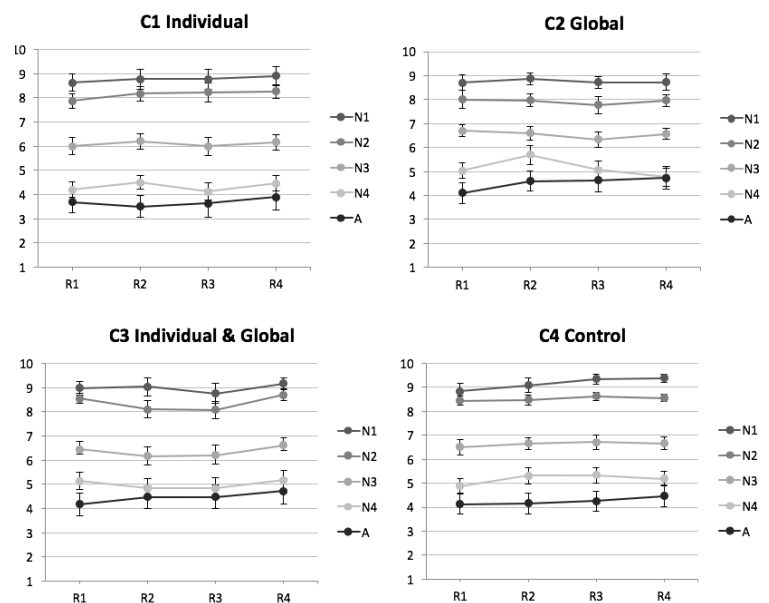


Figure 2: Mean ratings of the employees over test rounds in the four story conditions.

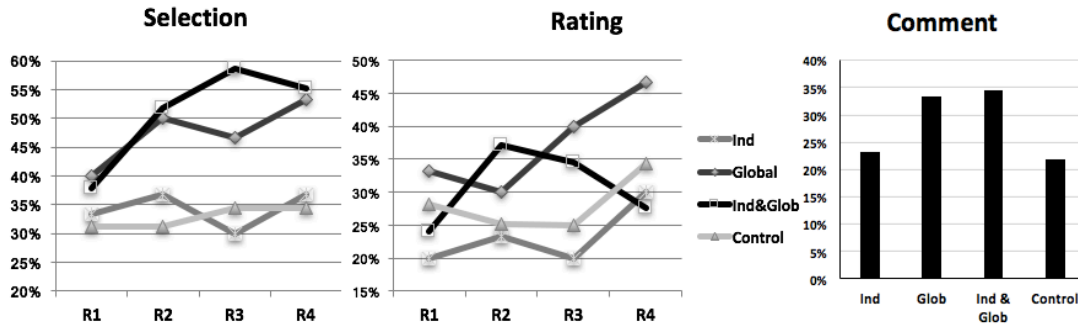


Figure 3: Percentage of group-based selections, evaluations (rating A > all N), and comments by conditions in Experiment 1.

some small effects of the story, and that thinking about the group level can reduce the Tragedy of Personnel Evaluation at least with regard to some participants.

Experiment 2 – Selective Information Presentation

Experiment 2 varies what information (group versus individual versus both) is presented and at what time. It also increases the number of test phases, to explore the temporal dynamics in more detail and what people could learn even after ten trials. Finally, it tests the ability of participants to detect group-level (and individual-level) effects, also using ratings explicitly differentiating between individual effects, effects on others and overall team effects.

With regard to selective information presentation, there is a condition focusing participants on group-level information alone (*G*) in order to investigate whether and how quickly the ‘altruist’ could be detected (now using eight test phases). Presenting only individual-level information (*I*), provides the other extreme base-line condition. Always presenting both – individual and a team’s overall earnings (*B*) – replicates Experiment 1 but can now be compared to both benchmark conditions. Moreover, we added several further ‘mixed conditions, where the three information formats (*G*, *I*, *B*) changed over time (e.g., GIGIGIGI).

The increased number of test phases may allow participants more easily to realize the tension between individual-level success and overall group-level effects. The contrast to both extreme base-line conditions (only *I* or only *G*), should serve as controls for the level of performance in the intermediate conditions, that ‘only *B*’ or the mixed conditions. In a number of mixed conditions the shown information is varied over time (e.g., GIGIGIGI). We explore whether they may be adventitious in contrasting the global and individual level.

Method

Participants (recruited by Prolific Academic) came from English-speaking countries (i.e. the UK 52%, the US 32%, Ireland, Australia etc.). 172 participants passed all selection criteria and finished the experiment (cf. Exp. 1). Each participant obtained a compensation of £1.80. The mean age was 32 (59% male, 41% female). Regarding education, all participants had at least high school degree or A levels.

Procedure and material The materials and procedure apart from some differences strongly resembled the T-PET of Experiment 1. We used a similar, neutral introduction, but without stories. In the main part, participants again obtained, sequentially for each day, overview information (Figure 1) based on the same average earnings as before (Table 1). Thus, on the *individual* level, the rank-order of earnings for the normal (non-interacting) (*N_i*) workers and the altruist (*A*) worker was $N_1 > N_2 > N_3 > N_4 > A$, with small differences (400 €). The altruist impact on the other workers in a shift was larger, increasing each of their average earnings by 1000 € and the average earnings of the team by 2500 € resulting in a reversed rank order: $A > N_1 > N_2 > N_3 > N_4$.

Participants in the role of personnel managers were again shown information on four employees’ earning (out of five employees overall) for 80 shifts (days). We again counterbalanced presentation-order of persons shown (see Exp. 1). Participants could study the overview panels as long as required, but with a minimum of four seconds.

Table 2: Information in phases (P_x) and conditions (C_x).

	P1	P2	P3	P4	P5	P6	P7	P8
C1	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>
C2	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>
C3	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>
C4	<i>G</i>	<i>B</i>	<i>G</i>	<i>B</i>	<i>G</i>	<i>B</i>	<i>G</i>	<i>B</i>
C5	<i>G</i>	<i>G</i>	<i>B</i>	<i>B</i>	<i>G</i>	<i>G</i>	<i>B</i>	<i>B</i>
C6	<i>G</i>	<i>I</i>	<i>G</i>	<i>I</i>	<i>G</i>	<i>I</i>	<i>G</i>	<i>I</i>
C7	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>

Note: *B* = both individual & group; *G* = group only; *I* = individual only; C_x = condition; R_x = round

In contrast to Experiment 1, Experiment 2 uses eight (instead of four) test rounds and eight preceding learning phases, each composed of ten days. Moreover, we varied the information formats. In a learning phase 10 information panels (10 days, one for each shift) are shown, with information either only on the individual earnings (*I*; see Figure 1 without last row), on the group earnings (*G*; Figure 1 without middle row) or, finally, on both individual and group earnings (*B*; cf. Figure 1). Table 2 presents the information formats in different conditions and phases.

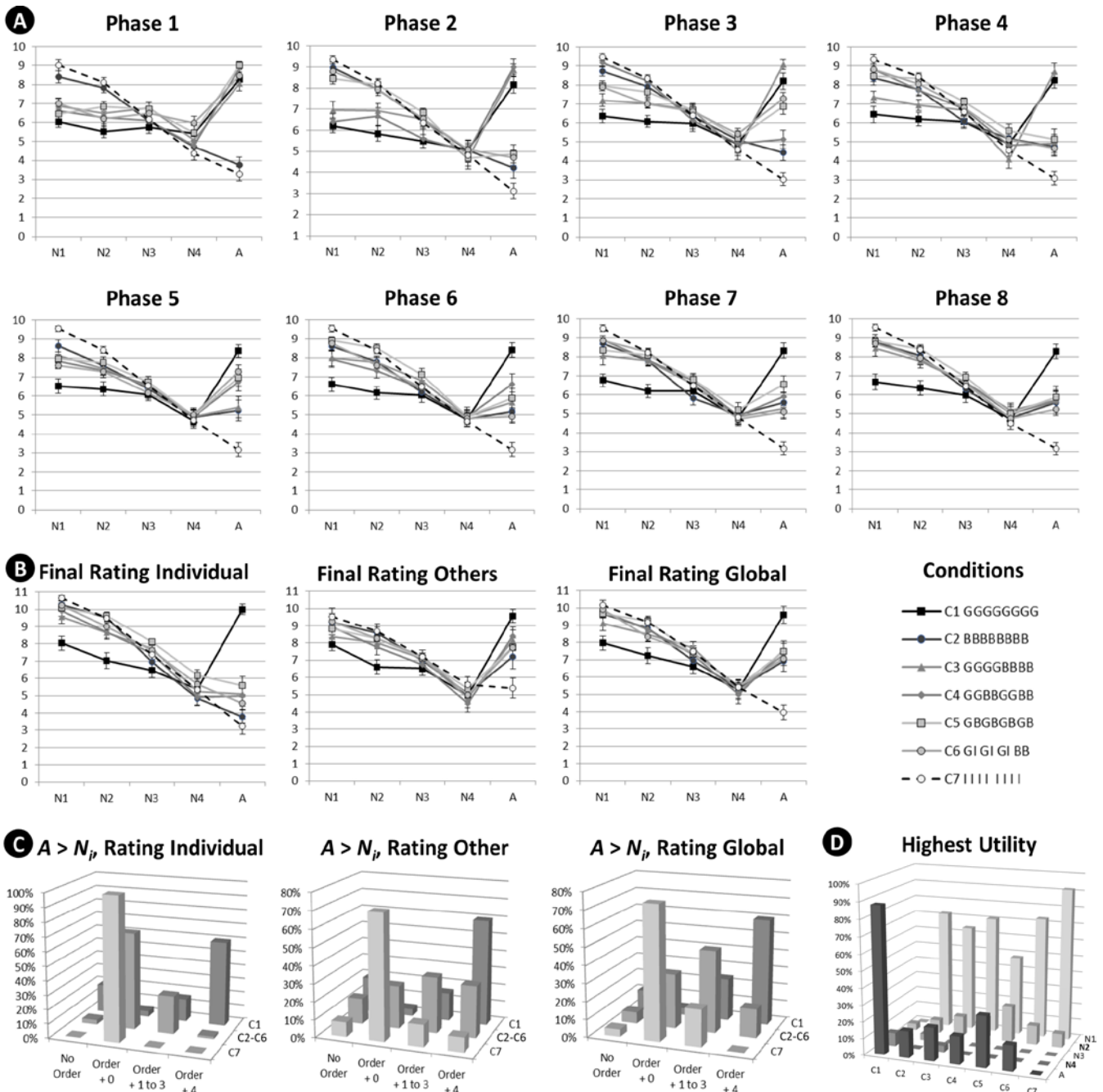


Figure 4: *Panel A* shows participants' mean ratings for the 'altruist' worker (A) and the four non-interacting normal workers (N) for the seven conditions and eight learning phases. *Panel B* presents the mean of the final ratings, differentiating between workers' individual earnings, effect on others and resulting overall effect (global). For these ratings, *Panel C* shows the percentage of participants who had no sense of individual differences between normal workers ('No Order'; $N1+N2 > N3+N4$). For those who detected this basic individual order ('Order'), we show the percentage, rating the altruist higher than none of the Ns ('+ 0'), then one to three Ns ('+ 1 to 3'), and then all four Ns ('+ 4'). *Panel D* shows the percentage of participants attributing the overall highest utility for the company to the 'altruist' or a particular normal worker.

In each test round, we administered rating scales and then a personnel selection task. At the end, participants should again provide ratings for all five workers but now explicitly using three rating scales differentiating between individual impact, impact on others and overall contribution. Finally,

participants had to choose the employee with greatest or lowest utility and comment on their own behavior.

Results

Figure 4 presents the main results of Experiment 2. *Panel A* shows the participants' mean ratings of the five workers'

contributions in the eight phases. These results show that (a) even after ten presentations (Phase 1) participants' mean judgments suggest a high altruist-detection tendency (at the expense of less clear individual differences between the normal workers); (b) between the phases there was a strong variation of the mean ratings of the mixed middle conditions, suggesting that at least in the beginning participants had difficulties in integrating the results and were strongly influenced by the recent phase; and (c) at the end (Phase 8), however, all middle conditions (C2 to C6) show a middle result between the benchmarks (C1 and C7), and there is no large difference between the middle conditions. Despite still underestimating the altruist, participants clearly do not focus only on individual-level information.

Panel B shows the results of the final ratings differentiating between workers' individual earnings, their impact on others, and the overall impact. (a) Participants in the middle conditions (C2 to C6) clearly differentiate between these ratings. (In the two extreme conditions this was unlikely, but there is even a small effect in C7.) In the middle conditions the individual ratings show $A < N_s$, but the 'other-ratings' show that A is, on average, rated higher than N_4 , and similarly high as N_3 and N_2 . The global rating may be a mixture that seems even more strongly to resemble the 'other-ratings' (but note the incommensurability of the other-rating that used a bipolar scale). In any case, the global rating in the middle conditions clearly does not reach the C1 benchmark, suggesting a remaining tragedy. (b) Comparing the final ratings with the ones in Phase 8 suggests that participants do interpret the latter largely as global ratings (with a small individual influence). (c) Given that the impact-on-others rating would correctly be answered following $A > N_1 = N_2 = N_3 = N_4$, there is not only a correct $A > N_s$, but also an incorrect impact of the individual order $N_1 > N_2 > N_3 > N_4$. This suggests the heuristic '(s)he who is good individually also helps others'.

Panel C investigates the individual differences of the final ratings. (a) Only few participants demonstrate no sense of individual differences ('No Order'; not $N_1 + N_2 > N_3 + N_4$). Though the individual condition C7 does have advantages here, also all other conditions fare reasonably well.

(b) Looking at the others who detected the basic order between the N_s , the individual and the group condition (C7 and C1) show highly similar results for the three ratings, suggesting a transfer in both directions. In contrast, the middle conditions (C2 to C6) differentiate between the conditions: In the individual ratings, most participants here detect correctly that $A < N_s$ (Order + 0); and in the 'others-ratings' they rate A higher than one, two, three (Order + 1 to 3) or even all four N_s (Order + 4). The 'others-ratings' and global-ratings make clear that the results of the middle conditions lie between both extreme conditions. Thus it is clearly wrong to claim that all participants completely ignore the group effect; but it is also apparent that only a few rate the altruist as high as would be appropriate based on A 's overall (direct and indirect) impact on overall earnings. *Panel D*, finally, shows who the participants judge to be of highest

overall utility for the company. (a) C1 shows that, with a focus on group-information, all participants learned the foremost utility of A , whereas none learned it in the individual C7. (b) Despite the cited positive effects in the middle conditions, the dependent variable shows that the altruist is still underrated. Nonetheless, a considerable minority also in 'C2 to C6', and more than in C7 ($p < .001$), rates altruist as high as would be appropriate based on A 's (direct and indirect) impact on overall earnings ($A > all N_s$).

General Discussion

Experiment 1 documents the stability of the Tragedy of Personnel Evaluation. The altruist's or team-player's outstandingly positive *overall* effects on a team were ignored or inadequately acknowledged by most participants; providing a context emphasizing the group had no large effect. However, Experiment 1 suggests that the postulated tragedy could – at least for some participants – be mitigated by contextual cues enhancing focus on the group-level.

Experiment 2 varied the information presented (individual information, group information, or both), used several test-phases, and at the end used rating-scales differentiating between individual earnings, effect on others, and overall impact on group earnings. The results show that participants in principle can learn the overall advantage of A as early as in Phase 1 (even after 10 'days') if forced to focus on the global information alone (C1). Second, toggling what information is presented over time did not provide the strong boost we hoped for. Third, and this seems important, the base-line conditions reveal that the tragedy is not at all a total one (at least here after 8 test phases). That is, participants' judgements in all middle conditions (including the only B condition) did differ also from an individual-focus condition. Nonetheless, the average ratings, the percentage of correct ratings, and the highest utility judgements show that most participants still substantially underestimate the overall impact of the 'altruist' team player.

Overall, the results are two-fold. They show that the postulated tragedy is neither completely immune to improvement (Exp. 1) nor as radical (Exp. 2) as perhaps originally suggested. However, despite using strong group-level effects, the results obviously do not allow for acquittal but rather corroborate a remaining (but reduced) tragedy. Even though this provides some first evidence for a means to mitigate the tragic, it also continues to underline the danger of potentially similar tragedies in the real world.

Future avenues of research should explore theoretical implications, mediating mechanisms, applications, and boundary conditions of these findings also in real-life settings. Although we here used highly educated participants and strict selection criteria, it would for instance be important to explore the stability of our findings with real personnel managers as well, with or without a number-based task (cf. von Sydow et al., 2018).

Moreover, this research may well be connected to several lines of more theoretically inspired research. For instance, our social-cognition two-level personnel evaluation tasks

may be understood more generally as studies of Simpson's paradox (Waldmann & Hagmayer, 2001; Fiedler et al., 2003; cf. von Sydow, Hagmayer, & Meder, 2016).

Second, we have suggested *some* rational basis for the apparently irrational reluctance to check for large correlations with a high overall outcome (the sum effect of many causal effects). We have suggested that this may be due to a concern with local causal relations rather than ephemeral overall outcomes (von Sydow et al., 2018; cf. Lagnado, Waldmann, Hagmayer, & Sloman, 2006; Sloman & Hagmayer, 2006; Hagmayer & Meder, 2013). Thus other interaction patterns may be more easily detected. For instance, it is known that people are well able to see some logical or causal interaction-patterns if focusing on two or three variables only (e.g., von Sydow, 2016).

In any case, the phenomenon seems of high importance, and the current research warns us that people, at least in the setting of number-based evaluations and perhaps beyond, may well tend to ignore or underestimate the strong overall group effects of team players in contrast to their individual effects.

Acknowledgments

This project was funded by the Humboldt Foundation's 'Anneliese Meier Research Award' to Ulrike Hahn and earlier, by a grant Sy111/2-1 from the Deutsche Forschungsgemeinschaft (DFG) to Momme von Sydow.

References

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

Brandl, J. (2002): Die Problematik der Kennzahlen in Personalinformationssystemen. *Personalführung*, 9, 42-47.

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.

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

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.

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.

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.

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.

Haslam, S. A., Steffens, N. K, Peters, K, Boyce, R. A., Mallett C. J., & Fransen, K. (2017). A social identity approach to leadership development. *Journal of Personnel Psychology*, 16(3), 113–124. DOI: 10.1027/1866-5888/a000176

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2006). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.

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

Murphy, R. O., Ackermann, K. A. & Handgraaf, M. J. J. (2011). Measuring Social Value Orientation. *Judgment and Decision Making*, 6(8), 771-781.

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

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.

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.

Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences*, 10, 407– 412. doi:10.1016/j.tics.2006.07.00

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

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., Braus, N. (2017). Altruist vs. Egoist Detection and Individual vs. Group Selection in Personnel Management. *Proceedings of the Thirty-Ninth Annual Conference of the Cognitive Science Society* (3466-3471). Austin, TX: Cognitive Science Society.

von Sydow, M., Braus, N., Hahn, U. (2017). Overcoming the Tragedy of Personnel Selection? *Proceedings of the Thirty-Ninth Annual Conference of the Cognitive Science Society* (pp. 3460-3465). Austin, TX: Cognitive Science Society.

von Sydow, M., Braus, N., & Hahn, U. (2018). On the Ignorance of Group-Level Effects – The Tragedy of Personnel Selection. *Journal of Experimental Psychology: Applied*. Advance online publication. doi: 10.1037/xap0000173

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.

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *Quarterly Review of Biology*, 82(4), 2007, 327–348. doi: 10.1086/52280