Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

Designing incentives and performance measurement for advisors: How to make decision-makers listen to advice

Robert M. Gillenkirch, Julia Ortner, Sebastian Robert, Louis Velthuis

May 3rd, 2023

Discussion paper number 2304

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

Robert M. Gillenkirch
Chair for Controlling
University of Osnabrück
Rolandstraße 8
49069 Osnabrück
Germany
robert.gillenkirch@uni-osnabrueck.de

Julia Ortner
Chair for Controlling
Johannes Gutenberg University
Jakob-Welder-Weg 9
55128 Mainz
Germany
julia.ortner@uni-mainz.de

Sebastian Robert
Faculty of Applied Health and Social Science
Rosenheim University of Applied Sciences
Hochschulstr. 1
83024 Rosenheim
Germany
sebastian.robert@th-rosenheim.de

Louis Velthuis
Chair for Controlling
Johannes Gutenberg University
Jakob-Welder-Weg 9
55128 Mainz
Germany
velthuis@uni-mainz.de

Designing incentives and performance measurement for advisors: How to make decision-makers listen to advice

Robert M. Gillenkirch
University of Osnabrück
robert.gillenkirch@uni-osnabrueck.de

Julia Ortner
University of Mainz
julia.ortner@uni-mainz.de

Sebastian Robert
Rosenheim University of Applied Sciences
sebastian.robert@th-rosenheim.de

Louis Velthuis
University of Mainz
velthuis@uni-mainz.de

Running Head: Designing incentives and performance measurement for advisors

Robert M. Gillenkirch, University of Osnabrück, Department of Law and Economics, Osnabrück, Germany; Julia Ortner, University of Mainz, Department of Law and Economics, Mainz, Germany; Sebastian Robert, Rosenheim University of Applied Sciences, Department of Law and Economics, Rosenheim, Germany; Louis Velthuis, University of Mainz, Department of Law and Economics, Mainz, Germany.

Data Availability: Data is available from the authors upon request.
JEL classification: D83; D91; M52.
Keywords: performance measurement, incentive design, reliance on advice; goal relatedness, incentives beliefs

Designing incentives and performance measurement for advisors: How to make decision-makers listen to advice

Abstract

In a sequence of experiments, this study investigates how the design of an advisor's performance-dependent pay affects a decision-maker's reliance on advice. In all experiments, the decision-maker forms an initial judgment, receives advice and then makes a final judgment. The advisor's compensation is manipulated to be fixed, based on individual performance, or based on group performance. We find that performance-dependent pay does not affect the decision-maker's reliance on advice unless performance measurement relates to group performance. Path model analyses show that the advisor's performance measurement affects the decision-maker's perceptions of responsibility and power, and that responsibility is the main driver of the decision-maker's cooperativeness, which mediates the relationship between performance measurement and reliance on advice. In contrast, a decision-maker's beliefs in the incentive effects of financial compensation on the quality of advice do not drive the results.

I INTRODUCTION

Before making a decision, decision-makers often receive advice – a recommendation or prediction – from another individual. How strongly a decision-maker relies on the advice she[1] receives depends on her belief about the advisor's skill and motivation to acquire, process, and communicate the relevant information Both the advisor's motivation to produce valuable advice and her willingness to give advice that is in the decision-maker's best interest can be affected by the financial incentives he has. For their design, performance measurement, i.e., the choice of performance measures as a basis for incentive compensation, is central (e.g., Baker 1992; Gibbons 1998; Prendergast 1999). Accordingly, performance measurement is a key issue in any advisory relationship, and both analytical and empirical research have investigated performance measurement of advisors. For example, analytical studies have derived characteristics of optimal

[1] Throughout this paper, we use the female form for the individual making the decision and the male form for the individual advising the decision-maker. This choice has purely expositional reasons; it neither intends to exclude any gender identity nor to make any judgmental distinction.

performance measures for subordinates providing forecasts in budgetary planning (e.g., Weitzman 1976; Osband 1989), and empirical studies have investigated the effects of performance measurement in the financial advisory industry (e.g., Hackethal, Haliassos, and Jappelli 2012). While most of these studies have addressed the direct effects of performance measurement on the behavior of those agents whose performance is measured, this study takes a different perspective in that it asks how performance measurement in the advisor's financial compensation affects *the decision-maker's reliance on advice*.

We thus investigate, in a preliminary and two main experiments (total N = 841), how a decision-maker's reliance on advice is affected by the type of performance measure used in the advisor's compensation. We model a setting in which the task is to make a judgment about an unknown fact. Initially, both the decision-maker and the advisor make their own judgments. Then, the decision-maker receives the advisor's judgment as advice and can revise her initial judgment to arrive at a final judgment. Across the two main experiments, we vary the task such that participants either face an estimation task (Experiment 1) or a choice prediction task (Experiment 2). All experiments have the same treatment manipulations: While the decision-maker is always paid based on the accuracy of her final judgment, we manipulate, between subjects, the advisor's compensation to be either independent of task performance (control condition with fixed payment), or performance-dependent. Nested within the condition with the advisor's compensation being performance-dependent, we manipulate the performance measure to reflect either the accuracy of the advice given, which represents individual performance measurement, or to reflect the accuracy of the decision-maker's final judgment after receiving advice, which represents group performance measurement. Importantly, none of the treatment manipulations creates an incentive for the advisor to strategically mislead the decision-maker.

Regarding the experimental manipulations, we develop two competing hypotheses: The *incentives beliefs hypothesis* focuses on the decision-maker's beliefs about how the financial compensation of the advisor affects the accuracy of advice. According to this hypothesis, the decision-maker will rely more strongly on advice when the advisor's compensation is performance-dependent relative to fixed, and will more strongly rely on advice when the performance measure used in the advisor's incentive pay reflects individual performance rather than group performance, as with individual performance measurement, the pay-performance sensitivity is higher than with group performance measurement. In contrast, the *goal relatedness hypothesis*, which is based on Deutsch's (1949a, 2006) theory of cooperation and competition, focuses on how incentive design affects the decision-maker's perceptions of goal relatedness. According to this hypothesis, the decision-maker will most strongly rely on advice if the advisor's compensation is based on group performance, because group performance measurement aligns financial incentives and thus reinforces the perception of positive goal interdependence, which positively affects the decision-maker's cooperativeness, i.e., her intention to listen to advice. In contrast, neither a fixed compensation nor individual performance measurement align the advisor's goal with that of the decision-maker and thus have no comparable effect on the decision-maker's reliance on advice.

Our findings give support to the goal relatedness hypothesis, but reject the incentives beliefs hypothesis. In all experiments, we find accordingly that decision-makers rely most strongly on advice in the group performance measurement condition. Path model analyses show that the effect of group performance measurement on reliance of advice is mediated, in two steps, by the decision-maker's feeling of responsibility, which is considerably stronger when performance measurement is such that the advisor's payment is aligned with that of the decision-maker, and by the decision-maker's cooperativeness. While we find that group performance measurement also affects a decision-maker's feeling of power, this feeling has no effect on cooperativeness in our first main experiment, but has a negative effect on cooperativeness in the second experiment; the effect is dominated, though, by the positive effect of responsibility. With respect to the incentives beliefs hypothesis, we find no supportive evidence in either experiment.

This study contributes to the literature and informs practice in several ways. First, it informs researchers and practitioners about incentive design in hierarchical planning and budgeting. For example, budgeting is usually characterized by a process relying on decentralized forecasts made by subordinates being responsible for organizational units (e.g., Mookherjee 2006), and research has investigated incentive designs (Weitzman 1976; Gonik 1978; Osband 1989; Chow, Cooper, and Waller 1988; Chow, Hirst, and Shields 1994) as well as other means (e.g., Chen Rennekamp, and Zhou 2015; Brüggen, Grabner, and Sedatole 2021) to improve forecast accuracy. Analytical models deriving optimal performance measures for an advisor focus on individual task performance, i.e., on paying the advisor based on how accurate his judgment turns out to be. Our findings shed new light on this research because as it shows that group performance measurement may outperform individual performance measurement once not only the effect of a financial incentive on the supplier of a forecast is considered, but also the effect on the receiver who makes subsequent decisions. More generally, our results complement research on designing incentives in hierarchical organizations in that it informs practitioners to be aware of the indirect effects incentive design has on the individuals which cooperate with those who receive the financial incentive.

Second, this study contributes to research contrasting individual versus group performance measurement. There is a large body of evidence with respect to their effects on behavior in work teams (e.g., Wageman and Baker 1997; Van Dijk, Sonnemans, and Van Winden 2001; Irlenbusch and Ruchala 2008), or on creativity and knowledge sharing in groups (e.g., Chen, Williamson, and Zhou 2012; Haesebrouck, Cools, and Van den Abbeele 2018). While these studies find evidence for performance measure design to affect behavior, they investigate the *direct effect* on the individuals receiving the respective compensations. Instead, we focus on the *indirect effect* of individual versus group performance measurement in the advisor's compensation on the decision-maker's use of advice. Our results show that performance measurement design is relevant in an advisory relationship even if the decision-maker neither expects advice to be strategically biased, nor expects the value of advice to be significantly affected by the type of the advisor's performance evaluation.

Third, our study contributes to research on advice discounting. Evidence shows that decision-makers discount advice in that they overweight their own prior beliefs and underweight the advice they receive (Bonaccio and Dalal 2006; Van Swol, Paik, and Prahl 2018). Studies have investigated various antecedents and determinants of advice discounting such as the advisor's expertise and confidence (e.g., Sniezek and Van Swol 2001; Jungermann and Fischer 2005), the decision-maker's feeling of power (e.g., Tost, Gino, and Larrick 2012), social bonds between decision-maker and advisor (e.g., Kadous, Leiby, and Peecher 2013), or the cost of advice (e.g., Gino 2008). We extend research on advice discounting in that we identify the type of performance measurement in the advisor's compensation as a determinant of advice discounting. While it has been tested how a decision-maker's option to reward the advisor affects advice discounting (Sniezek and Van Swol 2001; Sniezek, Schrah, and Dalal 2004; Van Swol and Sniezek 2005), we are not aware of any study looking at the type of performance measurement used for the reward.[2] We show that a financial incentive provided to the advisor does not affect advice discounting unless the performance measure used aligns the goals of decision-maker and advisor.

Finally, we also contribute to research on conflicts of interest between advisors and decision-makers. Examples are financial advice driven by advisors' incentives to create fees (Bolton, Freixas, and Shapiro 2007; Hackethal, Haliassos, and Jappelli 2012) or sales misguidance through sales agents (Inderst and Ottaviani 2009). Few of the respective studies look at the actual use of advice, though. Hackethal, Inderst, and Meyer (2010) provide survey evidence that private investors rely less strongly on advice when they perceive more of a conflict of interest. In an experiment, Jodlbauer and Jonas (2011) observe similar behavior. We complement these studies in that we show a positive effect of performance measurement which goes beyond avoiding or making transparent conflicts of interest.

II Basic Setting and Hypotheses

The setting of this study is the standard judge-advisor system (Bonaccio and Dalal 2006), in which two individuals, a decision-maker (the judge) and an advisor, interact with each other. We model a situation where advice is intended to help the decision-maker improve her judgment about an unknown fact or uncertain event. The decision-maker makes an initial judgment $\hat{x}_{DM1}$, receives advice $\hat{x}_A$ from the advisor, and then potentially revises her judgment to arrive at a final judgment $\hat{x}_{DM2}$. The more accurate the decision-maker believes the advice $\hat{x}_A$ to be compared to

2 Patt, Bowles, and Cash (2006) test an experimental condition equivalent to our group performance measurement condition, but do not contrast it to individual performance measurement. Decision maker and advisor interact repeatedly over two rounds, and the decision makers learns whether advice was accurate after round one. As the authors observe an effect in the second round only, it is unclear which effect (performance measurement or learning) is the main driver of the result.

her own initial judgment $\hat{x}_{DM1}$, the more strongly she should rely on it. Such a belief may be subject to bias, though: The decision-maker may overweight her own prior belief and underweight the advice she receives, because she generally perceives herself to make judgments that are superior to others' judgments (Krueger 2003), or, having better access to her own reasoning, because she sees a better justification for her own belief than for the advice (Yaniv 2004a).

While we keep constant the decision-maker's financial incentive to arrive at an final judgment $\hat{x}_{DM2}$ that is as accurate as possible, we manipulate the advisor's compensation across three experimental conditions, which we test in both the preliminary and the two main experiments we conducted: In the control condition, the advisor's pay is fixed and thus independent of performance, whereas in the other two conditions, it is performance-dependent, depending either on the accuracy of the advisor's own judgment $\hat{x}_A$ or the accuracy of the decision-maker's final judgment $\hat{x}_{DM2}$. The former case represents individual performance measurement, whereas the latter case represents group performance measurement in that the decision-maker's final judgment aggregates both her own initial and the advisor's judgment. Both types of performance measurement create an incentive for the advisor to provide the decision-maker with valuable advice. But they differ in two respects. First, individual performance measurement provides a more powerful incentive compared to group performance measurement, as the advisor's compensation is then more sensitive to the accuracy of his advice. Second, with group performance measurement the decision-maker's and the advisor's payments are congruent, whereas they differ with individual performance measurement. In the following, we derive two alternative hypotheses about the effects that these two differences have on the decision-maker's reliance on advice. We first derive a hypothesis focusing on the decision-maker's beliefs with respect to the effects of the power of financial incentives on task performance, which we term the *incentives beliefs hypothesis*. Then we develop a competing hypothesis focusing on the actors' perceptions of goal relatedness which we term the *goal relatedness hypothesis*.

The *incentives beliefs hypothesis* refers to the decision-maker's belief whether a financial incentive provided to the advisor affects his behavior such that his advice becomes more valuable. Even though financial incentives can be expected to improve task performance in general (e.g., Jenkins, Mitra, Gupta, and Shaw 1998; Camerer , and Hogarth 1999; Bonner , and Sprinkle 2002), results from studies implementing physical tasks cannot be generalized to the cognitive task considered in this study, for mainly two reasons (Camerer and Hogarth 1999; Bonner 2008; Krug and Braver 2014). First, skill is an important moderator of the relationship between a financial incentive, effort, and performance in cognitive tasks (Ashton 1990; Bonner, Hastie, Sprinkle, and Young 2000; Bonner and Sprinkle 2002). For example, when skill is low, the advisor may not expect increased effort to make his judgment more accurate and may thus not respond to the incentive. Second, the advisor may be (highly) intrinsically motivated, and the financial incentive may have a crowding-out effect such that there is no increase or even a decrease in cognitive task performance (e.g., Gneezy and Rustichini 2000; Rydval and Ortmann 2004).

The question, though, is whether the decision-maker believes in the effectiveness of a financial incentive provided to the advisor. A particular driver of this belief is an extrinsic incentives bias (Heath 1999): Even though the decision-maker may feel herself to be sufficiently intrinsically motivated to perform the task well, she may have the belief that the advisor needs an extrinsic incentive to perform. The existence of an extrinsic incentives bias has been shown in several studies. Heath (1999) demonstrates that participants' predictions of how individuals would respond to various extrinsic incentives are biased. Other studies find that participants acting as superiors tend to rely on financial incentives more than they should (e.g., Falk and Kosfeld 2006), or choose to introduce an additional financial incentive even though their subordinates' efforts would be higher without the incentive (e.g., Fehr and List 2004). If the decision-maker believes in the power of financial incentives, she will expect performance-dependent pay to motivate the advisor to produce more accurate advice, and she will expect the financial incentive to have a stronger effect the more powerful it is, i.e., the higher its pay-performance sensitivity (defined as the sensitivity of the performance measure to the accuracy of advice given). As a consequence, a decision-maker with the respective incentives beliefs will rely least strongly on advice when the advisor's compensation is fixed, and most strongly on advice when the advisor's compensation is based on individual performance, as in this case, the pay-performance-sensitivity is equal to one, whereas, with group performance measurement, pay-performance-sensitivity is lower than one unless the decision-makers gives full weight to advice and ignores her own initial judgment. This leads to the *incentives beliefs hypothesis*:[3]

H1 (incentives beliefs): Decision-makers more strongly rely on advice when advisors receive a performance-dependent financial compensation relative to a fixed payment. Reliance on advice is stronger for individual compared to group performance measurement.

The alternative hypothesis is derived from the theory of cooperation and competition proposed by Deutsch (1949a; 1949b; 2006), which has been applied to investigate a broad range of business relationships (Johnson and Johnson 1989, 2005; Tjosvold 1984, 1998). The theory states that individuals' interactions are determined by their perceptions of goal interdependence,

[3] Prior research provides consistent evidence that judges rely more strongly on advice that they evaluate to be more accurate. See, e.g., Yaniv and Kleinberger 2000; Yaniv and Milyavsky 2007.

i.e., whether their goals are positively related (cooperation), negatively related (competition), or independent. Individuals who perceive their goals to be positively related encourage and assist each other, value their respective inputs, have more positive attitudes towards each other, and show better communication (Tjosvold 1984; Deutsch 2006). A major factor determining perceptions of goal interdependence is how the respective individuals are rewarded. Accordingly, studies have investigated the effects of individual versus group performance measurement on motivation and cooperation in work teams (reviewed in DeMatteo, Eby, and Sundstrom 1998; Garbers and Konradt 2014; Nyberg, Maltarich, Abdulsalam, Essman, and Cragun 2018) and on information exchange and trust in groups (e.g., Ferrin and Dirks 2003; Taylor 2006; Kelly 2010; Super, Li, Ishqaidef, and Guthrie 2016; Hasebrouck, Cools, and Van den Abbeele 2018).

Even though the relationship between a decision-maker and an advisor is cooperative by design, as there is task interdependence (Wageman 1995) in that the advisor's task is to support the decision-maker in the decision-making process, the decision-maker's willingness to rely on advice will be affected by her perception of goal relatedness, and thus by the design of incentives and performance measurement for the advisor: Group performance measurement aligns compensations, and thus the decision-maker's perception of goal relatedness is likely to be most cooperative in the group performance measurement condition, so that she will have the most positive attitude towards the advisor, be most open to advice and have the highest appreciation for the advisor's input. Relative to the group performance measurement condition, the decision-maker will perceive her and the advisor's goal to be less positively related, or even unrelated, when the advisor's payment is fixed. We thus expect the decision-maker to have a less cooperative attitude towards the advisor in the control condition. When the advisor's compensation is based on individual performance, we also expect the decision-maker's attitude to be less cooperative. It is not clear, however, whether the decision-maker's perception of goal relatedness remains in the cooperative domain or is even shifted towards being competitive: Providing the advisor with a financial incentive to produce accurate advice is in the best interest of the decision-maker, and so the decision-maker may still perceive the situation as cooperative. But she may also perceive the situation as one of social comparison ("you have your judgment, I have mine, let's see who is better") and thus show an attitude of competitiveness (Festinger 1954; Garcia, Tor, and Schiff 2013) that makes her less inclined to give weight to advice. Consequently, there are two potential effects of individual performance measurement, one making the decision-maker perceive the situation as more cooperative, the other as less cooperative relative to the control condition. As we see no dominance of either argument, we only hypothesize that the most cooperative attitude of the decision-maker, resulting in the strongest reliance on advice, will be observed in the group performance measurement condition:

H2 (goal relatedness): Decision-makers more strongly rely on advice when advisors receive a financial compensation depending on group performance, both relative to a fixed payment and relative to a financial compensation based on individual performance.

III First Results and Design of Main Experiments

A preliminary experiment was conducted with a total of 120 student participants. The task was to estimate the percentage of individuals, from a sample of university students having specific personal characteristics such as playing a music instrument or having travelled to a non-European country. The experiment had 20 rounds in which each decision-maker interacted exactly twice with each advisor based on a random matching schedule. In each round, the respective question was presented, independent estimations $\hat{x}_{DM1}$ and $\hat{x}_A$ were made, the decision-maker was informed about $\hat{x}_A$, and then had the opportunity to revise her estimation from the initial $\hat{x}_{DM1}$ to a final $\hat{x}_{DM2}$. While decision-makers were always paid according to the accuracy of their final estimations, according to the treatment conditions, advisors either received a fixed payment (*FIXED*), or a payment that depended on the accuracy of their own estimations $\hat{x}_A$ (individual performance measurement, *IPM*), or a payment that depended on the accuracy of their respective decision-maker's final estimations $\hat{x}_{DM2}$ (group performance measurement, *GPM*). The weight a decision-maker gives to advice is defined by

$$woa = \frac{\hat{x}_{DM2} - \hat{x}_{DM1}}{\hat{x}_A - \hat{x}_{DM1}} \quad (1)$$

woa (weight of advice) is equal to zero when the decision-maker ignores advice ($\hat{x}_{DM2} = \hat{x}_{DM1}$), whereas it is equal to one when she fully relies on advice ($\hat{x}_{DM2} = \hat{x}_A$). Results give support to treatment manipulations affecting decision-makers' reliance on advice: While decision-makers give similar weight to advice in the *FIXED* and the *IPM* conditions (*FIXED*: $woa = 0.430$; *IPM*: $woa = 0.416$), they considerably more strongly rely on advice in the *GPM* condition ($woa = 0.506$; Wilcoxon tests: *GPM* vs. *IPM*, $z_{38} = 2.16$, $p = 0.030$; *GPM* vs. *FIXED*, $z_{38} = 1.68$, $p = 0.097$) [4].

Based on these results, we planned two main experiments.[5] Between the two experiments, the experimental task was changed so that the experiments would complement each other in providing evidence about the process that translates the differences in performance measurement for the advisor into decision-makers' reliance on advice (Asay, Guggenmos, Kadous, Koonce, and Libby 2022). The tasks differ in that effort plays a minor role in arriving at an accurate

[4] Throughout the paper, all p-levels are reported two-tailed.

[5] A power analysis was performed in order to derive sample sizes. Calculations were based on observed effect sizes in *woa* (*GPM* vs. *FIXED*: d = 0.5027; sample size for $\alpha = 0.05$ and $1-\beta = 0.8$, as recommended by Cohen 1992: N = 52 per condition)

judgment in the first experiment, but a (relatively) major role in the second experiment. In the first experiment, participants faced a standard estimation task in that they had to estimate rents for *airbnb* apartments in Berlin. While decision-makers received only basic information about the respective apartments, advisors were additionally given a short list of apartment characteristics driving rents. In the second experiment, participants faced a prediction task in that they had to predict the ranks of consumption alternatives according to the frequencies with which the alternatives were chosen by test subjects. Again, advisors received additional information, but now the information was a long list of statements that would help advisors to make better predictions if they made efforts to extract and correctly interpret the relevant statements. At the end of each experiment, participants completed a questionnaire with sets of items reflecting the constructs behind the two hypotheses. Both experiments were run online due to COVID-19 pandemic restrictions.[6]

IV Experiment 1

Design, Procedures, and Data

The task in Experiment 1 is a standard estimation task: Participants (20 in each session) make estimates about the weekly rate for an *airbnb* apartment for two individuals in Berlin. Information about the respective apartments (location, list of amenities, photos) was taken from the *airbnb* website and pseudonymized. After three practice rounds, roles were assigned randomly. Advisors received additional information about which apartment characteristics (location, amenities, individual versus shared bathroom or kitchen) would be especially relevant

6 We ensured that conditions were as close to lab conditions as possible: Recruited subjects logged into a video conference (with both micro and camera switched off), where they remained during the whole experiment and received all instructions, which were presented on screen and read out loud. Procedures ensured an identical level of anonymity for participants as a regular lab experiment.

for estimating their rents. Decision-makers were only informed that advisors would receive such additional information. Then, decision-makers and advisors interacted over ten rounds in a stranger design. Finally, all participants completed a questionnaire. Accuracy was measured based on the absolute deviation of the respective estimation from the actual prices on the airbnb website. Decision-makers received a payment that depended on the accuracy of their final estimations $\hat{x}_{DM2}$ (ranging from 0.00 EUR to 1.00 EUR per round)[7]. Depending on the treatment condition, advisors' compensations were either fixed (*FIXED*), or depended on the accuracy of their own estimations $\hat{x}_{A}$ (*IPM*) using the same payment scheme as for decision-makers, or were congruent with the decision-makers' payments (*GPM*). All participants received an additional show-up fee of 5 EUR.

Overall, 339 students from two large universities participated, with 52 decision-makers and 55 advisors in the *FIXED* condition, 58 decision-makers and 58 advisors in the *IPM* condition, and 57 decision-makers and 59 advisors in the *GPM* condition, which gives a total of 1,670 observations.[8] Following prior studies analyzing estimation tasks (e.g., Harvey and Fisher 1997, Yaniv 2004b; Logg, Minson, and Moore 2019), we omitted faulty or outlier data, losing 20 observations.[9] The measure *woa* = $(\hat{x}_{DM2} - \hat{x}_{DM1})/(\hat{x}_{A} - \hat{x}_{DM1})$ has some measurement issues, as it is not defined when $\hat{x}_{A} = \hat{x}_{DM1}$, can be negative (if, for example, the decision-maker increases her estimate even though $\hat{x}_{A} < \hat{x}_{DM1}$), and can exceed 1. We dealt with the latter two cases by winsorizing *woa* to lie in [0,1].[10]

7 The payment was 1 EUR if the deviation did not exceed 50, 0.75 EUR for any deviation between 51 and 100, 0.50 EUR for deviations between 101 and 150, 0.25 EUR for 151-200, and 0 EUR else.

8 Matching protocols were programmed such that drop-outs would not affect remaining participants. There were three incidents of a drop-out in the *FIXED* condition and two in the *GPM* condition.

9 We dropped two observations that could only be explained by decision-makers' mistyping, and we applied a three-sigma filter on advisors' estimations, eliminating observations where advice was more than three standard deviations distant from the average advice.

10 For 17 observations *woa* < 0 was changed to *woa* = 0, for 18 observations *woa* > 1 was changed to *woa* = 1.

Reliance on Advice: Descriptive and Test Results

Table 1 presents summary statistics on weight of advice (*woa*) and mean absolute errors (*MAE*) of estimations in Panel A, and results from non-parametric tests in Panel B. The data show that in the *GPM* condition, where *woa* = 0.488, decision-makers give significantly higher weight to advice than in either the *FIXED* (*woa* = 0.415) or the *IPM* conditions (*woa* = 0.400), whereas there is no significant difference between the two latter conditions. This contradicts H1 (incentives beliefs) which states that *woa* is highest for the *IPM* condition and lowest for the *FIXED* condition, and supports H2 (goal relatedness) in that *woa* is highest in the *GPM* condition. As assignment of roles was random, decision-makers could expect advisors, on average, to make estimations at least as accurate as their own initial estimations. In contrast, advisors' estimations $\hat{x}_{A}$ were, on average, slightly less accurate than decision-maker's initial estimations $\hat{x}_{DM1}$ in both the *IPM* and the *GPM* conditions. Using *woa* = 0.5 as a benchmark to check for advice discounting, we see that the average *woa* is significantly lower than the benchmark for both the *FIXED* and the *IPM* conditions, but not for the *GPM* condition. The data in Table 1 further show that over all conditions, decision-makers significantly improve their estimations from $\hat{x}_{DM1}$ to $\hat{x}_{DM2}$ in all three conditions. While decision-makers' average final estimations are most accurate in the *GPM* condition, the difference to the other conditions is statistically insignificant.[11]

Process Evidence on Treatment Effects

[11] To check whether decision-makers actually should have given more weight to the advice they received, we simulated estimations $\hat{x}_{DM2}$ that would have resulted from *woa* = 0.5 applied to every round and decision-maker. We find the accuracies of these simulated estimations being significantly higher than the accuracies of actual estimations $\hat{x}_{DM2}$ in each condition (t-tests, $p < 0.01$ in all cases), which is in line with previous results about the benefits of averaging estimations (e.g., Soll and Larrick 2009).

To provide process evidence on treatment effects, we performed analyses based on questionnaire data addressing various factors that we expected to mediate or moderate the relationship between the treatment manipulations and reliance on advice according to the alternative hypotheses. Panel A of Table 2 presents items from the post-experimental questionnaire[12], descriptive statistics, and summarizes results of non-parametric tests of treatment effects. H1 implies a moderating effect of incentives beliefs in that a decision-maker's beliefs about the effects of financial incentives on the accuracy of advice affect her behavior. Incentives beliefs were elicited using four items (Q_1-Q_4). H2 implies a mediation effect of goal relatedness in that group performance measurement (*GPM*) increases the decision-maker's cooperativeness, which in turn strengthens her reliance on advice. The mediation is potentially one-step in that *GPM* directly affects cooperativeness, or two-step in that additional factors mediate the effect of *GPM* on cooperativeness. In particular, as the relationship between decision-maker and advisor is hierarchical (the decision-maker makes the final judgement), the treatment manipulations can be expected to affect the decision-maker's feeling of responsibility and power. The two feelings can have opposing effects on cooperativeness: Prior research has emphasized that cooperation activates group members' feelings of being responsible for their work outcomes, and that such feelings of responsibility in turn positively affect cooperativeness (Harvey and Fischer 1997; Van Der Vegt, Emans, and Van De Vliert 1998; Poile 2017). GPM makes the decision-maker responsible for the advisor's compensation, whereas there is no such effect with either IPM or fixed pay. In contrast, if the decision-maker feels more powerful, she may be not more, but less inclined to listen to advice

[12] The questionnaire had two parts, a first part with items that specifically referred to the situation of the experiment, and a second part in which participants were asked about their general attitudes towards cooperation and responsibility. The evidence we present is based on data from the first part. We checked whether factors extracted from the second part of the questionnaire contribute to explaining the moderations and mediations implied by the two hypotheses, but they don't.

(Tost, Gino, and Larrick 2012).[13] Accordingly, questionnaire items address perceptions of goal relatedness, cooperation, and intention to accept advice (Q_5-Q_{10}), responsibility (Q_{11}-Q_{13}), and power (Q_{14}).

The descriptive data show that decision-makers' incentives beliefs do not differ across treatments. In contrast, the majority of the other items show pronounced treatment effects, having significantly higher levels for the *GPM* condition compared to both the *IPM* and the *FIXED* conditions. For all questionnaire items, a principal component factor analysis and subsequent evaluations of construct measurements were performed; results are presented in Panel B of Table 2. The analysis gives four factors explaining 72 percent of the total variance. Decision-makers' incentives beliefs are reflected in two factors, differentiating between the beliefs about how financial incentives affect effort and task performance, respectively (*inbe-effort*, Q_1 & Q_3, and *inbe-taskperf*, Q_2 & Q_4). All items addressing the decision-maker's perception of goal relatedness and cooperation are reflected in the single factor *cooperativeness* (Q_5 - Q_{10}). The factor combines a decision-maker's feeling of relatedness with her intention to listen to advice and her belief in the usefulness of working together with the advisor on the estimation task. All three items addressing the decision-maker's feeling of being responsible for the advisor reflect the factor *responsibility* (Q_{11} - Q_{13}). Finally, *power* is a single item (Q_{14}) variable.[14]

Based on the exploratory factor analysis, we performed various analyses providing evidence about the processes hypothesized in H1 and H2. First, we checked for the potential moderation

[13] Van Swol and Sniezek (2005) model a setting in which either the judge or the advisor has the power to decide about the advisor's reward. Here the effect of giving rewarding power to the judge is likely reverse, as she will justify her rewarding decision by using the advice.

[14] Measurement reliability is assessed by computing Cronbach's alpha for each factor (see Table 2). As construct measurement for *cooperativeness* is slightly improved after dropping Q_{10}, only items Q_5-Q_9 are included.

of incentives beliefs, performing both analyses of variance based on median splits and regression analyses. In none of the analyses do we find evidence for a moderation; Appendix A gives details. Second, we performed path model estimations (in AMOS) based on the mediation effects implied by the goal relatedness hypothesis H2. Figure 1 shows estimation results (standardized path coefficients and significance levels) of the main model we estimated. The model includes cooperativeness (Q_5-Q_9), responsibility (Q_{11}-Q_{13}), and power (Q_{14}) as mediators, but leaves out any incentives beliefs moderation. The independent variable is coded according to hypothesis H2, i.e., 0 for both the *FIXED* and the *IPM* condition, and 1 for the *GPM* condition. The dependent variable is the average weight the respective decision-maker gives to advice as defined in (1). The model has good fit, as it is insignificantly different from the perfect fit model ($\chi^2_{37} = 46.7$, $p = 0.13$) and as goodness of fit indices have satisfactory levels (goodness of fit index: 0.951; incremental fit index: 0.989; comparative fit index: 0.988). The path coefficients show that the process explaining how group performance measurement translates into stronger reliance on advice is a two-step mediation involving cooperativeness on the second step and responsibility, but not power, on the first step: Even though group performance measurement has a significantly positive effect on both responsibility and power, only its effect on responsibility translates into cooperativeness, which drives weight of advice. For all mediators, by including direct paths to weight of advice, the model checks whether the mediation is full or only partial. The estimation results show full mediations: Neither the independent variable *GPM* nor the first step mediators responsibility and power have a significant direct effect on *woa*.

In order to check for the robustness of these results, we estimated alternatives to the path model presented in Fig. 1. First, we included the incentive beliefs moderation into the path model. We tested several model variants, with different proxies for incentives beliefs (*inbe-*

effort, *inbe-taskperf*, or both factors) and different types of moderation (moderation of the path $GPM \rightarrow woa$, or of the path $GPM \rightarrow$ cooperativeness). Results (untabulated) show no significant moderation in any of the model variants tested, while all mediation results remain virtually unchanged. We conclude from these results that we have to reject H1, and find strong support for H2.

V Experiment 2

Design and Procedures

In Experiment 2, the task is changed such that advisors can, by investing some effort, gather valuable information to improve their judgments. For this reason, the task is made more complex in that participants have to predict consumer choices by assigning ranks to four consumption alternatives a - d. A prediction is thus a vector of ranks, $\{\hat{r}(a), \hat{r}(b), \hat{r}(c), \hat{r}(d)\}$. Actual ranks $\{r(a), r(b), r(c), r(d)\}$ are derived from the frequencies with which the respective consumption alternatives (e.g., computer accessories, pasta dishes, or day trip offers) were chosen by 100 test subjects. Participants of the experiment are given the same information about the alternatives as the test subjects received, including a photo and a table of product characteristics. Payments are calibrated as in Experiment 1, based on a different measure for accuracy, though: Accuracy is measured as the Euclidean distance $d(\hat{r}, r)$ between the respective predicted rank $\hat{r}$ and the actual rank r:[15]

$$d(\hat{r}, r) = \sum_{i=a,b,c,d} (\hat{r}(i) - r(i))^2 \tag{2}$$

[15] Questions for the 10 experimental rounds were selected such that there were no tied ranks. In this case, the Euclidean distance is a linear transformation of the Spearman rank correlation coefficient.

With four alternatives and four ranks, the Euclidean distance has a minimum of zero and a maximum of 20.[16] A decision-maker's use of advice is measured based on the Euclidean distance between her final judgment and the advice she receives:

$$d(\hat{r}_{DM2}, \hat{r}_A) = \sum_{i=a,b,c,d} (\hat{r}_{DM2}(i) - \hat{r}_A(i))^2 \qquad (3)$$

The Euclidean distance as defined in (3) is an indirect measure of weight of advice: The greater weight a decision-maker gives to advice, the smaller is the distance between her final prediction and the advice she received. $d(\hat{r}_{DM2}, \hat{r}_A)$ makes no difference between ranks. We will provide additional measures in the data analyses to account for these issues.

After three pratice rounds, all participants were informed that advisors would receive specific information about the test subjects' preferences. The information was a list of statements from which advisors could infer the ranks relatively precisely. Decision-makers knew that advisors would see a list, but were not informed about the quality of the information. As in the first experiment, after the ten prediction rounds all participants completed a survey. Overall, 382 students from two large universities participated.[17]

Reliance on Advice: Descriptive and Test Results

Table 3 presents summary statistics for Experiment 2 in Panel A, and results from non-parametric tests in Panel B.[18] It first shows that decision-makers make initial rank predictions ($\hat{r}_{DM1}$) which are almost equally accurate across conditions: $d(\hat{r}_{DM1}, r)$-values show only very small differences. Advisors' predictions are considerably more accurate than decision-maker's

16 A performance-dependent payment of a round was set to equal 1 EUR for a distance of zero and was reduced by 0.05 EUR for every unit of distance.

17 64 decision-makers and 65 advisors in the *FIXED* condition, 63 decision-makers and 64 advisors in the *IPM* condition, and 63 decision-makers and 63 advisors in the *GPM* condition.

18 A post-hoc power analysis gives (with $\alpha = 0.05$), for the actual values of the weight of advice proxy $d(\hat{r}_{DM2}, \hat{r}_A)$, a power of $1-\beta = 0.991$ for the effect *GPM* vs. *IPM* and $1-\beta = 0.768$ for *GPM* vs. *FIXED*.

initial predictions, indicating that the additional information was helpful. Furthermore, advisors make significantly more accurate predictions with performance-dependent pay present relative to absent. This result is consistent with performance-dependent pay increasing the advisors' efforts. The relation between the *IPM* and *GPM* condition, though, does not support the view that the power of the financial incentive is the driver of advisor behavior, as advice is most accurate not in the *IPM* condition, where the pay-performance sensitivity is the highest, but in the *GPM* condition instead.

Our main measure for reliance on advice is the Euclidean distance $d(\hat{r}_{DM2}, \hat{r}_A)$ between the decision-maker's final prediction $\hat{r}_{DM2}$ and the advice $\hat{r}_A$ she receives. The respective data show that final predictions are significantly closer to advice in the *GPM* condition than in either the *FIXED* or the *IPM* condition, whereas, even though advisors give significantly more accurate advice in the *IPM* condition than in the *FIXED* condition, decision-makers final predictions are insignificantly closer to that advice.[19] As a second measure, we use the frequency with which a decision-maker accepts advice. We say that a decision-maker accepts advice for a specific rank if she revises her initial prediction to the advice given. Panel A of Table 3 shows frequencies of advice acceptance per rank and overall. The frequency data confirm that decision-makers rely most strongly on advice in the *GPM* condition, and that they accept advice significantly more frequently in the *GPM* compared to the *IPM* condition, irrespective of the rank being advised. The differences in advice acceptance translate into differences in final predictions' accuracies: Even though decision-makers start with approximately equally accurate predictions in all three conditions, they end up with the most accurate final predictions in the *GPM* condition. That is,

[19] We checked whether these results can be explained by the distances between advice and initial predictions, $d(\hat{r}_{DM1}, \hat{r}_A)$, i.e., whether decision-makers' final rank predictions are closer to advice in the *GPM* compared to the *IPM* condition simply because initial predictions are already closer to advice. This is not the case. Initial predictions are closer to advice, though, in the *GPM* relative to the *FIXED* condition.

decision-makers significantly profit from giving more weight to advice. These results contradict hypothesis H1 (incentives beliefs) and give support to hypothesis H2 (goal relatedness).

Process Evidence on Treatment Effects

As for Experiment 1, we used data from the post-experimental questionnaire to investigate the mediation and moderation effects implied by the two alternative hypotheses. The questionnaire had the same items Q_1-Q_{14} as presented in Table 2 except for changes in wordings due to the different task. Table 4 presents descriptive data and results from a principal component analysis of items. Results are very similar to the results from Experiment 1: The same four factors are identified (*cooperativeness*, *responsibility*, two factors for *incentives beliefs*), explaining 66 percent of the total variance, and *power* is again represented by a single item. Item averages are very similar to the data from Experiment 1: decision-makers' *incentives beliefs* do not differ across treatments, whereas all items reflecting *cooperativeness*, *responsibility*, and *power* show considerably higher values for the *GPM* condition.

As for Experiment 1, we checked for the moderating effect of incentives beliefs on the decision-makers' reliance on advice, both by use of ANOVA based on median splits and of regression analyses. As in Experiment 1, we find no indication of a moderating effect of incentives beliefs; see Appendix A for details. That is, even though providing a financial incentive to advisors actually makes advice more valuable, decision-makers' beliefs in this effect do not drive their reliance on advice. We also performed path model estimations (in AMOS) to investigate the mediation process hypothesized by H2 (goal relatedness). Figure 2 shows the estimation results (standardized path coefficients and significance levels). The model is identical to the model we estimated with data from Experiment 1, i.e., it includes *cooperativeness* (items Q_5-Q_9), *responsibility* (items Q_{11}-Q_{13}), and *power* (single item Q_{14}) as mediators, but no incentives beliefs moderation. The independent variable is an indicator variable which is equal to 1 for the *GPM* condition and 0 else. As the dependent variable we use the proximity of the decision-maker's final rank prediction to the advice she receives, i.e., the reverse measure of $d(\hat{r}_{DM2}, \hat{r}_A)$, so that positive path coefficients indicate a reinforcing effect on decision-makers' reliance on advice. Even though the model is significantly worse than a perfect fit model ($\chi^2_{37} = 55.6$, $p = 0.03$), goodness of fit measures are close to or above the usual thresholds (goodness of fit index: 0.948; incremental fit index: 0.979; comparative fit index: 0.979). The estimation results give support to the goal relatedness hypothesis H2, as group performance measurement positively affects cooperativeness, which in turn increases reliance on advice. As in Experiment 1, all mediations are full: no direct path towards the dependent variable is significant. In contrast to Experiment 1, we now see that power has a negative effect on cooperativeness, which means that group performance measurement affects a decision-maker's reliance on advice not only through its positive effect on the decision-maker's feeling of being responsible for the advisor, but also negatively through the decision-maker's feeling of power. Comparing the standardized path coefficients shows, though, that the effect on responsibility clearly dominates; the aggregate effect of the *GPM* indicator variable on cooperativeness is $0.763 \cdot 0{,}503 - 0.205 \cdot 0.227 = 0.337$, which is significantly positive ($p < 0.001$). We conclude from these results that, as in Experiment 1, incentives beliefs do not drive decision-makers' reliance on advice, whereas responsibility has a strong, positive effect on cooperativeness, and is highly significantly affected by the choice of the performance measure for the advisor's compensation.[20]

[20] As for Experiment 1, we positively checked for the robustness of the results by estimating variants of the path model. In particular, we included the incentives beliefs moderation into the model, but still found no evidence for an incentives beliefs moderation.

VI Discussion and Conclusion

We report results from a preliminary and two main experiments that examine how incentive and performance measurement design for an advisor affect a decision-maker's reliance on advice. The study does not focus on the direct effects of incentives and performance measurement on the individual providing a decision-maker with valuable information, but on their *indirect effect* on the use of such information by the decision-maker. Our results reject the incentive beliefs hypothesis which states that a decision-maker more strongly relies on advice, the more powerful the advisor's financial incentives are, as we find that, relative to giving advisors a fixed compensation, paying them according to the accuracy of their advice (*IPM*, giving the highest pay-performance-sensitivity) does not affect decision-makers' reliance on advice. Instead, we find that a decision-maker gives significantly stronger weight to advice with group performance measurement (*GPM*), when the advisor's compensation is aligned with her own. The effect of *GPM* on reliance on advice is mediated by the decision-maker's feeling of being responsible for the advisor, which positively affects her cooperativeness. Our results thus give strong support for the goal relatedness hypothesis. In the following, we briefly discuss three aspects of our results.

First, our study has important implications for designing budget-based incentives in organizations. While research has focused on how such incentives should be designed to elicit accurate forecasts from subordinates (e.g., Young 1985; Chow, Cooper, and Waller 1988; Chow, Hirst, and Shields 1994; Chen, Rennekamp, and Zhou 2015; Scheele, Thonemann, and Slikker 2018), available evidence relates to these forecasts only, but not to the use of these forecasts by superiors. Our experimental evidence suggests that such a focus on individual forecasting accuracy is too narrow, as corporate performance equally depends on accurate forecasts made by subordinates and on the optimal use of these forecasts by superiors in their corporate planning and decision making. Our results indicate that this use may critically depend on how incentives and performance measurement for subordinates are designed.

Second, our study informs about the role of performance measurement design in promoting cooperation in hierarchical relationships. Deutsch's theory of cooperation and competition refers to how the goals of individuals involved in a specific situation are related: positive (cooperation) or negative (competition) (Deutsch 1949, 2006; Johnson and Johnson 1989; Tjosvold 1984). The theory has been initially developed without explicit reference to hierarchies, but has been applied to superior-subordinate relationships. Here, positive goal interdependence does not guarantee cooperative behavior, because it also affects perceptions of power and influence. Prior studies (See, Morrison, Rothman, and Soll 2011; Tost, Gino, and Larrick 2012) directly manipulated a decision-maker's power, but not the monetary consequences of judgment accuracies for either the decision-maker or the advisor. They find that the feeling of power leads decision-makers to discount advice, and that an important driver of advice discounting is the negative effect of power on a decision-maker's cooperativeness. In our study, the manipulation of the advisor's compensation affects not only the decision-maker's power, but also her responsibility, and we find that the driving force behind her cooperativeness is her feeling of responsibility, not power. Tjosvold (1989) points out the importance of defining power not a way that creates a situation of competitiveness. We show that hierarchical relationships can be designed in such a way.[21]

Third, our study complements research on the role of individuals' beliefs about other individuals' financial incentives for decision making. Data from Experiment 1 indicate that

[21] Haesebrouk, Cools, and Van den Abbeele (2018) examine how individual versus group performance measurement affects knowledge sharing in groups with status differences, when these differences do not imply differences in power or responsibility. They find that status differences may have a positive effect on knowledge sharing. Our results complement this finding in that we show that GPM serves as a debiasing measure for decision-makers.

decision-makers have an extrinsic incentives bias, as they show considerably stronger incentives beliefs than advisors in this experiment; however, this bias does not drive their behavior.[22] In contrast, we find that both decision-makers and advisors are equally aware of the cooperative effect of *GPM* in both experiments, as items reflecting *cooperativeness* show strong differences in means across treatments not only for decision-makers, but also for advisors.[23] This joint awareness of the benefits of goal alignment, in combination with the prediction accuracy being highest under *GPM*, is consistent with prior evidence on cooperative social interactions in groups increasing individual motivation(e.g., Hertel, Kerr, and Messé 2000):

We conclude by pointing out three limitations of this study, each of which implies future research opportunities. First, in our experiments the information content and the process of information exchange are kept simple whereas in real relationships, both are likely more complex.[24] Additional studies will be needed to examine the group performance measurement effect we observe in such more complex settings. Second, the interaction between decision-maker and advisor was such that decision-makers received no feedback about the accuracies of their initial and final judgments. As a consequence, decision-makers could not learn to listen to advice, and it would be interesting to see whether the performance measurement manipulation of our study extends to learning behavior. Third, our study is designed such that neither decision-makers nor advisors would expect estimations or predictions to be strategically biased. In reality though, individuals may have incentives to intentionally bias the information they provide to

---

22 We find no extrinsic incentives bias in Experiment 2: Questionnaire data show that advisors have almost equally strong incentives beliefs as decision -makers in this experiment.

23 Factor *cooperativeness*, *GPM* vs. *IPM*: 5.33 vs. 4.76 for decision-makers and 4.47 vs. 3.47 for advisors in Experiment 1, 5.66 vs. 4.88 for decision-makers and 5.36 vs. 4.05 for advisors in Experiment 2.

24 For example, advisors will give justifications for their advice, indicate their confidence, or provide additional information. Also, information exchange will not be one-directional, but bi-directional, and the decision-maker's and the advisor's judgments will likely affect each other in the judgment and decision-making process.

others. We see the absence of strategic motives as a strength of our study since it provides a benchmark for the effects of performance measurement design for advisors on a decision-maker's feeling of responsibility and power, cooperativeness, and reliance on advice. However, all of the effects observed will likely be affected by strategic motives being present.

References

Asay, H. S., R. D. Guggenmos, K. Kadous, L. Koonce, and R. Libby. 2022. Theory testing and process evidence in accounting experiments. *The Accounting Review* 97 (6): 23-43.

Ashton, R. H. 1990. Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback, and Justification. *Journal of Accounting Research* 28: 148-180.

Baker, G. P. 1992. Incentive contracts and performance measurement. *Journal of Political Economy* 100 (3): 598-614.

Bolton, P., X. Freixas, and J. Shapiro. 2007. Conflicts of interest, information provision, and competition in the financial services industry. *Journal of Financial Economics* 85 (2): 297-330.

Bonaccio, S., and R. S. Dalal. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101 (2): 127–151.

Bonner, S. E. 2008. *Judgment and Decision Making in Accounting*. Upper Saddle River, NJ: Pearson Education.

Bonner, S. E., and G. B. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27 (4): 303-345.

Bonner, S. E., R. Hastie, G. B. Sprinkle, and S. M. Young. 2000. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management and Accounting Research* 12 (1): 19–64.

Brüggen, A., I. Grabner, and K. L. Sedatole. 2021. The folly of forecasting: The effects of a disaggregated demand forecasting system on forecast error, forecast positive bias, and inventory levels. *The Accounting Review* 96 (2): 127-152.

Camerer, C. F., and R. M. Hogarth. 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19: 7–42.

Chen, C.X., K. M. Rennekamp, and F. H. Zhou. 2015. The effects of forecast type and performance-based incentives on the quality of management forecasts. *Accounting, Organizations and Society* 46: 8-18.

Chen, C.X., M. G. Williamson, and F. H. Zhou. 2012. Reward system design and group creativity: An experimental investigation. *The Accounting Review* 87 (6): 1885-1911.

Chow, C. W., J. C. Cooper, and W. S. Waller. 1988. Participative budgeting: Effects of a truth-inducing pay scheme and information asymmetry on slack and performance. *The Accounting Review* 63 (1): 111-122.

Chow, C. W., M. K. Hirst, and M. D. Shields. 1994. Motivating truthful subordinate reporting: an experimental investigation in a two-subordinate context. *Contemporary Accounting Research* 10 (2): 699-720.

Cohen, J. 1992. Statistical power analysis. *Current Directions in Psychological Science* 1 (3): 98-101.

DeMatteo, J. S., L. T. Eby, and E. Sundstrom. 1998. Team-based rewards: Current empirical evidence. *Research in Organizational Behavior* 20, 141-183.

Deutsch, M. 1949a. A theory of co-operation and competition. *Human Relations* 2 (2): 129-152.

Deutsch, M. 1949b. An experimental study of the effects of cooperation and competition upon group process. *Human Relations* 2 (3): 199-231.

Deutsch, M. 2006. Cooperation and competition. In *The Handbook of Conflict Resolution: Theory and Practice*, edited by M. Deutsch, P. T. Coleman, and E. C. Marcus, 23-42. Hoboken NJ, John Wiley and Sons.

Falk, A., and M. Kosfeld. 2006. The Hidden Costs of Control. *The American Economic Review* 96 (5): 1611-1630.

Fehr, E., and J. A. List. 2004. The Hidden Costs and Returns of Incentives-Trust and Trustworthiness among CEOs. *Journal of the European Economic Association* 2 (5): 743-71.

Ferrin, D. L., and K. T. Dirks. 2003. The use of rewards to increase and decrease trust: Mediating processes and differential effects. *Organization Science* 14 (1): 18-31.

Festinger, L. (1954). A theory of social comparison processes. Human relations, 7(2), 117-140.

Garbers, Y., and U. Konradt. 2014. The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of Occupational and Organizational Psychology* 87 (1): 102-137.

Garcia, S. M., A. Tor, and T. M. Schiff. 2013. The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science* 8 (6): 634-650.

Gibbons, R. 1998. Incentives in organizations. *Journal of Economic Perspectives* 12(4): 115-132.

Gino, F. 2008. Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes* 107 (2): 234-245.

Gneezy, U., and A. Rustichini. 2000. Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115 (3): 791-810.

Gonik, J. 1978. Tie salesmen's bonuses to their forecasts. *Harvard Business Review* 56 (3): 116-123.

Hackethal, A., M. Haliassos, and T. Jappelli. 2012. Financial advisors: A case of babysitters? *Journal of Banking and Finance* 36 (2): 509-524.

Hackethal, A., R. Inderst, and S. Meyer. 2010. Trading on advice. *Working Paper*.

Haesebrouck, K., M. Cools, and A. Van den Abbeele. 2018. Status differences and knowledge transfer: The effect of incentives. *The Accounting Review* 93 (1): 213-234.

Harvey, N., and I. Fischer. 1997. Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes* 70 (2): 117-133.

Heath, C. 1999. On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives. *Organizational Behavior and Human Decision Processes* 78 (1): 25-62.

Hertel, G., N. L. Kerr, and L. A. Messé. 2000. Motivation gains in performance groups: paradigmatic and theoretical developments on the Köhler effect. *Journal of Personality and Social Psychology* 79 (4): 580.

Inderst, R., and M. Ottaviani. 2009. Misselling through agents. *American Economic Review* 99 (3): 883-908.

Irlenbusch, B., and G. K. Ruchala. 2008. Relative rewards within team-based compensation. *Labour Economics* 15 (2): 141-167.

Jenkins, G. D. Jr., A. Mitra, N. Gupta, and J. D. Shaw. 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology* 83 (5), 777-787.

Johnson, D. W., and R. T. Johnson. 1989. *Cooperation and Competition: Theory and Research.* Interaction Book Company.

Johnson, D. W., and R. T. Johnson. 2005. New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs* 131 (4): 285-358.

Jodlbauer, B., and E. Jonas. 2011. Forecasting clients' reactions: How does the perception of strategic behavior influence the acceptance of advice? *International Journal of Forecasting* 27 (1): 121–133.

Jungermann, H., and K. Fischer. 2005. Using expertise and experience for giving and taking advice. In *The Routines of Decision Making*, edited by T. Betsch and S. Haberstroh, 157-173. New York, NY: Psychology Press.

Kadous, K., J. Leiby, and M. E. Peecher. 2013. How do auditors weight informal contrary advice? The joint influence of advisor social bond and advice justifiability. *The Accounting Review* 88 (6): 2061-2087.

Kelly, K. 2010. The effects of incentives on information exchange and decision quality in groups. *Behavioral Research in Accounting* 22 (1): 43-65.

Krug, M. K., and T. S. Braver. 2014. Motivation and Cognitive Control: Going Beyond Monetary Incentives. In *The Psychological Science of Money*, edited by E. Bijleveld and H. Aarts, 137-162. New York NY: Springer.

Logg, J. M., J. A. Minson, and D. A. Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151: 90-103.

Mookherjee, D. 2006. Decentralization, hierarchies, and incentives: A mechanism design perspective. *Journal of Economic Literature* 44 (2): 367-390.

Nyberg, A. J., M. A. Maltarich, D. D. Abdulsalam, S. M. Essman, and O. Cragun. 2018. Collective pay for performance: A cross-disciplinary review and meta-analysis. *Journal of Management* 44 (6): 2433-2472.

Osband, K. 1989. Optimal forecasting incentives. *Journal of Political Economy*, 97(5): 1091-1112.

Poile, C. 2017. Why would I help my coworker? Exploring asymmetric task dependence and the self-determination theory internalization process. *Journal of Experimental Psychology: Applied* 23 (3): 354-368.

Patt, A. G., H. R. Bowles, and D. W. Cash. 2006. Mechanisms for Enhancing the Credibility of an Advisor: Prepayment and Aligned Incentives. *Journal of Behavioral Decision Making* 19 (4): 347-359.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1): 7-63.

Rydval, O., and A. Ortmann. 2004. How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Economics Letters* 85 (3): 315-320.

See, K. E., E. W. Morrison, N. B. Rothman, and J. B. Soll. 2011. The detrimental effects of power on confidence, advice taking, and accuracy. *Organizational Behavior and Human Decision Processes* 116 (2): 272-285.

Scheele, L. M., U. W. Thonemann, and M. Slikker. 2018): Designing incentive schemes for truthful forecast information sharing within a firm. *Management Science* 64 (8), 3690-3713.

Sniezek, J. A., and L. M. Van Swol. 2001. Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Decision Processes* 84 (2): 288-307.

Sniezek, J. A., G. E. Schrah, and R. S. Dalal. 2004. Improving judgment with prepaid expert advice. *Journal of Behavioral Decision* Making 17 (3): 173–190.

Soll, J. B., and R. P. Larrick. 2009. Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35 (3): 780-805.

Super, J. F., P. Li, G. Ishqaidef, and J. P. Guthrie. 2016. Group rewards, group composition and information sharing: A motivated information processing perspective. *Organizational Behavior and Human Decision Processes* 134: 31-44.

Taylor, E.Z. 2006. The effect of incentives on knowledge sharing in computer-mediated communication: An experimental investigation. *Journal of Information Systems*, 20(1): 103−116.

Tost, L. P., F. Gino, and R. Larrick. 2012. Power, Competitiveness, and Advice Taking: Why the Powerful Don't Listen. *Organizational Behavior and Human Decision Processes* 117 (1): 53–65.

Tjosvold, D. 1984. Cooperation theory and organizations. *Human Relations* 37 (9): 743-767.

Tjosvold, D. 1989. Interdependence and power between managers and employees: A study of the leader relationship. *Journal of Management* 15 (1): 49-62.

Tjosvold, D. 1998. Cooperative and competitive goal approach to conflict: Accomplishments and challenges. *Applied Psychology* 47 (3): 285-313.

Van Der Vegt, G., B. Emans, and E. Van De Vliert. 1998. Motivating effects of task and outcome interdependence in work teams. *Group and Organization Management* 23 (2): 124-143.

Van Dijk, F., J. Sonnemans, and F. Van Winden. 2001. Incentive systems in a real effort experiment. *European Economic Review* 45 (2): 187-214.

Van Swol, L. M., and J. A. Sniezek. 2005. Factors affecting the acceptance of expert advice. *British Journal of Social Psychology* 44 (3): 443–461.

Van Swol, L. M., J. E. Paik, and A. Prahl. 2018. Advice recipients: The psychology of advice utilization. In *The Oxford Handbook of Advice*, edited by E. L. MacGeorge and L. M. Van Swol, 21-37. New York, NY: Oxford University Press.

Wageman, R. 1995. Interdependence and group effectiveness. *Administrative Science Quarterly* 40 (1): 145-180.

Wageman, R., and G. Baker. 1997. Incentives and Cooperation: The Joint Effects of Task and Reward Interdependence on Group Performance. *Journal of Organizational Behavior* 18 (2): 139-158.

Weitzman, M. L. 1976. The new Soviet incentive model. *The Bell Journal of Economics* 7 (1): 251-257.

Yaniv, I. 2004a. The benefit of additional opinions. *Current Directions in Psychological Science* 13 (2): 75–78.

Yaniv, I. 2004b. Receiving other people's advice: influence and benefit. *Organizational Behavior and Human Decision Processes* 93 (1): 1–13.

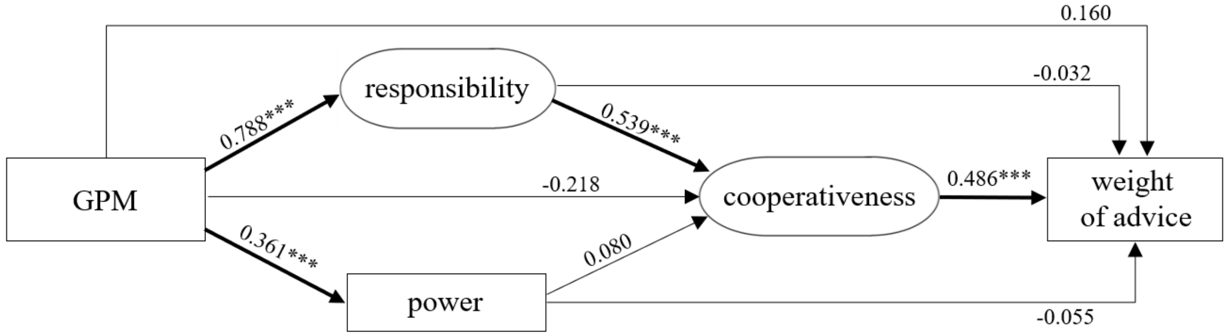
Yaniv, I., and E. Kleinberger. 2000. Advice taking in decision making: egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes* 83 (2): 260–281.

Yaniv, I., and M. Milyavsky. 2007. Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes* 103 (1): 104-120.

Young, S. M. 1985. Participative budgeting: The effects of risk aversion and asymmetric information on budgetary slack. *Journal of Accounting Research* 23 (2): 829-842.

Figures and Tables

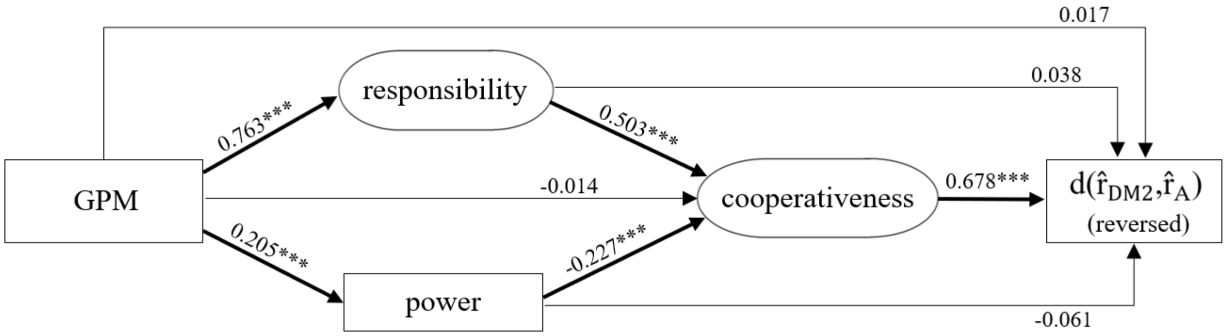
Figure 1. Path analysis for Experiment 1



The Figure shows standardized path coefficients and significance levels. ***, **, * indicate $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively. The unit of observation is the individual decision-maker. Sample size: N = 167.

- GPM is an indicator variable equal to 1 if the respective observation is from the GPM condition, and zero if it is from either the FIXED or the IPM conditions. *FIXED*, *IPM* and *GPM* represent the treatment conditions with advisor's payment being fixed, or depending on the accuracy of his own judgment (*IPM*), or depending on the accuracy of the decision-maker's final judgment (*GPM*), respectively.
- *weight of advice* is the weight the decision-maker assigns to advice.
- *responsibility* and *cooperativeness* are factors derived from the factor analysis in Panel B of Table 2, *power* is a single-item variable.

Figure 2. Path analysis for Experiment 2



The Figure shows standardized path coefficients and significance levels. ***, **, * indicate $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively. The unit of observation is the individual decision-maker. Sample size: N = 190.

– GPM is an indicator variable equal to 1 if the respective observation is from the GPM condition, and zero if it is from either the FIXED or the IPM conditions. *FIXED*, *IPM* and *GPM* represent the treatment conditions with advisor's payment being fixed, or depending on the accuracy of his own judgment (*IPM*), or depending on the accuracy of the decision-maker's final judgment (*GPM*), respectively.
– $d(\hat{r}_{DM2},\hat{r}_A)$ is the Euclidean distance between the decision-maker's final rank prediction and the advice she receives. The measure is reversed (i.e., calculated as $20 - d(\hat{r}_{DM2},\hat{r}_A)$ to reflect the proximity of predictions.
– *responsibility* and *cooperativeness* are factors derived from the factor analysis in Panel B of Table 2, *power* is a single-item variable.

Table 1. Results of Experiment 1

Panel A: Descriptive statistics							
Mean *(standard deviation)*		*FIXED*		*IPM*		*GPM*	
weight of advice (woa)		0.415	(0.156)	0.400	(0.156)	0.488	(0.170)
Estimation Error:	mae($\hat{x}_A$)	163.7	(41.0)	167.1	(44.7)	168.3	(38.4)
	mae($\hat{x}_{DM1}$)	166.6	(59.8)	158.4	(53.6)	155.2	(73.5)
	mae($\hat{x}_{DM2}$)	143.4	(46.8)	139.7	(46.2)	133.8	(53.8)
Groups		52		58		57	

Panel B: Test results			
One sample tests: z-score *(p-value)*	*FIXED*	*IPM*	*GPM*
woa = 0.5	z_{51} = -3.30 ($p < 0.001$)	z_{57} = -4.40 ($p < 0.001$)	z_{56} = -0.85 ($p = 0.398$)
mae($\hat{x}_{DM1}$) = mae($\hat{x}_{DM2}$)	z_{51} = 4.90 ($p < 0.001$)	z_{57} = 4.04 ($p < 0.001$)	z_{56} = 3.65 ($p < 0.001$)
Two sample tests: z-score *(p-value)*	*FIXED* vs. *IPM*	*FIXED* vs. *GPM*	*IPM* vs. *GPM*
woa	z_{108} = 0.61 ($p = 0.545$)	z_{107} = -2.01 ($p = 0.045$)	z_{113} = -2.93 ($p = 0.003$)
mae($\hat{x}_{DM2}$)	z_{108} = 0.46 ($p = 0.645$)	z_{107} = 1.55 ($p = 0.123$)	z_{113} = 1.21 ($p = 0.227$)

The Table shows descriptive statistics for the weights decision-makers give to advice (*woa*) and for mean absolute estimation errors (*mae*) in the three treatments of the experiment in Panel A, and results of Wilcoxon rank sum tests in Panel B. The unit of observation is the individual decision-maker.

– $\hat{x}_{DM1}$ is the decision-maker's initial estimation, $\hat{x}_{DM2}$ her final estimation, and $\hat{x}_A$ is the advisor's estimation which the decision-maker receives as advice.
– *FIXED*, *IPM* and *GPM* represent the treatments with advisor's payment being fixed, or depending on the accuracy of his own estimation $\hat{x}_A$ (*IPM*), or depending on the accuracy of the decision-maker's final estimation $\hat{x}_{DM2}$ (*GPM*), respectively.
– *Weight of advice* is the weight the decision-maker assigns to advice, defined as $woa = (\hat{x}_{DM2} - \hat{x}_{DM1})/(\hat{x}_A - \hat{x}_{DM1})$; the measure is not defined in case the advice is exactly equal to the decision-maker's first estimation ($\hat{x}_A = \hat{x}_{DM1}$). It is equal to 0 when the decision-maker does not adjust her initial estimation ($\hat{x}_{DM2} = \hat{x}_{DM1}$), and equal to 1 when the decision-maker fully adopts the advice ($\hat{x}_{DM2} = \hat{x}_A$). The *woa* data is truncated, i.e., *woa* values are set equal to 0 if the actual value is negative, and equal to 1 if the actual value exceeds 1.
– The e*stimation error mae* is the mean absolute deviation of the estimation from the true value.

Table 2. Post-experimental questionnaire items from Experiment 1

Panel A: Items, item means (standard deviations), and Wilcoxon test results: p-values		*FIXED* [N=52]	*IPM* [N=58]	*GPM* [N=57]	*FIXED* vs. *GPM*	*IPM* vs. *GPM*
Incentives beliefs: Financial incentives lead to ...						
Q_1	people invest more effort in estimation tasks	6.06 (1.38)	5.86 (1.36)	5.98 (1.23)		
Q_2	better performance in estimation tasks	4.54 (1.98)	4.64 (1.84)	4.65 (1.62)		
Q_3	people invest more effort into tasks in general	6.15 (1.06)	5.95 (1.41)	6.12 (1.10)		
Q_4	better task performance in general	4.29 (1.58)	4.41 (1.77)	4.36 (1.68)		
Intention to take advice, relatedness, responsibility, power:						
Q_5	Key to good estimations was good coordination between the advisor and me	4.69 (1.37)	4.55 (1.44)	5.21 (1.40)	0.026	0.005
Q_6	I was sure that I would improve my estimations by adjusting them towards the advice	5.17 (1.41)	5.29 (1.46)	5.74 (1.17)	0.022	0.108
Q_7	My aim was to cooperate with the advisor on making estimations as accurate as possible	5.60 (1.61)	5.83 (1. 19)	6.21 (1.13)	0.040	0.022
Q_8	I felt I should listen to the advisor's advice	4.83 (1.34)	4.64 (1.37)	5.19 (1.26)	0.112	0.046
Q_9	I perceived the advisor to be responsible for my success	4.29 (1.18)	4.43 (1.39)	4.86 (1.27)	<0.001	0.075
Q_{10}	I felt that the advisor and me were more acting in a team than as separate individuals	3.69 (1.55)	3.79 (1.80)	4.77 (1.74)	<0.001	0.003
Q_{11}	I felt responsible for the advisor's success	1.96 (1.58)	2.48 (1.79)	5.54 (1.44)	<0.001	<0.001
Q_{12}	I felt responsible for the advisor	1.63 (1.10)	1.93 (1.39)	5.09 (1.78)	<0.001	<0.001
Q_{13}	I felt responsible for myself only (reversed)	2.12 (1.45)	2.41 (1.71)	5.00 (1.98)	<0.001	<0.001
Q_{14}	Due to the roles assigned, I felt power	2.02 (1.51)	2.17 (1.58)	3.53 (2.11)	<0.001	<0.001

Panel B: Principal component factor analysis of items

	cooperativeness	*responsibility*	*inbe-effort*	*inbe-taskperf*
Q_1			0.90	
Q_2				0.83
Q_3			0.89	
Q_4				0.89
Q_5	0.80			
Q_6	0.80			
Q_7	0.69			
Q_8	0.82			
Q_9	0.65			
Q_{10}	0.59			
Q_{11}		0.92		
Q_{12}		0.92		
Q_{13}		0.85		
Eigenvalue	4.27	2.09	1.74	1.26
Cronbach's alpha	0.839[a]	0.892	0.816	0.742

The Table shows, in Panel A, descriptive statistics for the post-experimental questionnaire items from Experiment 1, and p-values from pairwise Wilcoxon (Mann-Whitney-U) tests contrasting treatment conditions *FIXED* vs. *GPM* and *IPM* vs. *GPM*. Only p-values < 0.20 are displayed. Results of corresponding tests contrasting *FIXED* vs. *IPM* are statistically insignificant with p-values exceeding 0.1 except for Q_{11} ($p = 0.096$).

Panel B of the table reports factor loadings from a principal component factor analysis with orthogonal varimax rotation, eigenvalues of factors, and values for Cronbach's alpha. Only factor loadings > 0.3 are displayed.

All items are measured on a scale from 1 to 7.

[a] Cronbach's alpha increases to 0.841 if Q_{10} is dropped.

Table 3. Results of Experiment 2

Panel A: Descriptive statistics							
Mean (standard deviation)		*FIXED*		*IPM*		*GPM*	
Reliance on advice:	$d(\hat{r}_{DM2}, \hat{r}_A)$	2.71	(2.08)	3.04	(1.55)	1.92	(1.49)
Advice acceptance:	overall	59.5%		49.3%		65.7%	
	1st rank	66.7%		55.2%		67.0%	
	2nd rank	57.1%		47.5%		60.3%	
	3rd rank	52.1%		41.2%		64.6%	
	4th rank	57.0%		50.4%		66.3%	
Accuracies of predictions:	$d(\hat{r}_{DM1},r)$	7.65	(1.87)	7.45	(1.66)	7.58	(1.80)
	$d(\hat{r}_A,r)$	4.75	(1.37)	4.23	(1.13)	3.82	(1.47)
	$d(\hat{r}_{DM2},r)$	5.18	(1.67)	5.10	(1.20)	4.57	(1.67)
Groups		64		63		63	

Panel B: test results				
Two sample tests: z-score (p-value)		FIXED vs. IPM	FIXED vs. GPM	IPM vs. GPM
Reliance on advice: $d(\hat{r}_{DM2}, \hat{r}_A)$		$z_{125} = -1.69$ (p = 0.092)	$z_{125} = 2.02$ (p = 0.044)	$z_{124} = 3.95$ (p < 0.001)
advice acceptance (overall)		$z_{125} = 2.69$ (p = 0.007)	$z_{125} = -1.41$ (p = 0.159)	$z_{124} = -4.17$ (p < 0.001)
Prediction accuracies:	$d(\hat{r}_{DM1},r)$	$z_{125} = 0.83$ (p = 0.408)	$z_{125} = 0.39$ (p = 0.698)	$z_{124} = -0.58$ (p = 0.561)
	$d(\hat{r}_A,r)$	$z_{125} = 2.16$ (p = 0.031)	$z_{125} = 3.62$ (p < 0.001)	$z_{124} = 2.17$ (p = 0.030)
	$d(\hat{r}_{DM2},r)$	$z_{125} = 0.25$ (p = 0.805)	$z_{125} = 2.12$ (p = 0.034)	$z_{124} = 2.42$ (p = 0.015)

The Table shows, in Panel A, descriptive statistics for the distance between decision-makers final predictions and the advice they receive [$d(\hat{r}_{DM2}, \hat{r}_A)$], for the frequencies of advice acceptance, and for the accuracies of predictions. Panel B presents results of Wilcoxon rank sum tests. The unit of observation is the individual decision-maker.

- $\hat{r}_{DM1}$ is the decision-maker's initial rank prediction, $\hat{r}_{DM2}$ her final rank prediction, and $\hat{r}_A$ is the advisor's rank prediction which the decision-maker receives as advice.
- *FIXED*, *IPM* and *GPM* represent the treatments with advisor's payment being fixed, or depending on the accuracy of his own rank predictions $\hat{r}_A$ (*IPM*), or depending on the accuracy of the decision-maker's final rank prediction $\hat{r}_{DM2}$ (*GPM*), respectively.
- $d(\hat{r}_{DM2},\hat{r}_A)$ is the Euclidean distance between the decision-maker's final rank prediction and the advice she receives. The distance is 0 if all ranks are predicted identically and has a maximum of 20.
- The frequency of advice acceptance is the frequency with which a decision-maker receives advice that contradicts her initial rank prediction, and changes her prediction to the advice given.
- $d(\hat{r}_{DM1},r)$, $d(\hat{r}_{DM2},r)$, and $d(\hat{r}_A,r)$ are the Euclidean distances between rank predictions $\hat{r}_{DM1}$, $\hat{r}_{DM2}$, and $\hat{r}_A$, respectively, and actual ranks r. The distance is 0 for a perfect prediction and has a maximum of 20.

Table 4. Post-experimental questionnaire items from Experiment 2

	Descriptive data: means (st. dev.)			Principal component analysis: Factors			
	FIXED [N=64]	*IPM* [N=63]	*GPM* [N=63]	*coopera-tiveness*	*respon-sibility*	*inbe-effort*	*inbe-taskperf*
Incentives beliefs							
Q_1	6.17 (1.39)	6.03 (1.34)	6.37 (0.80)			0.805	
Q_2	5.33 (1.72)	4.95 (1.67)	5.41 (1.35)				0.839
Q_3	6.23 (1.32)	6.05 (0.96)	6.30 (0.78)			0.787	
Q_4	4.77 (1.69)	4.70 (1.67)	4.76 (1.46)				0.853
Intention to take advice, relatedness, responsibility, power:							
Q_5	4.92 (1.42)	4.76 (1.35)	5.49 (1.08)	0.622			
Q_6	5.11 (1.43)	5.38 (1.13)	5.83 (0.99)	0.755			
Q_7	5.73 (1.54)	5.73 (1.10)	6.30 (0.82)	0.652			
Q_8	5.09 (1.41)	4.86 (1.23)	5.63 (1.05)	0.823			
Q_9	5.05 (1.43)	4.52 (1.34)	5.52 (1.22)	0.734			
Q_{10}	3.81 (1.78)	4.03 (1.68)	5.16 (1.45)	0.478	0.475		
Q_{11}	2.14 (1.90)	2.33 (1.66)	5.60 (1.39)		0.866		
Q_{12}	1.67 (1.37)	2.14 (1.63)	4.87 (1.69)		0.882		
Q_{13}	2.38 (1.74)	2.70 (1.57)	4.62 (1.75)		0.740		
Q_{14}	2.48 (1.82)	2.63 (1.85)	3.38 (1.92)	(not included)			
Factor's eigenvalue				4.41	2.04	1.10	1.65
Cronbach's alpha				0.824[a]	0.853[a]	0.709	0.729

The left columns of the table show descriptive statistics for the post-experimental questionnaire items from Experiment 2. Item texts are identical to those in Panel A of Table 2 except for changes referring to the different task (rank prediction instead of estimation).

The last four columns of the table report factor loadings from a principal component factor analysis with orthogonal varimax rotation, eigenvalues of factors, and values for Cronbach's alpha. Only factor loadings > 0.3 are displayed.

All items are measured on a scale from 1 to 7.

[a] After dropping Q_{10}.

Appendix A: Moderation Analyses of incentives beliefs

Experiment 1

H1 implies a moderation as it states that a decision maker's beliefs about how a financial incentive provided to the advisor, and how the performance measure used for the advisor's compensation, affect the decision maker's reliance on advice. In the analyses of the Experiment 1 data we tested both a treatment coding according to pay-performance sensitivity (*pps*) and a coding simply contrasting a performance-dependent pay (*pdp*) being absent versus present. The *pps* coding is *pps* = 0, 1, 2 for the *FIXED*, *GPM*, *IPM* condition, respectively; the *pdp* coding is *pdp* = 0, 1, 1 for *FIXED*, *GPM*, *IPM*, respectively. Panel A of Table A.1 presents descriptive data for decision makers' incentives beliefs - factor *inbe-effort* with items Q_1 & Q_3, and factor *inbe-taskperf* with items Q_2 & Q_4; the factors are measured as unweighted item averages - based on median splits. As the data show, in both the FIXED and the GPM conditions, but not in the IPM condition, participants who more strongly believe that a financial incentive increases both effort and task performance (*high* incentives beliefs group) give more weight to advice. Panel B of the table presents ANOVA results for incentives beliefs about effort and the two alternative treatment codings. The results show that the differences in incentives beliefs observed in Panel A do not moderate the treatment effects, as in neither ANOVA there is a significant interaction effect. We repeated the ANOVA for the second incentives beliefs factor (*inbe-taskperf*) as well as for an all-item average of items Q_1 - Q_4, but did not find any interaction, either (results not tabulated). We further performed regression analyses (results not tabulated), and, again, results show no indication of a moderation effect of incentives beliefs: Irrespective of which proxy for incentives beliefs we use (factor *inbe-effort* or *inbe-taskperf*, or an all-item average), and irrespective of the treatment coding (*pps* or *pdp*), the interaction effect is not significant.

Table A.1. Incentives Beliefs Moderation Analyses for Experiment 1

Panel A: Descriptive statistics based on median splits (low vs. high)					
Mean weight of advice (woa) *(standard deviation) [groups]*		*FIXED*	*IPM*	*GPM*	*Total*
inbe-effort (Q1&Q3):	*low*	0.378 (0.16) [28]	0.381 (0.15) [33]	0.463 (0.18) [30]	0.407 (0.17) [91]
	high	0.458 (0.14) [24]	0.426 (0.16) [25]	0.517 (0.16) [27]	0.468 (0.16) [76]
inbe-taskperf (Q2&Q4):	*low*	0.391 (0.17) [25]	0.416 (0.17) [29]	0.464 (0.17) [29]	0.426 (0.17) [83]
	high	0.437 (0.14) [27]	0.384 (0.14) [29]	0.513 (0.17) [28]	0.444 (0.16) [84]

Panel B: ANOVA results				
	Sum of Squares	df	F-statistic	p-value
Model: inbe-effort and pps				
median split inbe-effort	0.148	1	5.81	0.017
pps	0.242	2	4.75	0.010
median split inbe-effort × pps	0.008	2	0.17	0.847
Residual	4.10	161		
Model: inbe-effort and pdp				
median split inbe-effort	0.126	1	5.01	0.027
pdp	0.235	1	9.32	0.003
median split inbe-effort × pdp	0.001	1	0.03	0.874
Residual	4.109	163		

The Table shows descriptive statistics for the weights decision-makers give to advice (*woa*) based on median splits for their incentives beliefs in Panel A, and ANOVA results in Panel B. The unit of observation is the individual decision-maker.

– *FIXED*, *IPM* and *GPM* represent the treatments with advisor's payment being fixed, or depending on the accuracy of his own estimation (*IPM*), or depending on the accuracy of the decision-maker's final estimation (*GPM*), respectively.
– *Weight of advice* is defined as $woa = (\hat{x}_{DM2} - \hat{x}_{DM1})/(\hat{x}_{A} - \hat{x}_{DM1})$ Data are truncated, i.e., *woa* values are set equal to 0 if the actual value is negative, and equal to 1 if the actual value exceeds 1.
– *inbe-effort* is the average of the two post-experimental questionnaire items Q_1 and Q_3 as defined in Table 2. *inbe-taskperf* is the average of the two post-experimental questionnaire items Q_2 and Q_4 as defined in Table 2.
– *pps and pdp* are treatment codings: *pps* = 0, 1, 2 for the *FIXED*, *GPM*, *IPM* condition, respectively; *pdp* = 0, 1, 1 for *FIXED*, *GPM*, *IPM*, respectively

Experiment 2

We repeated the moderation analyses reported for Experiment 1 (ANOVA based on median splits of incentives beliefs factors, regression analyses) for Experiment 2. Results (not tabulated) are very similar: While we find differences in reliance on advice between decision makers that have stronger versus weaker beliefs in the effectiveness of financial incentives, these differences do not moderate the effect of the treatment manipulation on decision makers' behaviors. In contrast to the results reported in Panel B of Table A.1, we do not even find the main effect of the incentives beliefs median split to be statistically significant.