

Errors in Group Performance Evaluation and Individual Effort Provision

An Experimental Investigation

Lucia MARCHEGIANI^{a,1}, Maria Rosario GUIDI^a, Tommaso REGGIANI^b, and Matteo RIZZOLLI^c

^a University Roma Tre

^b University of Cologne

^c Free University of Bolzano, Bozen

Abstract

Supervisors cannot always observe the exact effort provision of each team member, because they can only take notice of the group performance. Though they seek to achieve high effort from the group and from each team member, performance appraisal errors can be committed. A study was conducted to analyze the impact of such errors in the evaluation of team members' effort provision. The individuals' behavior is analyzed, in particular the impact of leniency and severity on the effort provision of best and worst performing team members. Type-I errors (severity) occur when agents are not rewarded though they have provided high efforts, while type-II errors (leniency) are generated when agents are rewarded despite of they have provided lower effort in comparison to other team members. The aim of the research is to examine and discuss some issues of individual behavior in team work with the support of behavioral and managerial theories. The results could be applied to human resources strategies and personnel economics, contributing to the debate about incentives and organizational performance.

Keywords. Performance appraisal errors, Team performance, Evaluation, Motivation, Organizational Citizenship Behavior

Introduction

Organizations are moving toward group-based systems, therefore the importance of research on motivation within groups is increasing (Ambrose and Kulkik, 1999). When group performance is evaluated, both the team and the individual effort provision are important, hence, appraisal systems are ineffective when one of these two ratings is incorrect (Levy and Williams, 2004). Scholars have often been concerned with the difficulties that might occur when allocating reward among team members. We believe it might be useful to study the behavior of team members when the supervisor commits errors in evaluating the members' effort provision. It can also be useful to study whether high performers react differently to the introduction of errors than low performers.

As the success of performance appraisal systems depends from the reaction of the agent (Levy and Williams, 2004), we believe it is interesting to analyze the behavior and the reactions of group members, when errors in evaluating the team occur.

Teamwork requires individuals to cooperate to reach the teams goals, but "the motivation to achieve a collective performance is regarded as derived from individual concerns and motives." When an organization adopts team work to reach some objectives, it has not only the difficulty of rewarding the individuals, but it has also the difficulty of rewarding the group. Most part of the research is focused on the identification of errors and how they can be reduced, but few scholars

¹ Corresponding Author.

studied the effects of errors on agent's behavior. Especially literature on teams and team member appraisal rarely considers errors and weaknesses due to the implementation of evaluation systems based on performance.

As the literature on performance appraisal errors explain, Type I errors are occur when the supervisor under rates the agent's performance, and the high effort of the employee is not rewarded. This is connected to the concept of severity, namely the situation in which a rater is severe in rating the agent's performance. Instead, Type II errors occur when the supervisor over rates the performance of an agent, meaning that the employee perceives a reward though he exerted low effort. This is connected to the concept of leniency, namely the situation in which a supervisor is lenient in his ratings of the agent's performance. The traditional model of law enforcement focusing on average deterrence considers the impact of these two types of errors as symmetric, meaning that deterrence is reduced by Type I error in the same way of Type II error. Instead, this research aims to provide evidence that team member's effort provision is affected in an asymmetric way by Type I and Type II errors.

Free riding impacts negatively not only on the performance of the team, but also on the performance of the organization, therefore we want to analyze how it is influenced by evaluation errors. When some undeserving agents are rewarded their team members feel unjustly rewarded. As shown by the equity theory of Adams (1965) an agent compares his input-outcome ratio with the one of another individual. In case of free riding the employee who is under rewarded might decide to lower his inputs to modify his ratio and obtain the same ratio of the individual who exerts low effort. In this case the productivity of the group is negatively affected.

The remaining of the paper is organized as follows: the next section is a discussion that presents our research propositions, each of which is sustained by previous theories. The experiment and the experimental design are explained, finally the results of the experiments are given and further steps are shown.

Discussion and hypothesis

The compensation systems are studied to reach two objectives, firstly they should motivate individual productivity, and secondly they should keep group harmony. Both goals are relevant and should be accomplished, but some studies regarding distributive justice demonstrate that it is hard to achieve both at the same time. Pfeffer and Langton (1993) analyzed the case when the salary distribution is high and they confirmed that in such situations agents are less satisfied and teamwork is worse. This effect is presumed not occur when reward is based on merit. Motivation to reach team performance is believed to come from individual interests and motives (e.g., Ilgen and Sheppard, 2001), the major source of motivation seems to be the individual reward.

Though appraisal based on performance is the most applied system to reward agents, some scholars suggested that they are at times unsuccessful (Grubb, 2007) and imply some difficulties. Behn (2003) states that performance and effort are difficult to measure, and that these measurements are subjective and prejudiced. Especially in teamwork it is difficult to distinguish the effort of each individual, because the teams outcome results from the cooperation within the group, and unfortunately not always do the supervisors consider the impact of team and organization on the agent's performance. Furthermore, supervisors value agents' behavior subjectively, which could induce agents to become dissatisfied and misdirect their contributions. Finally, it is suggested that sometimes it is difficult to adopt measures based on performance to evaluate the performance of a team, because the performance of a group is strongly affected by other groups' performance (Mohrman et al., 1992).

A traditional way of looking at evaluation accuracy is provided by the standard economic theory of deterrence, which suggests that the punishment of an innocent individual (Type I error) is not worse than the acquittal of a guilty one (Type II error), since both types of errors jeopardize deterrence by the same token (Polinsky and Shavell, 2007; Rizzolli and Stanca, Forth.).

Shavell(1980, 2006) and Shavell and Polinsky (2000) suggested that individual benefits and collective harms depend on individual's decisions on the level of activity and of precaution. Judicial errors decrease and have negative effects on marginal deterrence, therefore they are a focal point of studies on law enforcement (e.g.Png, 1986; Polinsky and Shavell, 2000). Scholars distinguish between two types of errors, first is Type I error when over-enforcement occurs and an enforcer finds an innocent person guilty, it is also referred to as false positive. Second is Type II error when the enforcer absolves a guilty person and under-enforcement occurs, it is also referred to as false negative.

Literature suggest that performance appraisal errors are committed intentionally (when evaluators are dishonest and inaccurate for political reasons to defend and reinforce their interests) or unintentionally, e.g. leniency error, severity error, central tendency and range restriction errors, halo and horn errors, recency error, fundamental attribution error, self-serving bias (Grubb, 2007).Inaccurate evaluations lead to mistrust of agents towards future appraisals and towards the evaluator's capabilities (Sholtes, 1987). Furthermore, when appraisal inaccuracy generates under evaluated employees, these may feel discouraged and guilty, instead over-rated agents may held undeserved personal recognition.

Performance appraisal research has focused on the agents motivation, such studies either search connections between performance appraisal and compensation, or they investigate on appraisal factors that increase agents' motivation. Pettijohn et al. (2001a, 2001b) investigated whether performance appraisal can be adopted to improve sales-force satisfaction and commitment at work. They found connections between participation and perceptions of fairness, which are important for agents to feel satisfied and be commitment to reach the organizations main goals. Their results also show that systems of performance appraisal can be implemented to rise agents' job satisfaction, organizational engagement, and work motivation. Other studies suggested that participation and justice are essential to the motivational function of a performance evaluation system (e.g. Bartol, 1999). Bartol (1999) suggested that Agency Theory-based compensation systems affect perceptions of objectives, compensation, and justice, which influence agent's levels of satisfaction and commitment and the purposes of performance and turnover.

RP1: In presence of evaluation errors (Type I and Type II errors), the team member's motivation to work diminishes and effort provision is lower.

As mentioned previously, the use of team work is pervasive and indispensable, therefore it is crucial to determine individuals' motivators and demotivators within the group contexts (Karau and Williams, 1993). One might think that team members should feel inspired by working with others, and hence, they should "maximize their potential and work especially hard." It is important, for practical and theoretical reasons, to determine the conditions under which team members tend to exert less effort.

The need for appraisal and the refinement of multi-source rater accuracy improved (Valle and Davis, 1999), as organization rely always more on group-level work (e.g. Katzenbach and Smith, 1993). Valle and Davis (1999) suggested that improving the reliability and validity of group performance appraisal could be considered as a method to show both accuracy and fairness. Furthermore, the more reliable and "content valid" are the tools, the more confident can be the organization to use them when making important work force decisions.Rating accuracy ought to be improved until there is a direct relation between the "metrics" and the desired outcomes (e.g. enhanced productivity, performance).

Equity theory anticipates that not rewarding employees in accordance with their contributions undermines performance (e.g. Folger, 2001). Equity theory (Adams, 1965) distinguishes between negative and positive inequity. Negative inequity happens in situations such as our Type I error: when individuals exerting high effort are not deservingly rewarded. In T2 all subjects exerting high effort are rewarded. Higher effort provision (as compared to T1) can still be

explained by equity theory in terms of positive inequity. Perception of unfairness persists when unfair distributions of outcomes are in favor of the employee.

The research of Rizzoli and Stanca (Forth.) has shown that deterrence is mostly affected when judicial errors are incremented individually instead of jointly. Crime increments when the probability of a judicial error increases. Both types of errors impact deterrence, but the effects aren't symmetric, type-I errors (the conviction of an innocent) are more detrimental than type-II errors (the acquittal of a guilty individual).

There is also a prolific stream of research in behavioral economics dealing with inequity aversion, described as a recurring preference for fairness and aversion to distributive inequality.

RP2: Type I errors and Type II errors have asymmetric impacts on individual effort provision in team work.

Scholars demonstrated that the behavior of individuals performing in teams, may be different than behavior they have when they carry out the task alone, and, therefore, the group performance is not necessarily the simple sum of the individual performances in case employees had worked alone (Levine and Moreland, 1998). Unfortunately, the adoption of teams has the potential to generate problems with their structures as well. Free riding is a typical behavior that could occur when employees work in teams, it has negative effects on the team itself and on the organization as a whole. As proposed by literature, free riding diminishes the group's performance, because the free rider puts less effort to reach the goals of the team.

The free riding effect is analyzed by Fuster and Meier (2010), the scholars made a research in which team members were questioned about their anger and frustration against the free rider on a seven-point scale (1 = "not at all" to 7 = "very much"). The results of the survey show that anger intensifies with the divergence between the contributions of the free rider and the ones of other team members. Also the experiment generated such results, group members tended to punish the free rider harder when they invested much more than him, instead when the divergence was smaller, they tended to be less angry.

Incentives can have negative effects on the norm enforcement mechanism. When incentives are distributed centrally and there is no decentralized norm enforcement, the incentives have positive impact, but when norm enforcement is relevant and powerful, the effects on "prosocial behavior" can be negative. This is because incentives that are proportional to reward induce team members to punish free riders in a less severe way. Furthermore, the behavior of free riders is less affected by punishment, when monetary incentives are given. The scholars invite managers to be cautious when adopting private incentives to encourage agent's behavior in a context in which norm enforcement and social pressure matter. Other studies show that the use incentives can be risky, the agent's motivations could be reduced to behave in the interest of the organization and not on their own (Deci 1975, Frey 1997). It is also suggested that, in a principal-agent relationship, the effects of adopting incentives could affect trust by diminishing it (Fehr and Falk 2002, Fehr and List 2004, Falk and Kosfeld 2006).

RP3: The free riding effect increases when a supervisor commits evaluation errors (Type I and Type II errors), especially when he commits a Type I errors.

Procedures for collecting data

Laboratory experiments have been successfully designed to test personnel issues, adding valuable insights to the existing body of knowledge. (e.g. Brügger and Strobel, 2007; Charness and Kuhn, 2010), Falk and Kosfeld (2006) show that not imposing a minimum effort level or allowing workers to delegate leads to better outcomes. On the other hand, Schnedler and Vadovic (2011) show that the "hidden costs of control" are mitigated if the principal's control is legitimate. In particular they show that when agents are given an initial endowment they do not resent effort requirements that simply prevent them from "pilfering" the endowment.

The use of a lab experiment to test our theoretical predictions provides several important advantages (Charness & Kuhn, 2010) in comparison with observational datasets that are typically used in labor/personnel economics or managerial case studies; above all the opportunity to control for all the crucial variables of the economic environment and the possibility to vary ad-hoc the precise variables of interest. On the other hand the external validity of lab findings can be questioned. However, the research question of the present work deals with a variable - evaluation errors - that is basically impossible to measure in the field because effort and the stochastic relation between performance and effort are unobservable. In the lab instead we can superimpose an exogenous probability of error in evaluating performance and at the same time we can perfectly observe effort. This ideally allows us to identify precisely the impact of errors on effort provision and thus on performance.

Experimental design

The experimental design is made up of three phases: the preliminary Phase I is used to elicit individuals' risk attitudes via a standard (Holt and Laurie, 2002) incentivized choice of lotteries. This is followed by Phase II, where individual productivity and attitude to free riding in the default task is measured, and then there is Phase III, where individuals can choose whether to carry out the task. This last phase is our actual main treatment phase. Each individual carries out the task at most no more than once but choices are elicited under all three treatments conditions. The experiment adopts a within-subjects design (that in its actual implementation is very close to the strategy method elicitation mechanism). To control for any possible ordering effect, the order in which subjects are asked to make their choices under Treatments 0, 1 and 2 is randomized across the different experimental sessions. Feedback information on the outcomes of the lotteries in Phase I and on whether the supervisor-automata makes an observational mistake are provided at the end of the experimental session. This is to assure full independency of the different treatments phases, free of historical contagion and therefore it is statistically independent across all subjects.

Between each of the phases, subjects have the opportunity to rest. Common instructions for the subsequent phase are read and described aloud while instructions concerning each treatment are delivered on screen, Control questions for each of the different phases are administered through the computer.

For the experiment students are randomly grouped in teams of four, participants are given a number of Experimental Currency Units equivalent to 10 € each. As in public good games, individuals have to decide how much of these 10 € they want to contribute to the team's good. Every member receives a quarter of the total prize, P , reached by the team. In the scenario without error, each of the two members exerting more effort receives $\frac{1}{4} P * 1.2$, meaning that he will receive a part of the team's total prize and a bonus of 20 % of P . Instead, each of the members exerting less effort receives $\frac{1}{4} P * 0.8$, which means that he will receive a part of the team's prize minus 20% of P . Three scenarios are considered, in the first errors are not committed, and higher performers are remunerated $\frac{1}{4} P * 1.2$, while the lower performers are remunerated $\frac{1}{4} P * 0.8$. In the second scenario the supervisor commits an error with 80% of probability, not recognizing high effort, a high performer the bonus he deserved. In the third scenario, the supervisor commits an error with 80% probability, he recognizes a bonus of 20% to an agent who performed poorly.

Discussion of the results

Before running the actual experiments sessions, we ran two pilot sessions of the experiment, through a strategy method. They already indicated that the three propositions find support. More in detail, it is interesting to notice that evaluation errors do exacerbate the free riding attitude, as non free riders in T0 tend to become free riders when exposed to evaluation errors. Moreover, the Type-II errors seem to have a higher negative impact on individuals and thus it seems to be more detrimental in team work (Fig. 1).

The results support the hypotheses. In particular, as per RP1: effort diminishes in presence of evaluation errors, this is straightforward confirmed.

As far as RP2 is concerned, our findings do show that free riding exacerbates under the scenarios with errors, since the correlation between free riding and the drop in the individual contribution are: $r_0=.02$; $r_S=-.3$; $r_L=-.4$.

Furthermore, we found it interesting that there seems to be a gender effect. In fact, males tend to drop their contribution more under T2 rather than T3, whereas female tend to do the opposite. This may have roots in the previous results (RP2), as it could be related to the sense of justice and to free riding.

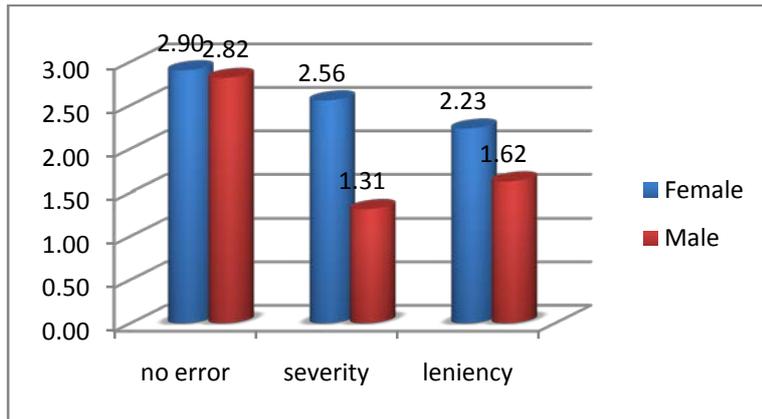


Figure 1: Average contribution under the three scenarios – pilot sessions

As far as the actual experiments are concerned, we ran three experimental sessions involving 65 participants. The experimental design resembles a strategy method, thus allowing for sufficient observations. The first phase was devoted to assess the risk aversion and the loss aversion of participants. This is useful to verify that participants' choices are not related to their risk aversion. Since the correlations of the choices taken in the first phase and the allocation choices in the second phase are very low, (.12 in the scenario without errors, -.07 in the scenario with type 1 errors and .01 in the scenario with type 2 errors) it can be concluded that participants' allocations do not depend from their risk aversion.

Figure 2 shows the average contributions under the three scenarios.

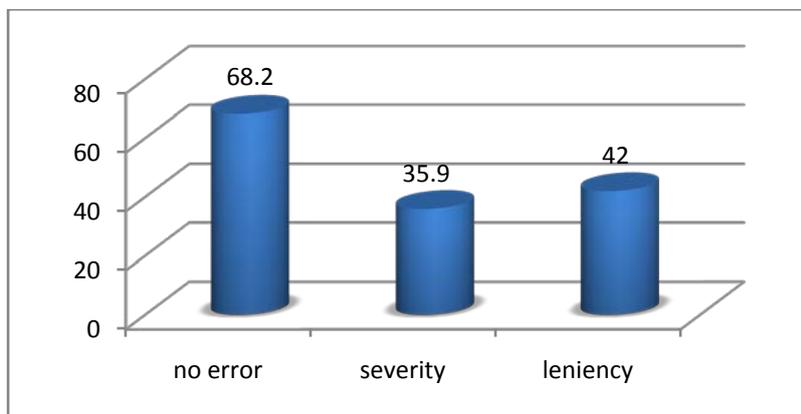


Figure 2: Average contribution under the three scenarios – experimental sessions

The final results confirm the outcomes of the pilot session, as the average contribution drops from T0 to T1 (-47%) and from T0 to T2 (-38%), providing further support to our RP1.

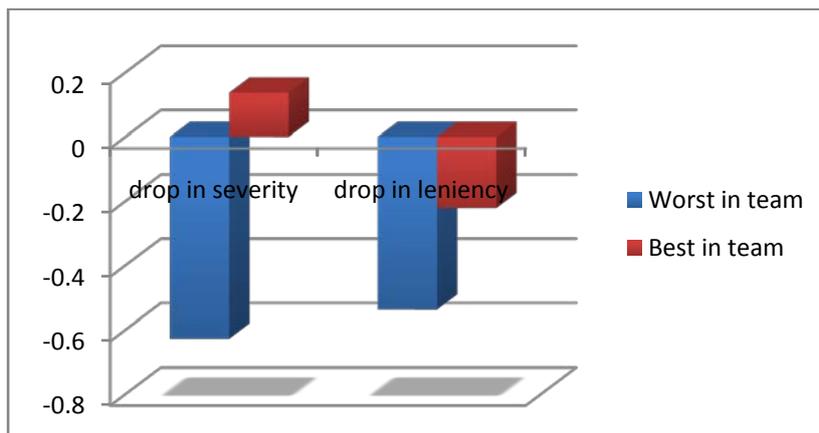
We find also support to RP2 as in the leniency scenario the average contribution is higher than in the severity case (+17%).

It must be said that the final results did not provide strong enough support to the gender effect that was hypothesized on the basis of the results related to the pilot sessions.

Moreover, if we consider the correlation between the contribution to team performance and the free riding attitude (which we assessed in the final phase of the experiment) there is a significant difference between the three scenarios. In fact, in absence of errors, the correlation is almost null, whereas in the severe and in the lenient treatment the coefficient lowers to -.5 and -.4 respectively. This indicates that the evaluation errors exacerbate the free riding attitude, and this gives support to our RP3.

We define the best performers as those team members who contributed the most to their team, whereas the worst performers are those who contributed the least. Taking into account the scenario with no errors and the ones with leniency and severity errors, there are significant differences in the best performers' vs. worst performers' behaviors. Figure 4 shows these strong dissimilarities in behavior under the different scenarios. On the basis of these results, we conclude that leniency and severity errors have an asymmetric impact on the behaviors of the individuals that tend to achieve higher results rather than those who tend to perform poorly.

Figure 3: Average drop in individual contribution to the team performance – best and worst performers



When we consider the best performers, they tend to contribute more under the severity scenario (+14% with respect to the average contribution with no errors); whereas they contribute less under the leniency scenario (-22%). At odds, the worst performers tend to lower their contribution more under the severe scenario (-63%) rather than under the lenient scenario (-53%).

We found explanations in the personality traits, as we assessed them through the 10-items Big Five personality measure (TIPI) (Gosling *et al.*, 2003)². Not surprisingly, emotional stability is negatively correlated to the willingness to contribute under the evaluation errors, especially in the severity scenario ($r=-.22$ in the severity scenario; $r=-.13$ in the leniency scenario). On the contrary, openness to new experience is positively correlated to the individual contribution under the severity scenario. Emotional stability is also positively correlated to the group commitment ($r=.26$). Consciousness indeed seems to be negatively correlated to group commitment ($r=-.17$), but it is also negatively correlated to free riding ($r=-.34$).

²Gosling *et al.* (2003) proposed a 10-item measure of the Big Five dimensions for situations when very short measures are needed, personality is not the primary topic of interest, or researchers can tolerate the somewhat diminished psychometric properties associated with very brief measures.

Conclusions

The experiment finds strong support for the existence of an asymmetric impact of errors on agents' willingness to exert high effort. In particular an agent exposed to evaluation errors is more sensitive to Type I errors. This result is not predicted by the theory even when considering risk aversion. Further work is needed to explain the result, which seems to be robust against some preliminary treatments manipulations. From a theoretical perspective, the experiment sheds new light on the relation between reward systems and motivation that should inform agency theory, organizational behavior and personnel economics. From an organizational perspective, our result expands the notion of organizational justice: Departures from the just treatment can be both advantageous and disadvantageous even in absence of distributional implications with third parties as assumed by equity theories; subjects react differently when they suffer (enjoy) disadvantageous (advantageous) injustice. Our results can also be interpreted as further evidence of gift-exchange behavior in the indulgent treatment and negative reciprocity in the mean treatment.

International and managerial implications

The experiment finds strong support for the existence of an asymmetric impact of errors on agents' willingness to exert high effort. Further work is needed to explain the result, which seems to be robust to some preliminary treatments manipulations. From an organizational perspective, our results expand the notion of organizational justice: injustice can be both advantageous and disadvantageous without distributional implications with third parties as assumed by the previous equity theory: subjects react differently when they suffer (enjoy) disadvantageous (advantageous) injustice. Our results can also be interpreted as further evidence of gift-exchange behavior in the indulgent treatment and negative reciprocity in the mean treatment.

Although the experiment has limited external validity it may have implications in real world contexts. Since intangibles are increasingly important in business organizations and knowledge-intensive jobs are difficult to assess, errors in evaluating the employees' performance may well be a relevant phenomenon. Our research seems to suggest that, when a perfect assessment of employees' outcome is not viable, it may be wise for the supervisor to be cautious when neglecting rewards and -in general- have a pro-employee bias in conducting her assessment as this will be beneficial for longer term employees' motivation and effort provision.

References

1. Ambrose, M. L., & Kulik, C. T. 1999. Old friends, new faces: Motivation research in the 1990s. *Journal of Management*, 25: 231–292.
2. Levy, P. E., & Williams, J. R. (2004). The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management* 30: 881
3. Adams, J. S. (1965) Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299). New York: Academic Press.
4. Pfeffer, J., & Langton, N. (1993). The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly*, 38, 382–407.
5. Ilgen, D. R., and Sheppard, L. 2001. *Motivation in work teams*. In M. Erez, U. Kleinbeck, and H. Thierry (Eds.), *Work motivation in the context of a globalizing economy*: 169–179. Mahwah, NJ: Lawrence Erlbaum Associates.
6. Grubb, T. (2007), Performance Appraisal Reappraised: It's Not All Positive. *Journal of Human Resource Education* 1 Vol. 1, No. 1, 1-22.
7. Behn, R. O. (2000). Performance, people, and pay. Bob Behn's Public Management Report.

8. Mohrman, A. M., Mohrman, S. A. and Lawler, E.E., (1992), *The performance management of teams*. In B. Burns (ed.), *Performance Measurement, Evaluation and Incentive*: 217-241. Boston: Harvard University Press.
9. Polinsky, A. Mitchell and Steven Shavell (2007), *The Theory of Public Enforcement of Law*, in A. Mitchell Polinsky and Steven Shavell, eds., *Handbook of Law and Economics*, Amsterdam: Elsevier.
10. Rizzolli, Matteo and Luca Stanca, 2012, Forthcoming. *Judicial errors and deterrence: theory and experimental evidence*. *Journal of Law and Economics*.
11. Shavell, Steven (1980), *Strict Liability versus Negligence*, *The Journal of Legal Studies*, 9, 1-25.
12. Shavell, Steven (2006), *Do Excessive Legal Standards Discourage Desirable Activity?*, Harvard John Olin Discussion Paper n. 540.
13. Polinsky, Mithcell A. and Steven Shavell (2000), *The economic theory of public enforcement of law*, *Journal of Economic Literature*, 38(1), 45-76.
14. Png I.P.L. (1986), *Optimal subsidies and damages in the presence of judicial error*, *International Review of Law and Economics*, 6, 101-105.
15. Sholtes, P. R. (1987). *An elaboration on Deming's teachings on performance appraisal*. Madison, WI: Joiner Associates.
16. Pettijohn, C., Pettijohn, L. S., Taylor, A. J., & Keillor, B. D. 2001. *Are performance appraisals a bureaucratic exercise or can they be used to enhance sales-force satisfaction and commitment?* *Psychology & Marketing*, 18(4): 337–364.
17. Pettijohn, C. E., Pettijohn, L. S., & Amico, M. 2001. *Characteristics of performance appraisals and their impact on sales force satisfaction*. *Human Resource Development Quarterly*, 12(2): 127–146.
18. Bartol, K. M. 1999. *Reframing salesforce compensation systems: An agency theory-based performance management perspective*. *Journal of Personal Selling & Sales Management*, 19(3):1.
19. Karau, S. J., and Williams, K. D. 1993. *Social loafing: A metaanalytical review and theoretical integration*. *Journal of Personality and Social Psychology*, 65: 681–706.
20. Valle, M. and Davis, K., (1999). *Teams and performance appraisal: Using metrics to increase reliability and validity*. *Team Performance Management*, Vol. 5 Iss: 8 pp. 238 – 244.
21. Katzenbach, J.R. and Smith, D.K. (1993). *The Wisdom of Teams*. Harvard Business School Press, Boston, MA.
22. Folger, R. (2001). *Justice as deonance*. In S. W. Gilliland, D. D. Steiner, and D. P. Skarlicki (Eds.), *Research in social issues in management* (Vol. 1, pp. 3–33). New York: Information Age Publishing.
23. Levine, J. M. and Moreland R. L. (1998) *Small groups*. In: Gilbert, D. T., Fiske, S. T., Lindzey, G. (eds) *Handbook of social psychology*, vol 2, 4th edn. Pp. 415-469.
24. Fuster, A., Meier, S. (2010) *Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement*. In *Mnagementscience* Vol. 56, No. 1, pp. 57–70.
25. Deci, E. L. (1975). *Intrinsic Motivation*. New York: Plenum.
26. Frey, B. S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. Edward Elgar, Cheltenham, UK.
27. Fehr, E., A. Falk. 2002. *Psychological foundations of incentives*. *Eur. Econom. Rev.* 46(4–5) 287–324.
28. Fehr, E., J. A. List. 2004. *The hidden costs and returns of incentives—Trust and trustworthiness among CEOs*. *J. Eur. Econom. Assoc.* 2(5) 743–771.

29. Falk, A., Kosfeld, M., 2006. *The hidden cost of control*. Amer. Econom. Rev. 96(5) 1611–1630.
30. Alexander Brügger, A. & Strobel, M., 2007. Real effort versus chosen effort in experiments. *Economics Letters* Vol. 96, Issue 2, pp. 232–236.
31. Charness, G. and P. Kuhn, “Lab labor: What can labor economists learn from the lab?,” NBER Working Paper, 2010.
32. Falk, A., M. Kosfeld. 2006. The hidden cost of control. Amer. Econom. Rev. 96(5) 1611–1630.
33. Schnedler, W. and Vadovic, R. (2011). Legitimacy of control. *Journal of Economics & Management Strategy*, Volume 20, Issue 4, pp. 985–1009.
34. Holt, C.A. and S.K. Laury, (2002) “Risk Aversion and Incentive Effects,” *The American Economic Review*, 2002, 92 (5), 1644–1655.
35. Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality*, 37, 504-528.