Bonn Econ Discussion Papers

Discussion Paper 12/2008

Whom will you choose? - Collaborator Selection and Selector's Self-Prediction

by

Marion Eberlein, Judith Przemeck

June 2008

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net
MAIL EXPRESS LOGISTICS FINANCE

Whom will you choose?

- Collaborator Selection and Selector's Self-Prediction

Marion Eberlein and Judith Przemeck*

June 2008

Abstract

Our experiment investigates managers' self-predictions of their subsequent performance and, based upon, their choice of a collaborator. Our results show that managers' self-predictions are not biased anymore after they are informed about the performance of a reference group. In spite of this, most managers do not rationally choose a collaborator given their beliefs. In a second treatment, superiors (who are assumed to be at a higher hierarchy level than the managers) obtain various information, e.g. about managers' self-predictions, and have to predict the managers' performances. Our data show that superiors adapt their predictions into the direction of the managers' self-predictions, although not completely. Particularly, superiors think that their managers' self-predictions are biased if they are lower than the average performance of the reference group. Based upon their predictions, superiors have to select a collaborator for their managers. We find that superiors' collaborator choices do not significantly differ from the managers' choices. This proves due to excellent information processing by both, managers and superiors, which on the whole leads to very similar predictions of managers' subsequent performance.

JEL Classification:

M5, C91

Keywords:

Employee selection, self-prediction, overconfidence, experiment

Marion Eberlein
University of Bonn
Department of Economics
Adenauerallee 24-42
53113 Bonn
marion.eberlein@uni-bonn.de
Phone ++49 228 739212

Judith Przemeck
University of Bonn
Department of Economics
Adenauerallee 24-42
53113 Bonn
judith.przemeck@uni-bonn.de
Phone ++49 228 739213

* We thank Jürgen Arns, Thomas Dohmen, Simon Gächter, Bernd Irlenbusch, Johannes Kaiser, Matthias Kräkel, Dirk Sliwka and the participants of the Brown Bag Seminar at the University of Cologne, Erik Theissen, Klaus Utikal, and Gari Walkowitz for valuable comments and discussions. Financial support from the Deutsche Forschungsgemeinschaft (DFG), grant KR 2077/2-3 and IR 43/1-1, is gratefully acknowledged.

1. Introduction

Hiring of new employees is an important task for managers of corporate enterprises. The principal wants his managers to engage the most competent collaborators because they produce the highest output. Sometimes it can, however, be observed that managers not necessarily choose the most qualified collaborators.[1]

In the theoretical economic literature there are some studies that explain why it may not be rational for managers to choose the most qualified candidate. One possible explanation is presented by Friebel and Raith (2004). They argue that a manager makes every effort to avoid becoming replaced by a more qualified collaborator if his ability is directly compared to that of the collaborator. Therefore, the manager uses his authority to recruit strategically. Nevertheless, he takes into account the underlying disadvantage: Recruiting a less qualified collaborator increases the risk of a low team output, which negatively affects his further career.

Glazer and Segendorff (2005) assume that a manager is interested in credit claiming for successful work. They show that choosing the worst candidate can be an equilibrium strategy. Segendorff (2000) demonstrates that, under certain circumstances, it can be rational for a competent risk-averse manager to choose an incompetent employee in order to have a scapegoat in case of failure. According to Beniers (2005), a further reason for choosing a less competent collaborator can be loyalty: Very competent collaborators may have good side opportunities and so are usually less loyal.

Furthermore, the so-called anti-herding-literature deals with situations in which a person wants to signal its own (high) competence. In our context mainly Levy's article of 2004 has to be mentioned. In her model a decision-maker can be supported by a consultant. She shows that the most able decision-makers do not confer with a consultant in order to signal their excellent ability.

Besides the already mentioned aspects, there are numerous other reasons for a manager to hire a less competent collaborator. It can often be observed that a manager has to choose a collaborator for his department to undertake a project but has only little experience in the field of that project. Thus he does not know if he will perform well. If he believes that his achievement is not noticeable, he may tend to choose a less qualified collaborator.[2] In particular, the professional reputation he enjoys in his department may be lessened if he

[1] This is for example the case when managers think that they do not deserve their position (see Rodriguez-Bailon et al. 2006).

[2] Choosing a less qualified collaborator means choosing one whose expected outcome is lower than the expected outcome of the manager.

chooses a candidate who is more qualified than he himself. He even may be replaced. Therefore, the manager's self-prediction and the conditions influencing his self-prediction play a significant role in such a hiring process. It is plausible that a collaborator who is better than the manager will have better prospects for his career if the superiority in ability can be identified in the long run by a principal or the firm owners. Thus, choosing a less qualified collaborator may prevent the manager from decreasing authority and from loosing his position in the worst case. In such a case, the principal probably anticipates this inefficient behavior of the manager. One possibility to reduce this inefficiency is to offer the manager incentives to choose the best collaborator. This can be done by organizing a tournament between the departments of a firm in order to stimulate outstanding working results. Then an award is given to the manager's department for obtaining a high output or a good achievement compared to all other departments of the firm. In that case, a manager has to take into account all aspects that influence the chance of winning the tournament when choosing a collaborator: First, the abilities of the other departments' managers, second, which collaborators these will engage, and, third, his own expected ability and his chosen collaborator's ability.

To sum up, a trade-off can be observed in the process of employing a new collaborator: On the one hand, an exceptionally good collaborator increases the output of a manager's department and therefore increases the probability of winning a tournament prize. On the other hand, a manager's superiority and even his position may then be at stake. If this is of higher importance for the manager, the manager should decide in favor of a less qualified candidate. However, a manager chooses a collaborator before the project is undertaken. Thus he has to predict his own ability. If a manager, who *overestimates* his ability, wants to employ a less able staff, he may unintendedly hire one who is more able than he himself. If a manager, who *underestimates* his ability, wants to employ a less able staff, it may happen that he engages a much too bad one. If he knew his real ability, he might choose a better collaborator (although still worse than he himself). Besides, it is also important how certain the manager is of his self-prediction. If he is not sure of the accuracy of his self-prediction and wants to hire a less qualified collaborator, he may choose a much less qualified one to ensure in any case that his prospective performance is better than the performance of his collaborator.

Our experiment investigates managers' self-predictions of their subsequent performance and, based thereupon, their choice of a collaborator. To our knowledge, the influence of self-prediction on collaborator choice has not yet been investigated. Next to this aspect, our experiment examines if the selection behavior differs if a superior (who is assumed to be at a higher level of the hierarchy than the manager) chooses a collaborator for a manager, based

upon his prediction of the manager's performance. By informing the superior of the manager's self-prediction, we can investigate how a superior evaluates this information. It is not clear whether a superior takes a manager's self-prediction for sure. According to psychological and economic literature on overconfidence, managers often state too high self-predictions.[3] If a superior anticipates this, he will predict the manager's performance to be lower than the manager's self-prediction. As a consequence, a superior may select another collaborator for the manager than the manager would do.

Regarding the results of our experiment, our data show that managers' self-predictions are not biased anymore after they are informed about the performance of a reference group. About 35% of the managers choose a collaborator who is worse than the self-prediction of their subsequent performance. Moreover, most managers do not rationally choose a collaborator given their beliefs. Concerning our second treatment, our data reveal that superiors adapt their predictions into the direction of the managers' self-predictions, although not completely. They adjust their predictions but keep their former prediction in mind and do not deviate too much from it. Interestingly, superiors think that their managers' self-predictions are too low if they are lower than the average performance of the reference group. We find that superiors' collaborator choices do not significantly differ from the managers' choices. This proves due to excellent information processing by both, managers and superiors, which on the whole leads to very similar predictions of managers' subsequent performance.

The question whether subjects know that others are overconfident has already been investigated by Cesarini et al. (2006) and Ludwig and Nafziger (2007). Cesarini et al. (2006) observe that subjects anticipate others' overconfidence. Ludwig and Nafziger (2007) also show this but only for an environment in which subjects are very familiar with the task. These studies considerably differ from ours because we investigate performance *predictions (before* the task has been carried out) and their evaluation by a third person. In Cesarini et al. (2006) subjects have to answer numerical questions. They have to state a lower and an upper limit for each question so that they think the stated interval actually contains the true answer to the question with a probability of 90% (confidence interval estimation). Moreover, subjects have to assess the other subjects' average accurateness, i.e., they have to assess in how many of the stated intervals the right value is on average contained. Because subjects assess that the number of these intervals are, on average, less than 90% of all stated intervals, Cesarini et al.

[3] The phenomenon of overconfidence has extensively been investigated in psychology (e.g. Alpert and Raiffa 1969, Svenson 1981, and Weinstein 1980). In the field of economics there is also a growing literature which deals with overconfidence empirically (e.g. Camerer and Lovallo 1999, Fellner et al. 2004, Russo and Schoemaker 1992) and theoretically (e.g. De Long et al. 1991, Daniel et al. 1998 or Gervais and Odean 2001).

(2006) conclude that subjects anticipate that others are overconfident. This approach is completely different to ours. In our experiment managers have to state how many multiplication tasks they will solve correctly. If, after being informed about the managers' self-predictions, superiors systematically predict a lower performance for the managers than the managers themselves, we say that the superiors think that managers are overconfident. In Ludwig and Nafziger (2007), subjects are informed about the average assessment of some other subjects and have to evaluate if these subjects are on average over- or underconfident. Then, they have to assess the average performance of these subjects. Contrarily, in our experiment superiors have to predict the managers' performance *before* they get to know their self-predictions as well as *afterwards*. Thus we can exactly investigate superiors' information processing and whether they expect that managers are overconfident or not. This is also a difference to Cesarini et al. (2006), where subjects do not get any information about the subjects they have to assess. Furthermore, in contrast to the other experiments superiors get to know the self-prediction of a certain associated manager in our experiment. Ludwig and Nafziger (2007) also run a treatment in which subjects assess the bias of a certain subject, but they use a completely different approach. In their treatment, subjects have to evaluate for each possible individual assessment (strategy method) whether a subject, who hypothetically states one of these assessments, is over-, underconfident or unbiased. In our experiment, superiors know that their managers really made a certain self-prediction. Therefore they more intensively think about the accurateness of this particular value when they assess the manager's performance. Next to these aspects, we go one step further than Cesarini et al. (2006) and Ludwig and Nafziger (2007) in examining a decision (the collaborator choice), which is based upon the performance predictions.

The remainder of the paper proceeds as follows. Section 2 describes the design of our baseline treatment and the design of the two main treatments. Section 3 presents our results of the baseline treatment, section 4 the results of the first main treatment, and section 5 the results of the second main treatment. In section 6 the results of both main treatments are compared, while the final section concludes.

2. Procedure, Experimental Design, and Hypotheses

Our experiment was conducted at the Bonn Experimental Economic Laboratory from September to December 2007. Subjects were recruited via the internet by using ORSEE software (Greiner, 2003) announcing the possibility to earn an amount of money dependent

on their behavior. 20 subjects took part in the baseline treatment and 17 in each of the two main treatments. One session lasted 70 minutes on average and subjects earned approximately 12 €.

2.1 Design of the Baseline Treatment (BT)

The BT is conducted to find a typical reference group that will be used for comparison in the two main treatments. The BT is divided into two distinct stages: At stage 1, subjects carry out simple multiplications. At stage 2, they have to choose a hypothetical collaborator and to assess the collaborator choices of the other subjects in their group.

Stage 1:

Subjects get 10 minutes of time to solve simple multiplication tasks. They are only allowed to use pen and paper. For each correctly solved task, a subject obtains 4 Cents. Thereafter, the subjects are arranged into groups of four and are told how many multiplications they themselves solved correctly.[4] Moreover, they are informed about the number of correctly solved tasks of the other three persons in their group.

Stage 2:

At stage 2, every subject chooses a hypothetical collaborator with an integer result from the interval [0,110]. Thus subjects can choose between numerous collaborators who differ in the number of hypothetically correctly solved multiplications ranging between 0 and 110.[5] As it seems impossible to solve 110 tasks correctly in 10 minutes, a chosen hypothetical collaborator with the result of 110 is better than all the subjects in our experiment. This is important for our experiment, because we are mainly interested in the question whether subjects choose a better or a worse collaborator compared to their own performance.

Each subject and his chosen collaborator form a team. The selection of a collaborator affects the team result and the subject's payoff. The team result is defined as the sum of the subject's result in the task and that of the hypothetical collaborator. Before subjects select a collaborator, they are asked which (hypothetical) collaborator they expect to be chosen by the

[4] Before dividing subjects into groups, we first determine their results (i.e. the number of correctly solved tasks) and the distribution of the results. Then we assign the subjects to quartiles according to their results. Afterwards we divide the subjects into five groups so that each four-person-group has exactly one member of each quartile. This procedure is used to make sure that each group is "typical".

[5] Note that each subject always gets his desired collaborator so that different subjects may choose collaborators with the same result.

other three members of their group.[6] Afterwards they choose a collaborator themselves. Thus, each four-person-group forms four teams, each team consisting of a subject himself and his chosen collaborator.

A subject obtains 5 € + 1 Cent · CR if he chooses a collaborator who has a *lower* result than he himself (with CR denoting half of the selected collaborator's result). In contrast, a subject is paid 2.50 € + 1 Cent · CR if he chooses a collaborator with a result *higher* than or equal to his own result. In our experiment the difference (5 € – 2.50 €) represents the additional department's esteem for its manager, if the manager has higher abilities than his collaborator. The payoff structure ensures that rational subjects choose the best candidate (the one with the result of 110) in case of the selection of a more qualified one. If a subject prefers a less qualified collaborator, he should choose the candidate with just one result less than he himself. A tournament between the teams of a four-person-group is organized. If a subject's team result belongs to the two highest of his four-person-group, 7.50 € will additionally be paid to him. Thus the two rewards of 7.50 € represent the winner prizes of the tournament.[7]

2.2 Design of the Managers Treatment (MAT)

One of the five four-person-groups in BT is determined to constitute the reference group for our two main treatments. Each of the 17 "managers" attending MAT is compared to this reference group. Hence, each person in this treatment represents the fifth person in a group consisting of the four persons of the reference group and the considered subject himself. Using this method we obtain a comparatively great number of independent observations because the decisions of the, so to say, "fifth persons" are independent of each other.

MAT is divided into different stages: At stage 1, managers have to state predictions concerning their subsequent performance, at stage 2 they choose a collaborator, and at stage 3 they carry out multiplication tasks. Thus, contrarily to BT, managers perform the tasks at a later point of time.

Stage 1:

At stage 1, each manager is shown a typical multiplication task of stage 3 (namely "6 · 79"). Afterwards, they have to state a self-prediction (SP1) of how many tasks they will be able to solve correctly at stage 3. They know that they shall later work on the task for 10 minutes and that auxiliary means (except for pen and paper) are not permitted. They know that they will

[6] Subjects get 50 Cents for each correctly stated belief.

[7] The term 1 Cent · CR is chosen so that 5 € + 1 Cent · CR is always higher than 2.50 € + 1 Cent · CR. Therefore, choosing a better instead of a worse candidate is only reasonable if expecting to win a winner prize.

earn 4 Cents for each correctly solved multiplication task and that they will get a payoff for their self-prediction that is the higher the more precise their self-prediction is: A manager receives 1.80 € if his self-prediction exactly corresponds to his subsequent result, i.e. to the actual number of correctly solved tasks at stage 3. If the self-prediction deviates from the subsequent result, a manager is paid 1.80 € – (2 Cents · |actual number of correctly solved tasks – predicted number of correctly solved tasks|). Thus, the payoff is the lower, the more the actual performance at stage 3 differs from the predicted value.[8]

After stating their first self-prediction (SP1), managers are informed about the number of tasks the members of the reference group have solved correctly. If the managers thereupon want to change their former self-prediction (SP1), they can do that. In this case, their second self-prediction (SP2) is taken as the new basis for the determination of their payment as described above. If they do not change their self-prediction, payoffs are calculated on the basis of their first self-prediction. By a change of self-predictions, it can be investigated if and how managers renew their prediction of their subsequent performance as reaction to new information.

Stage 2:

At stage 2, managers are informed about the design of BT and the incentives that were given to the reference group's persons. The managers have to estimate which collaborator has been chosen by a certain person of the reference group. 50 Cents are paid for a correct belief. Moreover, they are asked how sure they are of their beliefs. Thereafter, managers select a collaborator from the interval [0, 110]. Each manager and his chosen collaborator compose a team. A manager obtains 5 € + 1 Cent · CR if he chooses a collaborator with a *lower* result than his own at the subsequent stage 3. On the contrary, a manager obtains 2.50 € + 1 Cent · CR if he chooses a collaborator with a result *higher* than or equal to his own result at stage 3. A manager's team result is compared to the four team results of the reference group. If his team belongs to the two best ones, he is additionally paid a bonus of 7.50 €.

[8] Of course, we want to avoid that managers try to solve as many tasks as predicted before and finish working on the task when reaching the predicted score. We pay 4 Cents for each correct answer but payoff is only reduced by 2 Cents for a one unit deviation from the expected score. Thus, there is always an incentive to solve as many tasks as possible. Moreover, even if managers tried to reach their predicted score by counting the calculated tasks (which is not easy because the tasks are not numbered), they would not be sure of their score, because they would not know with certainty if their calculations are correct. The strategy to predict to solve no task correctly and not to work on the task is not profitable either because solving just one task correctly would lead to a higher payoff (neglecting extreme high effort costs).

Stage 3:

At stage 3, managers have to work on the multiplication tasks for 10 minutes. They get 4 Cents for each correct result. Thus we can determine whether a considered manager chooses a better or worse collaborator in relation to the manager's *actual* performance. The payments for stage 2 are calculated with regard to the *actual* (and not the predicted) performance. This regard is necessary because otherwise managers could strategically assess themselves very high in order to win the bonus of 7.50 €.

2.3 Design of the Superiors Treatment (SUT)

In this treatment, each of 17 "superiors" gets associated to a manager of MAT and has to choose a collaborator for him. By a comparison of MAT and SUT, we investigate whether collaborator choices differ when a decision-maker (superior) selects a collaborator for another person (in this case for the manager) and not for himself. As an example, one can imagine a superior in an enterprise who recruits a new employee for his department. The superior allocates the new collaborator to a certain other employee in his department with whom he shall work together. Because the superior chooses the collaborator for his manager, the *manager's actual achievement* is also relevant in this treatment. As will be explained more precisely below, a superior's payoff due to his collaborator choice depends on the actual achievement of his associated manager, the selected collaborator's result, and the relation of the result of his team to those achieved by the teams of the reference group. In this treatment the superior's team consists of the associated manager and the superior's chosen collaborator. As in MAT, the manager's achievement is not known previously to the choice of a collaborator so that assessing his performance is of considerable importance in SUT.

The SUT is divided into two stages: At stage 1, a superior gets associated with a manager of MAT and has to predict this manager's performance.[9] At stage 2, the superior chooses a collaborator for his manager.

Stage 1:

At stage 1, the superior is informed that a manager, who solved multiplication tasks for 10 minutes without using any auxiliary means except for pen and paper, has been associated to him. He is told that a typical task was "6 · 79" and that the manager got 4 Cents for each

[9] Any superior is matched to a different manager, which yields 17 matched pairs. In our statistical analysis, we have to account for the fact that a superior and his associated manager are not independent of each other from a statistical point of view. The most important argument for this is that the manager's performance influences the behavior and payoff of the matched superior. When comparing the superiors with each other, tests for independent observations can be used, because superiors do not interact and have nothing to do with each other.

correct solution. Afterwards, the superior has to assess how many tasks the manager solved correctly. For his assessment he is paid according to the same scheme applied in MAT.[10] After stating his first assessment of the manager's performance (A1), the superior is told that his manager also had to assess how many tasks he would solve correctly. Besides, the superior is told how the manager was paid in dependence of his self-prediction. After being informed about the manager's first self-prediction (SP1), the superior can replace his assessment by a new one (A2). Then only the new, adjusted assessment is used for calculating his payoff (applying the same scheme). Thereafter, the superior is informed about the four subjects of BT who constituted the reference group for the managers in MAT and now also constitute the reference group for the superiors in SUT. Furthermore, he is informed about their performance. After having received this additional information, the superior can again consider his stated assessment and – if he wants – replace it by a new one (A3). Then, again, the payoff is solely calculated on the basis of this new assessment. In a last period of adjusting the assessment, the superior is informed about the revised self-prediction (SP2) the manager stated after he had learned the reference group's results. Again the superior can anew his assessment (A4). With the help of these four assessments, it can exactly be investigated to what degree a certain kind of information influences the superior's assessment. Additionally, the superior is always asked how sure he is of his respective assessment.

Stage 2:

At stage 2, a superior chooses a collaborator for his manager. The superior's payoff in SUT is calculated in the same way as for the managers in MAT. The payoffs consequently depend on the manager's actual achievement and the chosen collaborator's result. A superior's team consists of his associated manager and the superior's chosen collaborator. A superior is paid 7.50 € if his team scores the best or second best result of the five teams (as in MAT, the other four teams are composed of the reference subjects and their chosen collaborators). If the superior chooses a collaborator for his manager, whose result is at least as high as that of the manager, he obtains 2.50 € + 1 Cent · CR. Analogously to MAT, 5 € + 1 Cent · CR are paid if he chooses a collaborator who is worse than his manager. Before the superior chooses a collaborator for his manager, he has to assess which collaborator was chosen by each of the four reference subjects. For each correct belief the superior obtains 50 Cents. Besides, the superior is asked how sure he is of his beliefs.

[10] 1.80 € – {|actual number of the manager's correct solutions – assessed number of correct solutions| · 2 Cents}

At the end of each treatment, subjects have to answer a questionnaire. The questionnaire includes questions concerning risk aversion. These questions are taken from the German Socio-Economic Panel (GSOEP) and deal with the overall risk behavior of subjects. We elicit subjects' risk aversion because their attitude towards risk can affect their collaborator choices: Winning the bonus is uncertain, while the payoff due to a worse collaborator is certain. If a subject is risk averse, he may prefer to voluntarily give up his chance of winning the bonus and employ a less qualified staff. Figure A1 in the appendix gives an overview of the experimental design and the connection between our treatments.

2.4 Hypotheses

In this subsection we present our hypotheses. The first hypothesis deals with the self-predictions of managers in MAT and the second one refers to the performance assessments of superiors in SUT. Moreover, we state hypotheses concerning the collaborator choices of managers and superiors.

If people have to assess their own performance, overconfidence is generally considered to be a robust finding as is claimed by experimental studies in psychology and economics.[11] Moreover, some psychological studies show that subjects are overconfident in self-predictions – in the sense of being too optimistic: People have optimistic predictions about the time for completing a task (Buehler et al. 1994, Buehler and Griffin 2003), and overestimate the likelihood that they would engage in desirable behaviors (Dunning and Epley 2006). This aspect of overconfidence is most interesting in view of our experiment because the managers have to assess their performance before working on the task.[12] Because of the previous results we expect that managers overestimate their subsequent performance also in our real-effort task and state a too high self-prediction SP1 compared to their subsequent performance.

After getting information about the reference group's performance, managers can renew their self-predictions. The question arises whether this information helps to reduce their overconfidence. There are some psychological studies which investigate methods to de-bias subjects' self-assessments, for example by giving them feedback about their performance or train them in self-assessing.[13] These studies show ambiguous results: Some studies find that subjects get de-biased after getting information; others find that subjects ignore information

[11] See e.g. Camerer and Lovallo (1999), Fellner et al. (2004), and Weinstein (1980).
[12] In our study we use the term "overconfidence" in the sense that subjects overestimate their subsequent performance.
[13] See e.g. Adams and Adams (1961), Lichtenstein and Fischhoff (1980), Pulford and Colman (1997), and Sharp et al. (1988).

and hence do not improve their self-assessments.[14] In these studies subjects get valuable information, i.e. information that can indeed help them with their self-assessments. In our experiment, it is not clear whether the information about the performance of the reference group can help managers to improve their self-predictions because this information is not specific to their own ability. As even some of the other studies find that subjects are not de-biased although obtaining valuable information, we expect that managers are still overconfident when stating their second self-prediction in our experiment. Therefore we state the following behavioral hypothesis concerning SP1 and SP2:

Hypothesis 1: Managers are overconfident when predicting their subsequent performance.

Superiors in SUT have to assess the managers' performances and can renew their assessments three times after getting information about SP1, the performance of the reference group, and SP2. Also in this treatment, it is interesting whether superiors regard the information about the reference group's performance as helpful when predicting the performance of their managers. Next to this question, we investigate how superiors evaluate the self-predictions of their managers. Do superiors believe that the self-predictions of their managers are accurate? Superiors may anticipate that the self-predictions of the managers are too high because overconfidence is a common bias. This idea is underlined by the results of Cesarini et al. (2006), who experimentally find that subjects anticipate the overconfidence of their peers. In our experiment we therefore expect that superiors think that managers are overconfident and thus do not take their self-predictions for sure and state lower performance predictions.

Hypothesis 2: Superiors believe that managers are overconfident.

Concerning the collaborator choices in our treatments, we have to differentiate between those subjects who state a SP2 respectively an A4 that is higher than the performance of the second best reference person and those who state one that is lower.

If a manager thinks that he can correctly solve more tasks than the reference group's person with the second highest result, he will probably choose the best collaborator (with a result of 110). Then, in the eyes of the manager, his team will belong to the two best teams and he will get the bonus independent of the factual collaborator choices of the reference persons. This

[14] While for example Pulford and Colman (1997) and Sharp et al. (1988) find that feedback does mostly not influence overconfidence, Adams and Adams (1961) and Lichtenstein and Fischhoff (1977) find that feedback reduces overconfidence or can improve calibration, respectively.

choice assures winning the bonus if the manager's subsequent performance indeed turns out to be higher than the result of the second best reference person. If the manager chooses a worse collaborator than he himself, he may obtain the bonus, too. However, this is combined with uncertainty because the manager does not know which collaborator the reference persons selected. Therefore we think that most managers, whose self-assessments are higher than the result of the second best reference person, choose the best collaborator. This is analogously true for the choices of the superiors if their A4 is higher than the result of the second best reference person. Therefore, we state the following hypotheses:

Hypothesis 3a: (Most) Managers whose self-predictions are above the performance of the second best reference person choose the best collaborator.

Hypothesis 3b: (Most) Superiors whose performance assessments are above the performance of the second best reference person choose the best collaborator.

Let us now consider the case in which the self-prediction of a manager or the assessment of a superior is below the result of the second best reference person. Here, the choice of a collaborator with a lower result than SP2 respectively A4 yields a sure payment of 5 € + 1 Cent · CR (instead of 2.50 € + 1 Cent · CR) independent of the collaborator choices of the reference group's persons. A subject with a low predicted performance might not think to have a chance to win the bonus. Therefore we expect that most of these selectors choose a collaborator with a lower result than their own performance prediction. Hence, we state the following hypotheses:

Hypothesis 4a: (Most) Managers whose self-predictions are below the performance of the second best reference person choose a worse collaborator.

Hypothesis 4b: (Most) Superiors whose performance assessments are below the performance of the second best reference person choose a worse collaborator.

Of course, when choosing a worse collaborator, it is optimal to choose one whose result is only one point lower than the actual result of the manager. However, subjects cannot be sure that their performance predictions are correct. They may worry that the performance turns out to be lower than SP2 respectively A4. Therefore we expect that they want to make sure to

really choose a worse collaborator by choosing one whose result is more than one point lower than their SP2 respectively A4.

What do the so far stated hypotheses suggest concerning a comparison between the collaborator choices of managers and superiors? If managers are indeed overconfident and superiors anticipate this, the average assessment of superiors will lie below that of the managers. If most of the superiors' assessments are below the second highest result of the reference group and most of the managers' predictions are above, we expect that fewer superiors than managers choose the best collaborator, given that most subjects believe that the two best reference persons have chosen the best collaborator. Therefore, we state our last hypothesis:

Hypothesis 5: More managers than superiors choose the best collaborator.

3. Results of BT

We conduct the BT to determine a typical reference group to which subjects in the main treatments are compared. In the group which we select as reference group the subjects correctly solved 43, 52, and 64 respectively 81 multiplication tasks (mean 60). This group is the most "typical" one, because the average number of correctly solved tasks of the chosen reference group corresponds to the average number of correctly solved tasks over all subjects in BT. Moreover, the result of the reference group's subject with the (second) highest result is very close to the average result of the subjects belonging to the (second) highest quartile. This is analogously true for the performances of the two worst subjects in the reference group. When investigating the collaborator choices of this group, we find that the best and the worst subject choose the best collaborator, while the other two subjects choose a collaborator who is one point worse than they themselves.

4. Results of MAT

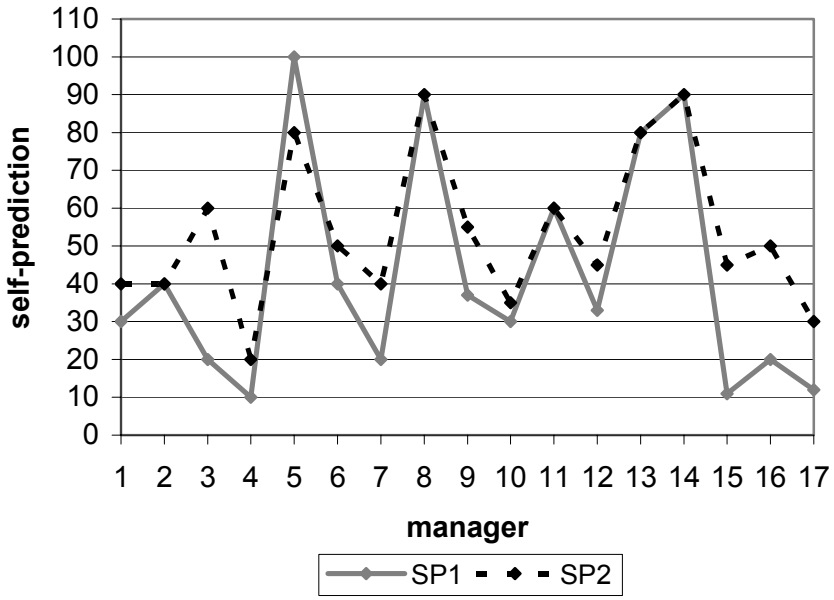
In this section we start with the investigation of the managers' performance predictions. We compare the successive performance predictions, analyze the reactions to new information and consider the accuracy of the predictions. Then we examine the beliefs about the reference persons' choices. Both, performance predictions as well as beliefs are decisive for the collaborator choices of the respective selectors. Afterwards, we investigate the selection

decisions in detail and study if they are rational from an ex ante point of view as well as from an ex post point of view.

Self-predictions

First, each manager is only informed about the kind of a typical multiplication task (namely "6 · 79") and states his first self-prediction (SP1). After being informed about the number of correctly solved tasks of the reference persons, managers state their second self-predictions (SP2). Figure 1 presents the self-predictions SP1 and SP2 of each manager. The average reported SP1 is 43, while the average reported SP2 is 54.[15] A Wilcoxon-Signed-Rank test for dependent observations shows that self-predictions 1 and 2 are significantly different from each other (p = 0.012, two-sided), so managers renew their self-predictions in the light of new information. We observe that 70.59% of all managers adjust their first self-predictions into the direction of the average number of correctly solved tasks of the reference group

Figure 1: Self-predictions SP1 and SP2 in MAT



(which is 60). When comparing the absolute value of the difference 60 – SP2 with the absolute value of the difference 60 – SP1 for all managers, a significant difference can be found (Wilcoxon-Signed-Rank test, p = 0.00, two-sided). Thus managers seem to perceive the information about the reference group as indication for their own performance in the multiplication task and adjust their SP1 into the direction of 60. This is particularly pronounced by those managers whose SP1 is lower than the average number of correctly

[15] The medians are 33 and 50; the modes are 20 and 40.

solved tasks of the reference group (70.59%). Interestingly, 91.67% of them increase their assessments after receiving information about the reference group. Out of those managers who have stated a SP1 higher than the average number of correctly solved tasks of the reference group (23.53%), only 25% decrease their prediction, while the others do not change their predictions.

Apart from the mean, managers may consider the distribution of results of the reference group. Note that 70.59% of the managers state a SP1 that is lower than the worst result of the reference group. In contrast to this, only 17.65% state a SP1 which is higher than the best result of the reference group. After being informed about the results of the reference group, these managers know that their self-predictions are rather extreme. It is plausible that these managers adapt their SP2 accordingly, unless they are completely confident of their self-prediction. Indeed, out of 88.24% having an extreme SP1, 80% adjust their self-prediction. 46.67% of the managers with an extreme SP1 state an adjusted SP2 lying in the interval [43, 81], which is the range of the reference persons' results. The other 33.33% change their self-predictions in the direction of the average number of correctly solved tasks of the reference group but still have rather "extreme" SP2s.

To test Hypothesis 1 and to investigate how accurately managers predict their subsequent performance in the task, we compare their self-predictions SP1 and SP2 with their actual number of correctly solved tasks. A Wilcoxon-Signed-Rank test shows that we have to reject Hypothesis 1 since SP1 as well as SP2 are not significantly higher than the actual performance. Since we nevertheless want to analyze if managers' self-predictions deviate from their subsequent performance, we additionally applied a two-sided Wilcoxon-Signed-Rank test. While the test indicates that the first self-predictions SP1 are significantly different from the actual performance (Wilcoxon-Signed-Rank test, $p = 0.004$, two-sided), the second self-predictions SP2 are not. Contrarily to Hypothesis 1, the first self-predictions are lower than the subsequent performance: 76.47% of managers underestimate their performance without the information about the reference group. Thus, managers are biased (i.e. make systematical mistakes) in predicting their subsequent performance, but in the opposite direction as hypothesized. This finding can be underlined by the strand of psychological literature which deals with the so-called "hard easy effect" when assessing one's performance. As many authors find[16] absolute overconfidence is most extreme at tasks of great difficulty[17]. However, when tasks get easier, overconfidence is reduced. Indeed, subjects responding to

[16] See for example Clarke (1960), Lichtenstein and Fischhoff (1977), and Gigerenzer et al. (1991).
[17] Almost impossible tasks, which have been investigated, are for example the distinction between European and American handwritings and between Asian and European children's drawings.

very easy tasks are often underconfident. According to this literature, our findings can be explained by the fact that subjects are familiar with multiplication tasks and do not have severe problems to work on them. Thus, most managers reveal underconfidence when predicting their performance in this (easy) task.

Some information about the reference group's performance seems to really help subjects to accurately predict their performance. As Table 1 indicates, the average difference between the self-predictions and the actual results becomes smaller after the managers receive the information about the reference group.

Table 1: Average accurateness of self-predictions

		SP1	SP2
Difference between self-prediction and actual result	mean	– 14.94	– 3.94
	standard deviation	17.53	10.05
Absolute value of the difference between self-prediction and actual result	mean	19.29	8.53
	standard deviation	12.20	6.34

Beliefs

If we want to interpret the choice of a collaborator, we have to pay attention to the managers' beliefs about the choices of the reference persons.[18] These beliefs are shown in Table 2.

Table 2: Beliefs in MAT

	score – 1	110	other choice
belief1	82.35%	0	17.65%
belief2	70.59%	17.65%	11.76%
belief3	11.76%	76.47%	11.76%
belief4	17.65%	70.59%	11.76%

For example, row 1 pictures the managers' beliefs about the choice of the reference person with the lowest result in the multiplication task. "Score – 1" means that a collaborator, who has a one-point-lower score than the particular subject of the reference group, has been chosen. "110" specifies the choice of the best collaborator, and "other choice" indicates the

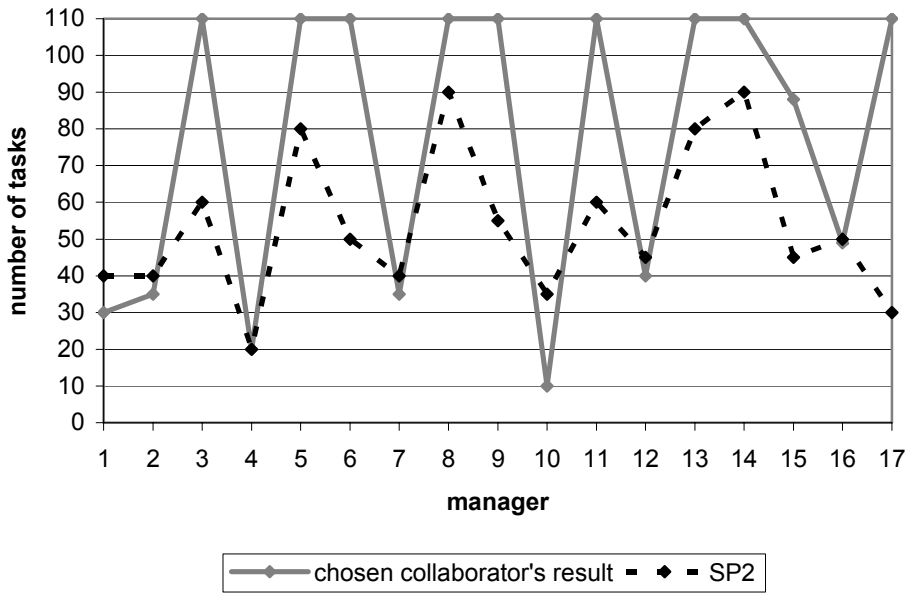
[18] Beliefs are denoted belief1 till belief4, with belief1 denoting the expected choice of the worst subject of the reference group and belief4 denoting the expected choice of the best subject of the reference group.

choice of another collaborator. First of all, note that interestingly at most 17.65% of the managers do not have rational beliefs and/or think that a particular person of the reference group has not rationally chosen a collaborator. While most of the managers expect that the two persons with the lowest results have chosen a worse collaborator, the two persons with the highest results are expected to have chosen the best collaborator.

Choice of collaborator

The managers' collaborator choices are shown in Figure 2. The best collaborator is chosen by 52.94% of the managers.[19] Interestingly, all managers who have a SP2 higher than the result of the second best reference person choose the best collaborator: A one-sided Binomial test reveals that the probability of selecting the best collaborator is significantly higher than 0.5 ($p = 0.0625$). Therefore Hypothesis 3a cannot be rejected. As these managers believe that the two best reference persons have chosen the best collaborator they want to win a bonus by choosing the best collaborator, too.

Figure 2: Choice of collaborator and SP2 in MAT



There is one manager in MAT who chooses a collaborator with a higher result than his own self-prediction, but not the best collaborator. This might be explained by the fact that this manager fails to recognize that the choice of the best collaborator is most profitable when choosing a better collaborator. Another manager chooses a collaborator who is equal to his

[19] One may think that the collaborator choice is influenced by a manager's risk attitude. However, no significant correlation is found between the applied risk aversion measure and the collaborator choice.

SP2. Thus his choice neither corresponds to a worse nor to a better collaborator. Because his SP2 is very low (just 20) we think that he actually hopes to solve more tasks correctly and therefore has the choice of a worse collaborator in mind.

A worse collaborator is chosen by 35.29% of the managers. Almost all of these managers choose a collaborator who is more than one point worse than their SP2. These managers want to assure that they really choose a worse collaborator compared to their subsequent performance. On average they choose a collaborator with a 10-point-lower result so that the average deduction for reasons of cautiousness is 9 points. Again we use a one-sided Binomial test to investigate whether most of the managers who have a SP2 lower than the result of the second best reference person choose a worse collaborator (Hypothesis 4a). However, in this case we cannot reject the null hypothesis that the probability of choosing a worse collaborator is lower than or equal to 0.5. Indeed, only 46.15% of the managers who have a SP2 lower than 64 choose a worse collaborator.

It seems plausible that the proportion of managers who choose a better collaborator is higher for managers with SP2 > 64 than for managers with SP2 ≤ 64.[20] The results of a one-sided Fisher exact test show that the proportion of managers with SP2 > 64, choosing a better collaborator, is not significantly higher than the proportion of managers with an SP2 ≤ 64, choosing a better collaborator ($p = 0.115$, one-sided). However, there is a clear tendency in this direction. While all managers with SP2 > 64 choose a better collaborator, only half of the managers with SP2 ≤ 64 choose a better collaborator.

The so far conducted analysis of collaborator choices does not take note of the managers' beliefs. In order to incorporate their beliefs and to analyze the managers' choices in terms of rational behavior, we develop two rationality concepts: one from an ex ante and one from an ex post point of view. If a manager chooses a collaborator in line with his beliefs (belief1 to belief4) and SP2, we call his choice *ex ante rational*: When a manager states certain beliefs and a certain SP2, we can calculate whether he should choose a better or worse collaborator to maximize his payoff from an ex ante point of view. We call a manager *ex post rational*, if he chooses a collaborator in line with his *actual* performance and the *actual* collaborator choices of the reference persons. Clearly, this is a very theoretical approach because managers do neither know their actual performances nor the choices of the reference persons. But with this concept we can investigate whether the lack of information leads to a severe distortion of collaborator choices.

[20] We exclude the manager whose collaborator choice equals his SP2 because he neither chooses a better nor a worse collaborator.

With both rationality concepts we investigate whether a manager chooses a better or a worse collaborator than SP2 and whether he should – from an ex ante or an ex post point of view – select a better or a worse one. Of course, we could also investigate the exact-point-choices and state whether these are ex ante (ex post) rational or not. For example, consider a manager who thinks that he will solve 50 calculations correctly and that the two best reference persons have chosen the best collaborator. Then he should – from an ex ante point of view – choose a 49-point-collaborator. But this choice cannot concretely be expected, because the manager is normally not totally sure that he will indeed correctly solve 50 calculations. This can be one reason for the choice of a collaborator who is a bit worse than 49. So, we think it would be misleading to describe this person as to be not ex ante rational. Thus, we use concepts of rationality which just control for the choice of a better or a worse collaborator and not for an exact-point-choice.

Examining the data reveals that only 37.5% of the managers behave ex ante rationally.[21] 33.33% of those choose a better collaborator. More than half of the ex ante irrational managers choose a better collaborator, although a worse collaborator would be rational, given their SP2 and beliefs. If managers state a relative low SP2 and irrationally choose a better collaborator, this may be explained by "wishful thinking". At the moment when managers choose a better collaborator, they may think that their stated self-prediction was too low. While hoping for a better performance, they may be tempted to try to win the bonus by choosing a better collaborator. If their subsequent performance turns out to be higher than 64, they will indeed obtain a bonus. This supposition can be corroborated by the observation that 66.67% of the managers, who irrationally choose a better collaborator and have a SP2 lower than 64, have self-predictions in the interval [50, 60]. These self-predictions are not far away from 64. Since these managers believe that the reference group's subjects with the two highest results choose the best collaborator, they try to win the bonus by choosing a better collaborator, too. Another explanation for irrationally choosing a better collaborator concerns managers with a very high SP2: Managers with an extreme high SP2 who should from an ex ante point of view choose a worse collaborator to maximize their payoff (given their statements they will win the bonus also by choosing a worse collaborator) are of course tempted to choose the best one to be sure to get the bonus. They may be afraid that their performance will turn out to be lower than predicted.

If we assess the collaborator choices from an ex post point of view, we observe that 41.18% of the managers behave ex post rationally. 60% of the managers, who do not behave

[21] We exclude the manager who chooses a collaborator equal to his own SP2 because the collaborator is neither better nor worse than SP2.

ex post rationally, choose a better collaborator. They would have done better by choosing a worse one.

In summary, the results of MAT show that most managers underestimate their subsequent performance without any additional information which seems to be a consequence of the so-called "hard easy effect". Information about the reference group is very valuable to managers: Most of them adjust their self-predictions in "the right direction". Thus we do not find a significant difference between the second self-predictions and the actual performance anymore. As hypothesized, most managers whose SP2 is above the result of the second best reference person choose the best collaborator. In contrast to Hypothesis 4a, only 46.15% of those managers whose SP2 is lower than the result of the second best reference person choose a worse collaborator. Furthermore, only 37.5% of managers choose their collaborator ex ante rationally. Interestingly, many of the ex ante irrational managers choose a better collaborator, although there seems to be no chance to win the bonus.

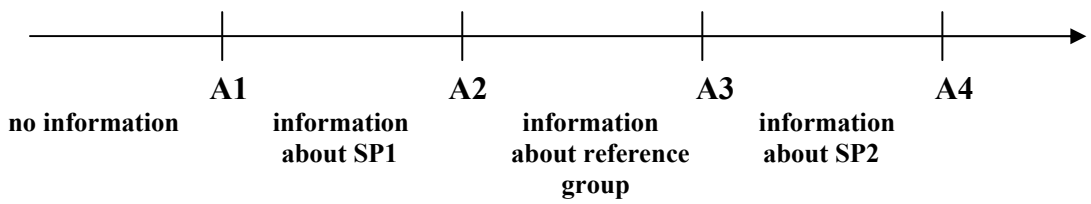
5. Results of SUT

In this section we first investigate the superiors' assessments of the managers' performances and the influence of information on these assessments. Afterwards, we analyze the superiors' beliefs and collaborator choices. Moreover, we examine whether the superiors' choices are ex ante rational, given their assessments and beliefs, and whether they are rational from an ex post point of view.

Superiors' Assessments

A superior has the opportunity to revise his assessments three times after getting more and more information (see Figure 3). Starting with an average assessment of 46, the second

Figure 3: Information and assessments in SUT



average assessment is somewhat lower (41). After receiving information about the performance of the reference group's subjects, the superiors' average assessment becomes higher (50), with the highest after the information about SP2 (51).[22]

How accurately do superiors assess their managers' performance? To get an insight into this issue, Table 3 shows the average differences between superiors' assessments A1 to A4 and their managers' actual results. We make two different computations: One takes account of the signs; the other regards the absolute values of the respective differences. Both cases show that the last assessment, which incorporates all information, corresponds best to the actual

Table 3: Average accurateness of assessments

		A1	A2	A3	A4
Difference between assessment and actual result	mean	– 11.41	– 16.59	– 7.29	– 6.59
	standard deviation	30.19	17.41	18.64	17.33
Absolute value of the difference between assessment and actual result	mean	26.94	19.53	16.12	14.47
	standard deviation	16.68	13.80	11.31	11.14

result. Moreover, the first row shows that superiors underestimate managers' performances on average. However, a two-sided Wilcoxon-Signed-Rank test reveals that there is only a significant difference between A2 and the actual performance ($p = 0.002$, two-sided). A2 is significantly lower than the actual performance. Keeping the payoff structure of our experiment in mind, the four assessments should only differ if the superiors believe to earn more if they renew their assessment after obtaining more information. To investigate the influence of information, we first examine whether the superiors' assessments differ from each other. We find significant differences between A2 and A3 and between A2 and A4, with A2 being lower in both comparisons (Wilcoxon-Signed-Rank test, $p = 0.004$ und $p = 0.001$, both two-sided).

Reactions to information about SP1

To investigate in more detail whether superiors adjust their assessments to the new information about SP1 we compare the absolute difference between SP1 and the former assessment A1 with the absolute difference between SP1 and the new assessment A2. If superiors take SP1 for sure, they should renew their assessment in the corresponding

[22] Medians are always lower: 38, 35, 48, and 50, respectively. The modes are 57, 35, 45, and 50.

direction. A comparison between the absolute values of the differences SP1 – A1 and SP1 – A2 reveals a significant difference (Wilcoxon-Signed-Rank test, p = 0.00, two-sided) with SP1 – A1 being higher. Most superiors take the information of SP1 into consideration and shift their second assessment in the direction of SP1.

It is an interesting aspect to analyze how much a superior trusts the self-prediction of his manager. On the one hand, SP1 is doubtful information because a superior does not know whether his manager is good in predicting his own performance. On the other hand, it is – at this stage of the experiment – the only information that is directly connected with the manager whom he has to assess. Examining the superiors' reactions to SP1 more closely shows that only 17.65% of them state an A2 equal to SP1. Moreover, 11.76% do not change their assessment and therefore ignore SP1. The others modify their assessment in the direction of SP1. Interestingly, none of these superiors changes his assessment in such a way that his new assessment A2 is "at the other side" of SP1 viewed from A1: If A1 is below SP1 (50% of these superiors), A2 is below SP1, too, but less distant. If A1 is above SP1, A2 is also above SP1 but less distant. As our data show, most superiors take the information SP1 for serious but do not completely take note of it and keep their first assessment in mind.[23] These superiors do not think that their managers are completely right with their self-prediction.

To test Hypothesis 2 regarding SP1, a one-sided Wilcoxon-Signed-Rank test shows that SP1 is not significantly higher than A2. Therefore, Hypothesis 2 can be rejected regarding SP1. Furthermore, also SP1 and A1 do not significantly differ. Hence, already the first predictions are quite similar.

Reactions to information about the reference group

To investigate superiors' reactions to the next information, we take the average performance of the reference group into consideration. Thus we take the average number of correctly solved tasks of the reference group as comparison value in our tests. One may argue that superiors consider the distribution of performances next to the average performance of the subjects (see MAT). This can be particularly relevant if a superior's A2 is higher (lower) than the result of the best (worst) subject of the reference group. But also in these cases, the superiors may adapt A2 into the direction of the mean of the reference group. Therefore, considering the average performance of the reference group in our tests is not misleading.

Comparing the absolute values of the differences 60 – A2 and 60 – A3 shows that they are significantly different (Wilcoxon-Signed-Rank test, p = 0.004, two-sided), with 60 – A2 being

[23] In psychology, there is much evidence that once people have formed an opinion, they cling to it too tightly and far too long (e.g. Lord et al. 1979). This phenomenon is commonly called "belief perseverance".

higher. A lot of superiors seem to revise their assessment in the light of new information. Investigating their reactions in detail, we observe that 47.06% of the superiors move A3 in the direction of 60 but remain below it, as are all their A2s. These superiors regard the average result of the reference group but do not forget their former assessments. Moreover, one superior states an A3 of 60. This superior seems to assume that the reference group is a really typical group. Contrarily, 47.06% of the superiors do not change their assessments and retain A2. These superiors either feel their assessment to be supported by the information about the reference persons or do not think that the performance of the reference group is helpful information for predicting the manager's performance.

Reactions to information about SP2

The last information a superior gets is the SP2 of his manager. Again, we observe a significant difference between the absolute values of the differences SP2 – A3 and SP2 – A4 (Wilcoxon-Signed-Rank test, $p = 0.039$, two-sided), with SP2 – A3 being higher. Thus the information about SP2 is very often taken for serious; 41.18% of the superiors take this information into consideration and adjust their next assessment in the direction of the new information, or even state an A4 which equals SP2 (23.53% of these subjects). However, 52.94% of the superiors do not change their A3. Indeed 17.65% state an A3 that is already equal to SP2. The others who do not change their A3 do not make use of SP2. To test Hypothesis 2, we conduct a one-sided Wilcoxon-Signed-Rank test that demonstrates that SP2 is not significantly higher than A4. Therefore, Hypothesis 2 can be rejected regarding SP2. Moreover, A3 and SP2 do not significantly differ.

An interesting finding can be detected if we divide the SP2s into two categories: $SP2 < 60$ and $SP2 \geq 60$. Superiors who get to know a SP2 lower than the average performance of the reference group may assume that their managers' self-consciousness is too low. A two-sided Wilcoxon-Signed-Rank test reveals that A4 is significantly different from SP2 for the first category ($p = 0.063$) but not significantly different for the second. If the SP2 of a manager is lower than 60, most of the matched superiors state an A4 that is higher than SP2. This hints at the fact that superiors, whose manager predicts a performance below average, expect that the manager is underconfident.

Beliefs

As can be seen from Table 4, most superiors believe that the worst and second worst reference subjects choose a collaborator with a one-point lower result. Moreover, most of

them believe that the two best reference subjects choose the best collaborator. Nonetheless, there are also many superiors who believe that the best two reference subjects choose a one-point-worse collaborator.

Table 4: Beliefs in SUT

	score – 1	110	other choice
belief1	82.35%	5.88%	11.76%
belief2	70.59%	17.65%	11.76%
belief3	29.41%	52.94%	17.65%
belief4	41.18%	47.06%	11.76%

Choice of collaborator

35.29% of the superiors choose the best possible collaborator, while 52.94% choose a worse collaborator compared to their stated A4.[24] There are two other superiors, who choose a better collaborator, but not the best one. This selection is not rational because the choice of the best collaborator would increase their payoffs. There are two different explanations for their behaviors. First, it may be that these superiors simply do not understand that the choice of the 110-point-collaborator is most profitable when choosing a better collaborator. Second, it may be that these superiors increase their assessments after stating A4 or forget what they exactly wrote down as A4 and have a higher value in mind. In these cases the collaborator choice of the two "irrational" superiors can be interpreted as if these superiors purposed the selection of a worse collaborator. This seems to be the case at least for one of these two superiors: This superior states an A4 of 50, while the chosen collaborator's result is 55. Moreover, this superior writes in the questionnaire that he has chosen a worse collaborator for his manager. This indicates that this superior indeed intended to choose a worse collaborator for his manager. Nevertheless, these superiors have chosen a better collaborator on their decision sheet, so we treat them as if they wanted to choose a better collaborator.

Examining the data in detail shows that 66.67% of the superiors who choose a worse collaborator select one who is more than one point worse than their stated A4. On average, the difference between the chosen collaborator's result and A4 is 18 for these superiors. Similar to some managers in MAT these superiors include a deduction of points for reasons of cautiousness. They really want to make sure to choose a collaborator who is worse than the

[24] One may think that the collaborator choice is influenced by a superior's risk attitude. However, no significant correlation is found between the applied risk-aversion measure and the collaborator choice.

actual performance of the associated manager. One of these superiors even chooses a 0-point-collaborator to ensure that his manager is better.

To examine whether we can find support for Hypotheses 3b and 4b, we separately analyze the choices of superiors whose performance assessments are above or below the performance of the second best reference person. Hypothesis 3b cannot be rejected: A Binomial test reveals that we can reject the null hypothesis that the probability for choosing the best collaborator is lower than or equal to 0.5 ($p = 0.0625$). All superiors who have an A4 that is higher than the performance of the second best reference person choose the best collaborator to assure to get the bonus if the manager's subsequent performance indeed turns out to be higher than the result of the second best reference person. Similar as in MAT we have to reject Hypothesis 4b: Although most of the superiors (69.23%), whose performance assessments are below the performance of the second best reference person, choose a worse collaborator, a one-sided Binomial test indicates that we cannot reject the null hypothesis that the probability of choosing a worse collaborator is lower than or equal to 0.5.

It seems plausible that the proportion of superiors, who choose a better collaborator, is higher for superiors with $A4 > 64$ than for superiors with $A4 \leq 64$. Indeed, a one-sided Fisher exact test shows that the proportion of superiors with an $A4 > 64$, who choose a better collaborator, is significantly higher than the proportion of superiors with an $A4 \leq 64$, who choose a better collaborator ($p = 0.029$, one-sided). This shows that the performance assessments have a crucial influence on the choices of superiors.

However, note that this analysis neglects the beliefs of the superiors. Thus in a next step we incorporate beliefs and investigate whether superiors' collaborator choices are *ex ante rational* and therefore consistent with their stated beliefs (belief1 to belief4) and A4.[25] 58.82% of the superiors behave ex ante rationally. Only 28.57% of those superiors, who do not behave ex ante rationally, choose a better collaborator. Thus, in this treatment too many superiors choose a worse collaborator from an ex ante point of view.

Another interesting aspect is the evaluation of the superiors' choices with regard to the managers' actual performances and the actual choices of the reference persons. As observed in BT, the best and the worst subject of the reference group choose the 110-point-collaborator and the other two subjects a collaborator with a one-point-lower result. If these choices and the actual performance of the associated manager in MAT were known by the superior, he could precisely choose the collaborator who maximizes his payoff. Our data show that

[25] Again we disregard the exact-point-choices and only take into account whether superiors choose a better or worse collaborator.

52.94% of the superiors behave ex post rationally. Of those who do not behave ex post rationally 50% choose a better collaborator although they should choose a worse one.

6. Comparison of MAT and SUT

In this section we compare the accurateness of performance predictions, the beliefs, and the collaborator choices across MAT and SUT.

Performance predictions

First, we regard the managers' and superiors' performance predictions. We cannot find any significant differences between SP1 and A1 as well as between SP2 and A4. We also do not observe a significant difference if we compare the accurateness of performance predictions across treatments, neither when taking note of the signs nor when comparing the absolute values.[26] Thus, managers do not predict their subsequent performance systematically more accurately than superiors, neither before getting any information nor after obtaining information.

Beliefs

Next, we compare the beliefs about the reference persons' collaborator choices: Concerning belief1 and belief2 most managers and most superiors think that the two worst reference subjects choose a worse collaborator. But regarding belief4 there is a noticeable difference: Much more superiors than managers believe that the best reference person chooses a worse collaborator. This tendency can also be seen for belief3. Nevertheless, a Wilcoxon-Signed-Rank test shows that we cannot find a significant difference if we compare the beliefs of managers and superiors about a certain reference person with each other.

Collaborator Choice

We now compare the collaborator choices across MAT and SUT. As has already been mentioned, 52.94% of the managers and 35.29% of the superiors choose the best collaborator. Contrarily, 35.29% of the managers and 52.94% of the superiors choose a collaborator with a

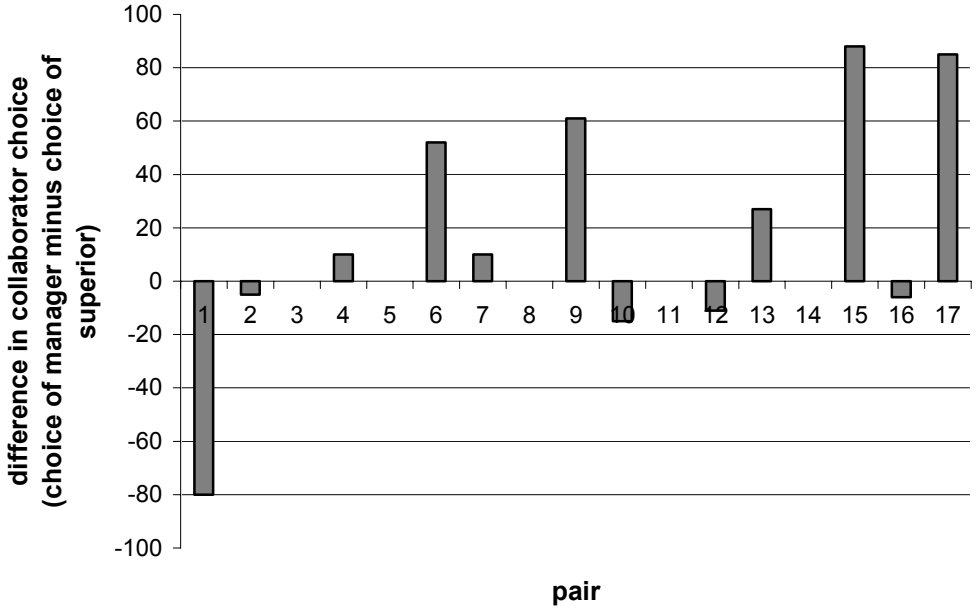
[26] While a sign test reveals no significant difference between |SP2 – actual performance| and |A4 – actual performance|, a Wilcoxon-Signed-Rank test does. However, one assumption of the Wilcoxon-Signed-Rank test is that the difference of the compared values has to be symmetrically distributed. The particular difference of the values of |SP2 – actual performance| and |A4 – actual performance| is rather asymmetrically distributed so that the results of the sign test are more meaningful and therefore presented here. For the other comparisons both tests do not show a significant difference.

worse result than SP2 respectively A4. The average result of the chosen collaborators in MAT is 76 and 64 in SUT. This shows that the collaborators in MAT have on average higher results. If we only compare the chosen collaborators, who are worse than the subject's SP2 respectively A4, their average results are very similar (33 in MAT, 31 in SUT). Interestingly, the average deduction due to reasons of cautiousness is much lower for managers (9) than for superiors (17). This may reflect that superiors, who want to be sure to choose a worse collaborator, are less sure of their managers' performance than managers are of their own performance.

If we compare the collaborator choices of managers and superiors, we also have to account for the fact that their self-predictions and assessments differ, even if not significantly. Our data show that 23.53% of the managers and also 23.53% of the superiors state a SP2 respectively A4 which is higher than 64. All of these managers and superiors choose the best collaborator. Thus, these subjects want to assure to win the bonus. 76.47 % of all managers and superiors state a SP2 respectively A4 that is lower than 64. In these cases, more managers than superiors choose a better collaborator than SP2 respectively A4 (46.15% of managers and 30.77% of superiors).

When considering the matched pairs of managers and superiors, there are noticeable differences in collaborator choices. Figure 4 illustrates the difference in collaborator choices

Figure 4: Differences in collaborator choices between MAT and SUT for matched pairs



(choice of a manager minus choice of the matched superior) for each matched manager-superior pair. As can be seen, only in 29.41% of all pairs, managers and superiors exactly

choose the same collaborator (which is the best collaborator in all these cases). In 41.18% of the matched pairs a manager chooses a better collaborator than the matched superior. In 57.14% of these cases the manager chooses a better collaborator than SP2, while the superior chooses a worse collaborator than A4. Contrarily, in 29.41% of the matched pairs the superiors choose a better collaborator than the managers. In 60% of these cases, both choose a collaborator worse than SP2 respectively A4 but the superiors' collaborators have higher results than those chosen by the managers.

To test Hypothesis 5 we conduct a one-sided McNemar test. It reveals that we have to reject Hypothesis 5 because there are not significantly more managers than superiors who choose the best collaborator. Furthermore, there is no significant difference when testing the collaborator choices across treatments with a Wilcoxon-Signed-Rank test.

Next to these aspects, we are interested in the question whether there are differences between managers' and superiors' behaviors concerning ex ante and ex post rationality. While only 37.5% of the managers behave ex ante rationally, 58.82% of the superiors are ex ante rational. Investigating this phenomenon in more detail, 80% of those managers who do not behave ex ante rationally choose a better collaborator. They should however choose a collaborator whose result is worse than their SP2 from an ex ante point of view. Having a look at the superiors, 28.57% of those who do not behave ex ante rationally choose a better collaborator although they state a very low A4. Therefore, this kind of mistake seems to happen more often in MAT than in SUT. A possible reason for this may be that more managers than superiors have a kind of "wishful thinking". Perhaps more managers hope to win the bonus – in contrast to their self-predictions and beliefs. They may hope to be better in the multiplication tasks than predicted before, or hope to have the luck that all subjects in the reference group choose a worse collaborator. Superiors do not seem to be so hopeful. They seem to condition their behavior better on the relevant statements.

If we compare the proportions of managers and superiors whose collaborator choice is ex post rational, we find that more superiors (52.94%) than managers (41.18%) are ex post rational.

7. Summary and Conclusion

Our results reveal that managers' self-predictions (SP1) are biased if their information only consists of an example of a typical multiplication task. In contrast to our Hypothesis 1 managers significantly underestimate their subsequent performance. Even about 70.59% of

them assess themselves worse than the worst reference person (whose result is not known to them at that time). This result can be explained by the "easiness" of our task and is in line with the so-called "hard easy effect". Interestingly, the accuracy of the managers' final predictions is remarkably improved, after managers are informed about the reference group's results. The very most of them adjust their predictions in the direction of the reference group's average result so that no significant difference to the subsequent performance remains.

Similarly, a very good information processing and adjustment of assessments can be observed for the superiors. Superiors significantly adjust all their assessments in the direction of the respective new information. However, they do not completely take note of the new information and keep their former assessment in mind ("belief perseverance"). They are especially sensitive to the information about the reference persons' performances, apparently regarding those as really factual data. As expected, A4 is the most accurate assessment of the managers' performances. It is not significantly different from SP2 and from the managers' factual performance. However, considering only superiors who are matched with a manager who states a SP2 that is lower than the average result of the reference group reveals that superiors' last assessments are significantly higher. Thus these superiors suppose that their managers are underconfident.

As hypothesized, most managers and most superiors whose SP2 respectively A4 is above the result of the second best reference person choose the best collaborator. Furthermore, only 46.15% of managers and 69.23% of superiors, whose predictions are lower than the result of the second best reference person, choose a worse collaborator.

Looking at the matched pairs shows that in 41.18% of the pairs the managers choose a better collaborator than the superiors but the differences are not statistically significant. Thus the collaborator choice does not seem to depend on the "authority" of the decision maker, i.e. whether he chooses a collaborator for himself or a collaborator selected by a third person is allocated to him. This result is a consequence of the excellent information processing and the excellent assessments of both, managers and superiors. Nevertheless, to a considerable degree we observe irrationality in choosing collaborators. Astonishingly, only 37.5% of the managers choose their collaborator ex ante rationally. In many cases, they prefer the best possible one although they have almost no chance to win a bonus according to their self-predictions and beliefs. Such behavior is most likely caused by a kind of "wishful thinking": They hope to win the bonus in spite of a rather bad precondition. Likewise the superiors do not choose collaborators for their managers in a notably rational way. 41.18% of the choices are irrational

in the light of their information and the given incentives. However, "wishful thinking" is less frequently observed for the superiors.

In our experiment, the collaborator choices depend on performance (self-)predictions, beliefs about the reference subjects' choices, and the given incentives. It would be interesting to examine in a further treatment, whether and how a change in incentives affects the collaborator choices. If the bonus is increased, we expect that more subjects try to win the bonus. Perhaps "wishful thinking" is more widespread in such an environment. Then even more subjects may choose the best collaborator, although their performance prediction and beliefs tell them to choose a worse one. This may also be the case if the bonus remains at 7.50 € but the amount of money, a subject gets for sure if choosing a worse collaborator, is decreased. More subjects may then compete harder by choosing the best collaborator, because the strategy to choose a worse collaborator does not lead to a comparably high payoff anymore.

References

Adams, J. and Adams P., (1961) "Realism of Confidence Judgments", *Psychological Review*, 68 (1), 33-45.

Alpert, M., Raiffa, H. (1969) "A progress report on the training of probability assessors", Reprinted in: Kahneman, D., Slovic, P., Tversky, A., (Eds), 1982. *Judgement Under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, 294-305. Unpublished manuscript.

Beniers, K. (2005) "Party Governance and the Selection of Parliamentarians", *Tinbergen Institute Discussion Papers* no. 05-080/1.

Buehler, R., Griffin, D. (2003) "Planning, personality, and prediction: The role of future focus in optimistic time predictions", *Organizational Behavior and Human Decision Processes*, 92, 80-90.

Buehler, R., Griffin, D., Ross, M. (1994) "Exploring the „planning fallacy": Why people underestimate their task completion times", *Journal of Personality and Social Psychology*, 67, 366-381.

Camerer, C., Lovallo, D. (1999) "Overconfidence and Excess Entry: An Experimental Approach", *The American Economic Review*, 89, No. 1, 306-318.

Cesarini, D., Sandewall, Ö., Johannesson, M. (2006) "Confidence interval estimation tasks and the economics of overconfidence", *Journal of Economics Behavior and Organization*, 61, 453-470.

Clarke, F.R. (1960) "Confidence Ratings, second-choice responses, and confusion matrices in intelligibility tests", *Journal of the Acoustical Society of America*, 32, 35-46.

Daniel, K., Hirshleifer, D., Subrahmanyam, A. (1998) "Investor psychology and security market under overreactions", *Journal of Finance*, 53, 1839-1885.

De Long, J., Shleifer, A., Summers, L., Waldmann, R. (1991) "The survival of noise traders in financial markets", *Journal of Business*, 64, 1-19.

Dunning, D., Griffin, D.W., Milojkovic, J.D., Ross, L. (1990) "The Overconfidence Effect in Social Prediction", *Journal of Personality and Social Psychology*, 58, No.4, 568-581.

Epley, N., Dunning, D. (2006) "The Mixed Blessings of Self-Knowledge in Behavioral Prediction: Enhanced Discrimination but Exacerbated Bias", *Personality and Social Psychology Bulletin*, 32, No. 5, 641-655.

Fellner, G., Güth W., Maciejovsky, B. (2004) "Illusion of Expertise in Portfolio Decisions: An Experimental Aproach", *Journal of Economic Behavior and Organization*, 55, No. 3, 355-376.

Friebel, G., Raith, M. (2004) "Abuse of Authority and Hierarchical Communication", *RAND Journal of Economics*, 35, No. 2, 224-244.

Gervais, S., Odean, T. (2001) "Learning to Be Overconfident", *The Review of Financial Studies*, 14, No. 1, 1-27.

Gigerenzer, G., Hoffrage, U., Kleinbolting, H. (1991) "Probabilistic Mental Models: A Brunswikian Theory of Confidence", *Psychological Review*, 98, 506-528.

Glazer, A., Segendorff, B. (2005) „Credit claiming", *Economics of Governance*, 6, 125-137.

Greiner, B. (2003) "An Online Recruitment System for Economic Experiments", In: Kurt Kremer, Volker Macho (Eds): *Forschung und wissenschaftliches Rechnen* 2003. GWDG Bericht 63. Ges. für Wiss. Datenverarbeitung, Göttingen, 79-93.

Levy, G. (2004) "Anti-herding and strategic consultation", *European Economic Review*, 48, 503-525.

Lichtenstein, S., Fischhoff, B. (1977) “Do those who know more also know more about how much they know? The calibration of probability judgements”, *Organizational Behavior and Human Performance*, 20, 159-183.

Lord, C., Ross, L., Leppner, M. (1979) “Biased Assimilation and Attitude Polarization: The effects of Prior Theories on Subsequently Considered Evidence”, *Journal of Personality and Social Psychology*, 37, 2098-2109.

Ludwig, S., Nafziger, J. (2007) “Do You Know that I Am Biased? An Experiment”, *Bonn Econ Discussion Papers*.

Pulford, B. and Colman, A., (1997) “Overconfidence: Feedback and Item Difficulty Effects", *Personality and Individual Differences*, 23 (1), 125-133.

Rodriguez-Bailon, R., Moya M.,Yzerbyt, V. (2006) “Cuando el poder ostentado es inmerecido : sus efectos sobre la percepción y los juicios socials (English abstract title: When power is underserved: Its effects on perception and social judgements)”, *Psicothema*, 18, no. 2, 194-199.

Russo, J. E., Schoemaker, P. J. H. (1992) “Managing overconfidence”, *Sloan Management Review*, 33, 7-17.

Segendorff, B. (2000): “A Signalling Theory of Scapegoats”, Stockholm School of Economics, *SSE/EFI Working Paper Series in Economics and Finance* No. 406.

Sharp, G., Cutler, B. and Penrod, S., (1988) “Performance Feedback Improves the Resolution of Confidence Judgments", *Organizational Behavior and Human Decision Processes*, 42, 271-283.

Svenson, O. (1981) “Are we all less risky and more skilful than our fellow drivers?”, *Acta Psychologica*, 47, 143-148.

Weinstein, N.D. (1980) “Unrealistic Optimism about future life events”, *Journal of Personality and Social Psychology*, 39, 806-820.

Appendix

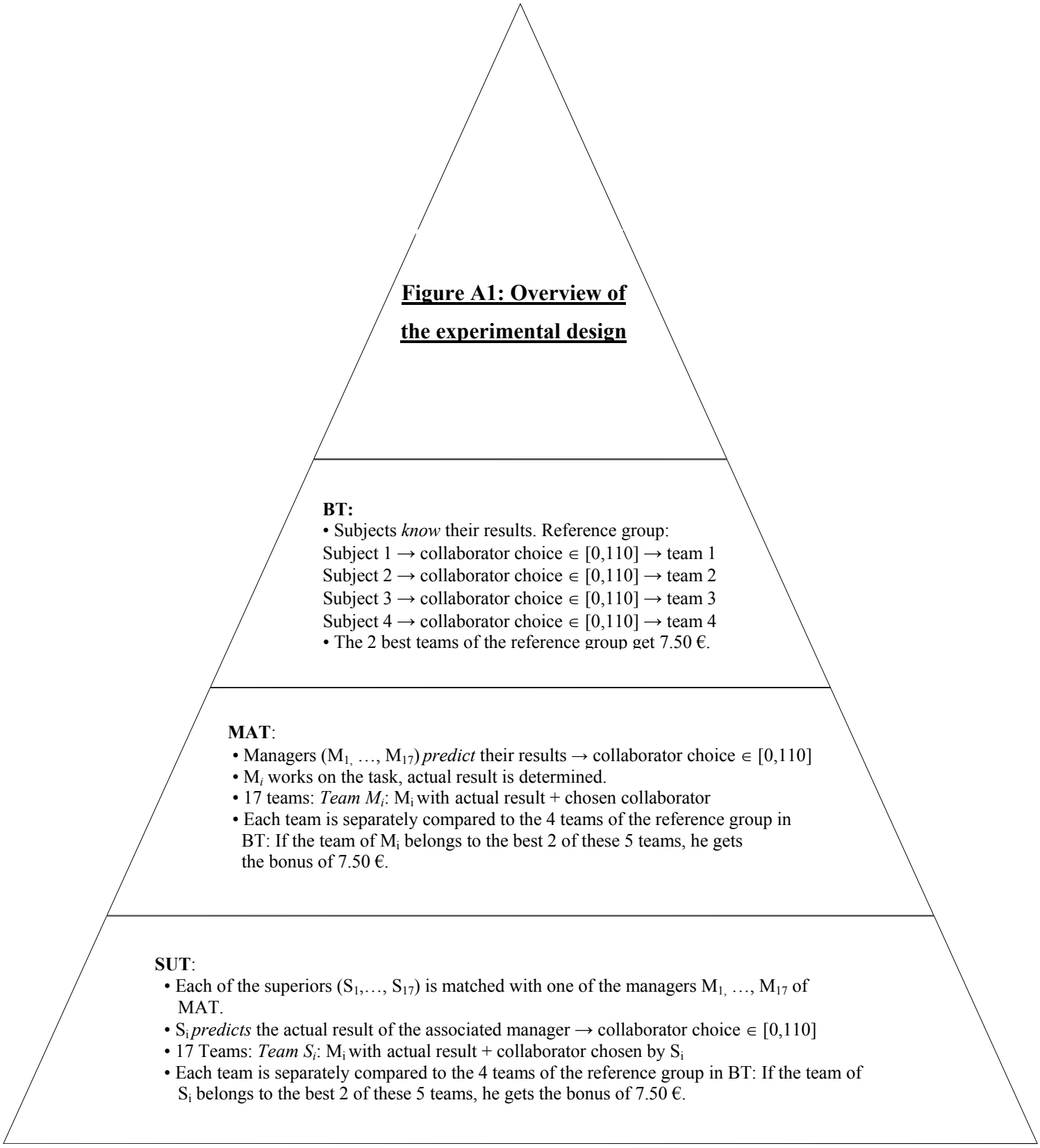


Figure A1: Overview of the experimental design
BT:
• Subjects *know* their results. Reference group:
Subject 1 → collaborator choice ∈ [0,110] → team 1
Subject 2 → collaborator choice ∈ [0,110] → team 2
Subject 3 → collaborator choice ∈ [0,110] → team 3
Subject 4 → collaborator choice ∈ [0,110] → team 4
• The 2 best teams of the reference group get 7.50 €.
MAT:
• Managers (M_1, …, M_{17}) *predict* their results → collaborator choice ∈ [0,110]
• M_i works on the task, actual result is determined.
• 17 teams: *Team* M_i: M_i with actual result + chosen collaborator
• Each team is separately compared to the 4 teams of the reference group in BT: If the team of M_i belongs to the best 2 of these 5 teams, he gets the bonus of 7.50 €.
SUT:
• Each of the superiors (S_1,…, S_{17}) is matched with one of the managers M_1, …, M_{17} of MAT.
• S_i *predicts* the actual result of the associated manager → collaborator choice ∈ [0,110]
• 17 Teams: *Team* S_i: M_i with actual result + collaborator chosen by S_i
• Each team is separately compared to the 4 teams of the reference group in BT: If the team of S_i belongs to the best 2 of these 5 teams, he gets the bonus of 7.50 €.