On the Tragedy of Personnel Evaluation

Momme von Sydow (momme.von-sydow@lrz.uni-münchen.de)
[1]University of Munich (LMU), Munich Center for Mathematical Philosophy (MCMP), Ludwigstr. 31, D-80539 München
[2]University of Heidelberg, Department of Psychology, Social Psychology, Hauptstr. 47-51, D-69117 Heidelberg, Germany
(momme.von-sydow@psychologie.uni-heidelberg.de)

Niels Braus (n.braus@stud.uni-heidelberg.de)
University of Heidelberg, Department of Psychology, Hauptstr. 47-51,
D-69117 Heidelberg, Germany

In social-dilemma situations (public-good games) people may pursue their local, egoistic interests and thereby lower the global, overall payoff of their group and, paradoxically, even their own resulting payoff. One may also speak of intra-individual dilemmas, where people pursue local goals at the expense of their overall utility. Our current experiments transfer this idea to the context of personnel evaluation and personnel selection. In our experiments, participants were put in the position of a Human Resources manager, who should for instance select workers who optimize the overall payoff of the company, rather than those who optimize only their specific payoffs. The results of the experiments, however, suggest that most, albeit not all, participants tended to focus on directly comparing individuals without considering the overall contribution to a group. Thus employees with the best overall effects for a company or organization may be evaluated the most negatively. This possible 'tragedy of personnel evaluation' may be linked to maladaptive incentive structures (personnel evaluation), advancement of employees (personnel promotion) and job offers (personnel selection), and may have a substantial negative impact on the effectiveness of companies or organizations.

Keywords: intra-individual dilemmas; social dilemmas; personnel evaluation; personnel selection, altruism; causal induction; global vs. local optimization; less-is-more effect

Inner-Organizational Dilemmas

Adam Smith (1776) famously argued that no altruism is needed to promote the common good: "By pursuing his own interest he [the individual] frequently promotes that of the society more effectually than when he really intends to promote it."

In contrast, in the wider context of evolutionary biology, economics and psychology, it has long been noted that social dilemma situations often arise when groups of organisms or human agents act 'optimally' – in the sense of pursuing their egoistic interests only. People in such situations may often behave contrary to the best interest of the whole, which paradoxically often results in disastrous outcomes for each individual as well. For instance, it has been argued that the over-exploitation and destruction of finite public resources by 'rationally' acting, selfish individuals is inevitable. This has been called the "tragedy of the commons" and has been discussed with regard to environmental pollution and sustainable development (Hardin, 1968). Social dilemmas have widely been studied, theoretically and empirically, in psychology and behavioural economics. Recent decades have yielded lively debate of possible solutions of the perhaps not always inevitable tragedies. Moreover, the debates have eroded the explicit or tacit strict egoism assumption without ignoring the limits of altruism (e.g., Fehr & Fischbacher, 2003, 2004; Fehr & Schmidt, 1999; Fehr & Gächter, 2002; Henrich, 2005; Ostrom et al., 1999). Likewise, evolutionary biology in recent decades has shifted from an emphasis on individual egoism (or gene-egoism) to an acknowledgement of multi-level approaches that generally suggests one should find egoistic as well as altruistic behaviour tendencies in social groups (Nowak & Sigmund, 2005; Sober & Wilson, 1999; Wilson & Wilson, 2007; cf. von Sydow, 2011).

In organizations it is also plausible, then, that each member frequently faces tension between egoistic and group-serving behaviours. The issue is not constrained to moral questions of pursuing the common good of a society, but arises within companies, for managers and workers alike, even when they merely aim to optimize certain economic key figures (e.g., net sales or operating profit). Solving such intra-organizational dilemmas, therefore, is one of the most crucial tasks of building efficient systems of co-operation.

One obvious potential solution of inner-organizational dilemmas seems to install an institutionalized, neutral third-party that tries to access justly the contributions of egoistic versus group-serving individuals. This seems linked to good leadership or should at least be a major concern of management control systems or Human Resources managers.

Yet do people really have the capacity for "altruist detection" or "egoist detection"?

First, one must concede that basic biological multi-level models of group altruism do not necessarily require the ability of organisms to detect altruist or egoists (Sober & Wilson, 1999; Wilson & Wilson, 2007). However, there are refined models that refer to such abilities, can be evolutionary stable without necessitating cultural transmission (Nowak & Sigmund, 2005). Thus it seems implausible to deny a general ability of altruist detection in humans.

Second, a *prima facie* potential for egoist or altruist detection – or, put more cautiously, of co-operator and cheater detection – seems quite plausible (at least for human beings) if we consider the importance of the topics in daily contexts. One might object that, in times when multi-level models had a bad standing in evolutionary biology, there was likewise psychological evidence in simple rule-testing tasks (Wason Selection Tasks) that people are only

successful at detecting cheaters but not co-operators or altruists (Cosmides, 1989). However, further research has suggested that this greater facility for cheater-detection depends on the question asked and the goals pursued (von Sydow, 2006; von Sydow & Hagmayer, 2006).

Third, in the specific context of our experiments on Human Resources management and personnel evaluation, it seems even clearer, at least from first sight, that people in management roles should be able roughly to discern who contributed the most to a company's success (if provided appropriate quantitative information). One reason is that, from the vantage point of game-theoretical analysis, the social dilemma becomes an "*intra*-individual dilemma" (von Sydow, 2015) that is normatively to be solved globally; the personnel manager should at least normatively optimize globally and refrain from selecting or rewarding employees with good individual operating figures over those who overall best serve the interests of the work group, company or organization. However, even for intra-individual dilemmas, it has been shown that people may optimize locally, at the expense of the global level (von Sydow, 2014). For instance, an individual may pursue too many projects, each with a positive utility, but may thereby reduce overall utility by ignoring those projects' negative external effects on other projects. Transferring the idea that people may often not realize or acknowledge the positive externalities if local effects, may in the context of a human recource management task imply that people may ignore interactions with other employees or the distributed indirect effects of group-serving altruism. The use of a simplified one-level model – considering only individual outputs – might be due to issues of complexity. Even if plausibly – as we argued before - people in daily life can distinguish egoistic from group-serving individuals, this need not apply to economic personnel evaluation contexts were evaluation is based on given numbers, where reporting units often have no direct acquaintance with the persons and processes involved, and in such contexts one might perhaps more easily ignore questions of group-serving behavior. Additionally, the context of personnel evaluation may elicit a competitive individual frame of mind.

This prompted our hypothesis that people in charge of personnel-evaluation may ignore the positive or negative effects of individuals on the work of other members of a larger group (externalities), sticking only to a person's direct operating figures. This, however, may imply a "tragedy of personnel selection", since people with the best overall effects on group-performance (e.g., by helping others and indirectly contributing the most to the overall goals of a company), might be evaluated the most negatively.

Experiment 1

In Experiment 1, participants put themselves in the role of Human Resources manager evaluating the utility of employees in a local shop. The managers obtained information on direct earnings of individuals and those of the shop daily. The overall profit strongly positively depended on the presence of an individual employee in a shift, suggesting his or her importance to the overall earnings of the team. We here call this employee "the altruist," since although not demonstrating particularly positive earning figures individually, he or she enhances the values for all other employees working during the same shift. Note, that although a functionally defined altruism would lead to the shown data, the number-based evaluation context does provide evidence for any intentions or effort. Hence, a perhaps more neutral term for the worker would have been 'facilitator'. We here nonetheless use the term 'altruist' as one easily conceivable interpretation. We aimed to explore how far participants as Human Resources managers take the essential indirect monetary contributions of an altruist worker into account, when evaluating workers.

Design

Experiment 1 has a two (rounds: 10 working days versus 20 working days) by two (difference of earnings between a normal worker with versus without the presence of the altruist; see Table 1), within-subjects design.

Table 1: Mean earnings of normal workers (NW) and altruist; and overall earnings

	Condition 1 / 2	Condition 3 / 4
NW without altruist	2200 €	2000 €
NW with altruist	2800 €	3000 €
Altruist	1600 €	1600 €
Overall without altruist	8200 €	7600 €
Overall with altruist	10000 €	10600 €

Method

Participants 124 participants from the US volunteered for the experiment via MTURK, each obtaining a reward of $1. 120 participants finished the computer experiment (59% male, 41% female; mean age was 34 years; highest education: 48% Bachelor or Master degree; 40% high school degree). Participants were randomly assigned to one of the four conditions (cf. Design).

Procedure and Material Participants in a computer experiment should imagine being in the role of a Human Resources manager evaluating the staff of a particular snack bar. There were five staff members, but in each day only four are working. Compared to many real personnel evaluation situations one with only five persons is relatively simple (cf. von Sydow, 2015, for a plausible inner-individual dilemma with ten nodes). Participants were instructed to establish which workers contributed most to the overall profit of the company, based on data provided by the reporting unit of the larger company.

First, participants in all conditions read the same instructions and overview description. Then, in the main part, they obtained for each day transparent overview information about the individual earnings of each of the four employees (presented by a picture) working in the shift (at this day), and information on the overall earnings (Figure 1).

Figure 1: Example of shown earnings at the individual and group level on a particular day (Condition 1 or 2).

The shown earnings of the normal workers and the altruist worker were based mainly on the mean earnings shown in Table 1. Additionally, we added some noise to each value (a normal distribution with SD = 600 €).

We randomly assigned the altruist-role to one of the pictures. The altruist randomly appeared in 6 of 10 rounds (days). The four normal workers appeared randomly (on average, 7.5 times in 10 rounds; or 15 times in 20 rounds). We counterbalanced the presentation-order in each shift. Participants could view the overview panels for each day as long as they wanted and proceeded to the next round by clicking a button when ready. After the 10 (or 20) rounds, the 'Human Resources managers' evaluated the employees of the snack bar in four tasks (each on a separate page).

First they had to choose which of the five workers was of "the *greatest* total utility for your business". If unsure, they had to choose intuitively. Second, they chose which person represented the lowest utility for the business. Third, they ranked employees according to "their total utility for the company in the present setting" (not reported here). Fourth, they rated the earnings of the five workers (presented randomly) on a scale of 1 to 10.

Finally, they had to provide further comments and demographic data. Additionally, they were tested on the Cognitive Reflection Task (Frederick, 2005) – a short task measuring the tendency to override an initial automatic albeit incorrect response by a reflectively corrected one – and a attention test item checking how careful or superficial people read instructions.

Results and Discussion

Figure 2A shows that, in all conditions, a clear majority of participants did not judge the altruist as representing the highest utility for the company and there were hence no large differences between conditions, the differences were reliable (exact $Chi^2(3, 124) = 3.33$, $p < .05$), with an average of 7% detecting that the altruist had the highest utility in Conditions 3 and 4, with only 1.5% in Conditions 1 and 2. The preponderant selection of normal workers over altruists over all conditions was statistically even more clearly above chance ($Chi^2(1, 124) = 17.8$, $p < .001$). Figure 2B shows that, across conditions, a clear majority of participants assigned even the lowest utility to the one who influenced the overall earnings of the work group the most positively. Given the higher number of normal workers, these judgments are most clearly above chance-level, even for the most critical condition: $Chi^2(1, 32) = 50,8$, $p < .0001$).

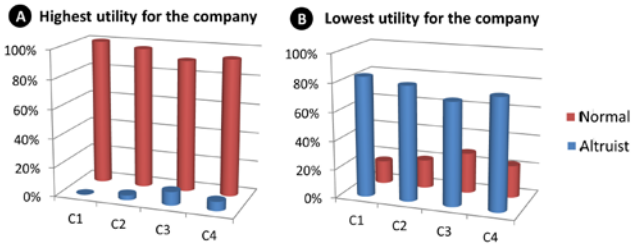

Figure 2: Percentages of choices of normal worker or altruist worker as representing the highest (Panel A) or lowest (Panel B) utility for the company (Condition 1 or 2).

The rating task (Figure 3) likewise showed that, in all conditions, the altruist was predominantly judged to be of lower utility to the company than the normal workers.

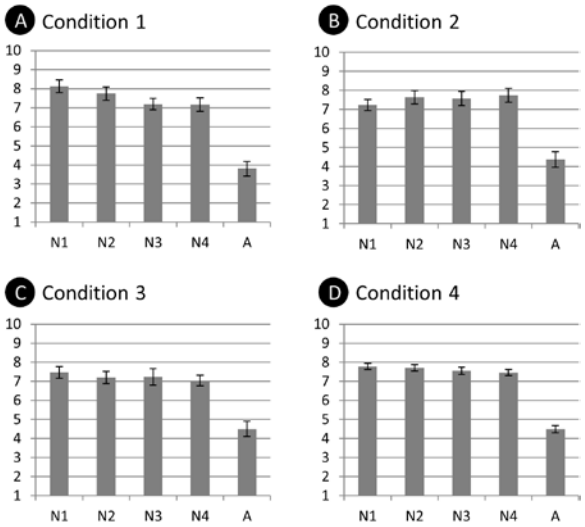

Figure 3: Average ratings (with SE) in Experiment 1 for the four normal workers (*N*) or altruist workers (*A*) of Conditions 1, 2, 3, and 4 (Panel A to D).

Overall, the results show that even if the 'altruist' worker causes a *decidedly* higher increase in the overall earnings than any 'normal' worker (Conditions 3 and 4) and this could be observed over 20 rounds (Condition 2 and 4) the altruist's overall contribution is not reflected by the evaluations of most participants (even though a small significant effect of conditions was found). Actually, the altruist is even evaluated most negatively by the vast majority of participants.

Experiment 2

Experiment 2 used the same personnel evaluation scenario as in Experiment 1, involving an altruist who is individually the lowest earner of the team but overall contributes the

most to the success of the group. However, Experiment 2 addressed several additional issues:

- *Number of days and repeated measurement*. We investigated whether most people's negative evaluation of the group-serving altruist changed, if people had to go through more learning-rounds. Additionally, the participants had to evaluate the workers not only once, but repeatedly.
- *Personnel selection*. We added a personnel *selection* task, wherein the manager determined the configuration of a shift. This should focus people more on the overall earnings of a group, allowing for easier detection of the altruist's highest overall contribution.
- *Selection of participants*. Due to bad results in the attention test, we aimed to rule out effects of unmotivated participants from MTURK. Hence we rigidly selected participants to exclude inattentive participants at the beginning of the experiment.
- *Sensitivity to distinguishing individual earnings*. Finally, we varied the earnings of the 'normal' workers (whose presence did not affect other workers' performance) to investigate whether participants were able to distinguish different individual earnings.

Design

Table 2 shows the varied average earnings of the normal workers in four conditions (C1 to C4). C1 has the same payoff-structure as conditions C3 and C4 of Experiment 1 (with the clearest altruist's contribution). Now in C2 additionally one normal worker (N1) stands out. In C3, N1 and N2 differ from N3 and N4, and in C4 all normal workers differ from one another.

Table 2: Mean earnings of normal workers (N1 to N4), the altruist worker, and the overall earnings, with and without the altruist in the four conditions

		C1	C2	C3	C4
Without Altruist	N1	2000€	2300€	2400€	2600€
	N2	2000€	1900€	2400€	2200€
	N3	2000€	1900€	1600€	1800€
	N4	2000€	1900€	1600€	1400€
With Altruist	N1	3000€	3300€	3400€	3600€
	N2	3000€	2900€	3400€	3200€
	N3	3000€	2900€	2600€	2800€
	N4	3000€	2900€	2600€	2400€
Altruist	(A)	1600€	1600€	1600€	1600€

In all conditions, the altruist's presence strongly affected the other workers' earning and correlates highly and consistently with group's high earnings. Nevertheless, the altruist individually had the lowest earnings. The overall average earnings were kept constant over the conditions.

Method

Participants and strict selection criteria 228 participants from the US checked the first page of the experiment in MTURK; 156 continued and passed a first participation-criterion (time spent on the first page > 20 sec. and < 6 min.). Only 140 (90%) passed Criterion 2, i.e. correctly rephrased the task. Of them, 120 (86 %) finished the experiment (52% male, 48 female; mean age: 33 years; highest education: 59% Bachelor or Master degree; 38% high school degree). The participants were volunteers, obtaining a reward of $1. Participants were randomly assigned to one of four conditions (cf. Table 2).

Material and procedure We used a similar task with a similar procedure and materials as in Experiment 1 (cf. Figure 1). However, instead of 10 or 20 rounds with information about the individuals' earnings and group-earnings, participants here obtained information about 40 working days (again each with four workers). For each day we required participants to stay a minimum time on the information screens (4 sec.), after which the 'continue' button could be pressed. We applied similar counterbalancing measures as in Experiment 1.

Additionally to the changed payoff-structure (Table 2) we now had *four* test-phases, where the workers should be evaluated, with the test-phases following every ten rounds. In the first three of them (following Rounds 10, 20 and 30) we presented only one evaluation task but we added a new personnel selection task. In the former, participants rated the workers' contribution to the group on a scale of 1 to 10. In the new personnel selection task, participants were asked which four of the five employees they would select to work in a further shift the next day. It was mentioned that all five employees wanted to work and that their choice should optimize the profit for the company on that day. In the fourth assessment phase we used all the evaluation measures used in the single test-phase of Experiment 1 as well as the described personal selection task.

Finally, we used (a) a brief Kimchi-Palmer–Test (similar to a Navon-Test) to explore global versus local perception preferences; (b) the attention test; and (c) participants supplied demographic data and comments on the task.

Results and Discussion

Since space precludes exhaustive treatment, we here present only the results for the fourth and final assessment phase of employees by the 'managers'. However, the results of all four phases are surprisingly similar.

Figure 4 presents the results of the rating task over all four rounds. It shows that the order of the average ratings reflects the individual mean differences shown between the normal workers. People seem to be sensitive to these individual differences, but not to the larger overall increase of group earnings occurring when the altruist was in the team. The ratings for the altruist in all conditions and phases were lowest from the first phase onwards. Moreover, the judgments appear quite stable over the time. Even in Phase 4 and the most critical condition, C4, the data was clearly at odds with a potential prediction linked to judgements based on the worker's overall contribution (A > N1 > N2 > N3 >

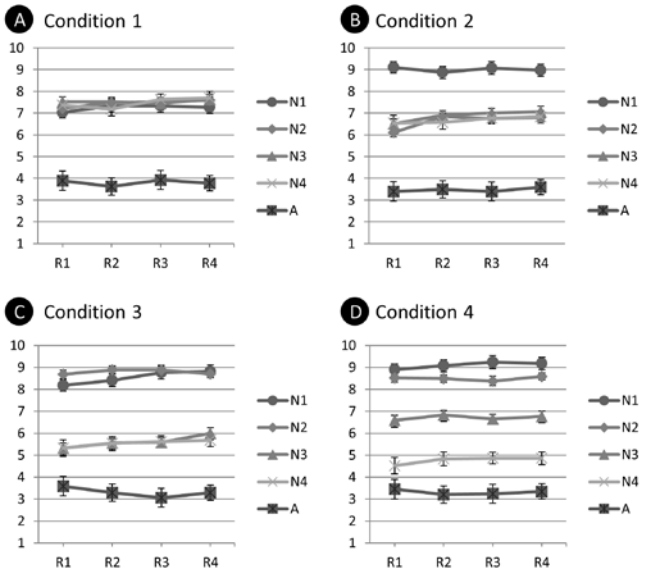

Figure 4: Average ratings (with SE) in Experiment 2 for the four normal workers (*N*) and altruist workers (*A*) in phases (R1 to R4) of Conditions 1, 2, 3, and 4 (Panel A to D).

N4) and contrasts showed significant results for all five mean differences predicted based on *individual* contributions (N1 > N2 > N3 > N4 > A): N1 > N2: $F(1, 28) = 5.5$; $p < .05$; N2 > N3: $F(1, 28) = 96.8$; $p < .001$; N3 > N4: $F(1, 28) = 34.5$; $p < .001$; N4 > A: $F(1, 28) = 37.9$; $p < 001$.

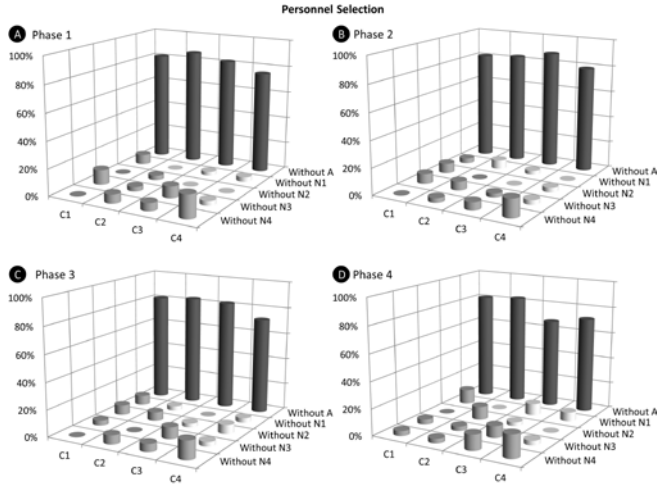

Figure 5: Results of the personnel selection task in the four test phases, showing the proportion of 'managers' choosing a team of four out of five, thus excluding worker N1, N2, N3, N4, or the altruist worker.

The added personnel-selection task (Figure 5) could have elicited better results, since it emphasizes the *overall* earnings of different group configurations. Participants may thus have realized the clearly lower outcome of configurations without the altruist relatively to the four other configurations. But Figure 5 shows that, even in this task, the participants tended to exclude the best player from the team (dark selections). This was quite stable over time. We also marked optimal selections that did not only involve the altruist, but the best normal workers (dark grey shading). But even without this distinction, and in the final Round 4, the altruist were excluded clearly more often (dark shading) than *all* other workers together, $\chi^2(1, N = 120) = 32.0$, $p < .001$), also in the most critical condition, C4 ($\chi^2(1, N = 29) = 5.83$, $p < .05$).

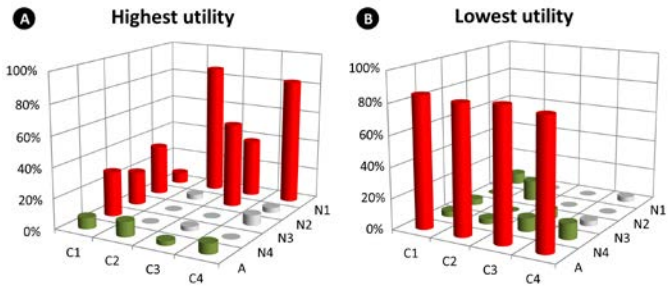

Figure 6: Percentage of 'managers' choosing either a normal worker (*N*) or the altruist worker (*A*) as of the highest (Panel A) or lowest (Panel B) utility for the company (Conditions C1, C2, C3, C4). The choices based on individual earnings are marked in red, those based on overall earnings in green.

Figure 6 presents the proportions of the judgments in the last test phase concerning which workers had the highest and which had the lowest utility for the company. A clear majority in all conditions did not assign the *highest* utility based on the overall impact on the earnings of the group (coloured green), but in line with individual earnings (coloured red) (C4: $\chi^2(1, N = 29) = 24.5$, $p < .001$). In all conditions, the majority of participants also assigned even the *lowest* utility to the 'altruist' (Figure 4B). This was the case in the most unclear condition (C4), since the red choices were significantly more frequent than all other choices taken together ($\chi^2(1, N = 29) = 24.5$, $p < .001$).

However, a small number of participants realized that the altruist should be on the team. This is strongly supported by ten comments of participants which showed an understanding of a difference between individual and overall performance. For instance, Participant 111 wrote: "The ones who sell the most aren't necessarily the top performers. The sales always increased on days when the girl in the green shirt was working although […] she didn't sell as well herself. She seemed to bolster everyone else's sales, making total sales increase by nearly 50%. This was consistent every day she worked, therefore I rated her as extremely important to the company." These participants were *all* among those who selected the altruist in the last round of the personnel selection task (Figure 5) and 80% of them made this selection already in Round 2 (only 30% in Round 1).

General Discussion

In sum, participants in all conditions of both experiments tended to focus on comparing the individuals' earnings without considering their overall contribution to the

earnings of the group. Although Experiment 1 showed a small effect of conditions, the illicit individualist judgments remained dominant both with a high-impact altruist and with 20 (instead of 10) rounds. Experiment 2 showed that participants were well able to distinguish even much smaller *individual* performance differences. Experiment 2 nonetheless corroborated the finding that people tend to wrongly evaluate the group-serving altruist lowest, despite the 40 rounds with four repeated test-phases. Even the personnel-selection task, which might have focused participants more closely on the overall earnings in different configurations of the team, the best team player was most often excluded from the team. As a conclusion, a "tragedy of personnel selection" may occur when people who contribute clearly the most to the overall performance are evaluated the most negatively due to their lower individual contribution.

The outcome was not 'tragic' for *all* participants, since some, but few, participant's selections and comments revealed that they clearly had detected the positive overall effect of the 'altruist'.

Nonetheless, the tragic results in our task with an ecologically rather low or at least common complexity may be due to various explanations. First, the tragedy may be tied to the abstract form of evaluation based on numbers only. But number-based evaluations play an increasing role in companies. Second, the evaluation-context might have elicited a kind of competition or individualism framing, preventing participants from considering helpful behaviour. But we conducted a further egoist-detection personnel evaluation experiment, which led to only partial improvements. Third, the tragedy of personnel selection may well be based on more general and cognitive mechanisms. For instance, it may be related to (a) a general difficulty to realize that many small externalities can add up to large payoffs (e.g., von Sydow, 2015, Dörner, 1993); (b) some general problems of understanding interaction effects (e.g., Novick & Cheng, 2004); or (c) general problems dealing with multilevel representations and the Simpson's paradox (Fiedler, Walther, Freytag, & Nickel, 2003; Waldmann & Hagmayer, 2001; von Sydow, Hagmayer, & Meder, 2016). Whatever the causes are, the practical importance of the problem of detecting group-serving employees and building most efficient teams (based on more than individual excellence) can hardly be overestimated, for both companies and other organizations.

Our findings suggest that there may be a tragedy of personnel selection in the real world as well, with implications for the sensibleness of incentive-structures (personnel evaluation), employee-advancement (personnel promotion) and job offers (personnel selection).

Acknowledgments

This work was 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 for support, to Klaus Fielder, Ulrike Hahn, Sandra v. Sydow, and Michael Waldmann for discussing related issues, and to Martha Cunningham for corrections. The pictures in Fig. 1 are from S. ter Burg, A. Hjartat, B. van Kleef, and T. Arbeider-parti, (creative commons licence; www.flickr.com/photos/ and ter-burg/8172735997, apotekhjartat/13406119904, mvon-ederland/12364694653, tromsarbeiderparti/4577244545). We thank anonymous reviewers for helpful comments.

References

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition, 31*, 187-276.

Dörner, D. (1993). *Die Logik des Mißlingens. Strategisches Denken in komplexen Situationen.* Rowolt: Hamburg.

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. doi: 10.1177/0146167202238368.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162 (3859), 1243–1248.

Fehr, E. & Schmidt, K. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics, 114*, 817–868. doi: 10.1162/003355399556151.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140. doi:10.1038/415137a.

Fehr, E. & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Science, 8*, 185–190.

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. doi: 10.1038/ nature04131.

Ostrom, E., Burger, J., Field, C. B., Norgaad, R. B., & Policansky, D. (1999). Revisiting the Commons: Local Lessons, Global Challenges. *Science, 284* (5412), 278–282.

Henrich, J. et al. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*, 795-855.

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

von Sydow, M. (2006). *Towards a Flexible Bayesian and Deontic Logic of Testing Descriptive and Prescriptive Rules*. Dissertation, Georg-August-Universität Göttingen.

von Sydow, M. & Hagmayer, Y. (2006). *Deontic Logic and Deontic Goals in the Wason Selection Task.* Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society (pp. 864-869). Mahwah, NJ: Erlbaum.

von Sydow, M. (2012). *From Darwinian Metaphysics towards Understanding the Evolution of Evolutionary Mechanisms*. Göttingen: Universitätsverlag Göttingen.

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., Hagmayer, Y., Meder, B. (2016, in press). Transitive Reasoning Distorts Induction in Causal Chains. *Memory & Cognition*, 1-19. 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.