TI 2008-074/1
Tinbergen Institute Discussion Paper

The Effect of Noise in a Performance Measure on Work Motivation: A Real Effort Laboratory Experiment

Randolph Sloof
C. Mirjam van Praag

Faculty of Economics & Business, University of Amsterdam, and Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

The effect of noise in a performance measure on work motivation: A real effort laboratory experiment

Randolph Sloof and C. Mirjam van Praag*

August 7, 2008

Abstract

This paper reports the results of an individual real effort laboratory experiment where subjects are paid for measured performance. Measured performance equals actual performance plus noise. We compare a stable environment where the noise is small with a volatile environment where the noise is large. Subjects exert significantly more effort in the volatile environment than in the stable environment. This finding is in line with standard agency theory and contrasts a distinct element of expectancy theory; noisier performance measures do not lower work motivation.

1 Introduction

A central problem within organizational economics is how to motivate employees to exert (well-directed) effort. Pay for performance is considered an important, if not key, instrument here. The extensive literature developed in the past decades focuses on the design of optimal pay for performance schemes. One of the main issues this literature addresses is how responsive pay should be to performance, given the actual characteristics of the performance measure(s) used. An important characteristic that has been identified in this regard is noise, besides alignment between the measure and organizational value (cf. Baker, 2002). In general, noisier (or riskier) performance measures reflect employee effort less accurately and optimally receive a lower weight in the employee's compensation scheme. Popularly put, people should not be held accountable for factors they do not control (cf. Roberts, 2004).[1]

*Both authors: Tinbergen Institute and University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. The excellent research assistance of Patrick Stastra and Simon Timmerman is gratefully acknowledged.

[1] As Roberts (2004, p. 137) carefully points out, there is a subtlety in this principle. Good performance measures may make valuable use of variables the employee cannot control, in order to filter out some extraneous randomness (the "informativeness principle"). For example, a company's total shareholder return (TSR) relative to an index (e.g., the S&P

The dictum that noisier performances measures should receive less weight in compensation follows from standard principal-agent theory. The optimal compensation contract strikes a balance between insuring the risk-averse agent against (uncontrollable) risk and providing him with incentives to exert effort. The principal pays a risk premium to the agent that increases with the intensity of incentives, the degree of risk aversion of the agent, and the noise in the performance measure. This premium is traded off against the benefits of additional effort that stronger incentives generate. Noisier performance measures are less attractive for the principal, because they require higher risk premia for a given amount of incentives. More generally, the main disadvantage of noisy performance measures is that they induce an inefficient allocation of risk.

In agency theory, apart from a higher required risk premium, noise *per se* does not have a direct adverse effect on incentives.[2] In the often used linear version of the agency model, noise does not directly affect effort incentives at all; for a given incentive intensity the incentive compatibility constraint is independent of the presence and the type of noise (see Sloof and van Praag (2008) for a full discussion). In more general specifications noise may have a direct impact, but under the standard assumptions typically made about the agent's preferences, it always *strengthens* effort incentives (cf. Section 2).

Expectancy theory as developed by psychologists has a keener eye for the direct detrimental impact noise may have on work motivation (cf. Vroom, 1964). One of the key motivational drivers on which this theory is based is the employee's *effort-performance expectancy.* The stronger the subjective perception that (more) effort leads to (better) performance, the more motivated the employee will be to put in effort. Now, a larger amount of noise in the performance measure implies that the relationship between effort and measured performance is weakened. A natural and intuitive prediction from expectancy theory, therefore, is that (ceteris paribus) a noisier performance measure *reduces* effort incentives. Employees will be less motivated when their evaluation and rewards are based on measures they are less able to control.

In this paper we test the above opposing predictions by means of a controlled laboratory experiment. Subjects are confronted with an individual real effort task, viz. adding three two-digit numbers. They are paid on the basis of their performance, with a given piece rate equal to (slightly more than) 5 eurocents per correct calculation (on top of their base salary). Noise enters the picture because, when calculating a subject's compensation, the number of correct calculations is not registered perfectly. In particular, there is a 50 percent chance that the subject is lucky and an amount of σ correct calculations is added to his actual number of correct calculations. Yet there is also an equal 50% probability

500), is often regarded a less noisy performance measure for CEOs than the absolute value of the company's TSR, because it filters out business cycles effects. Thus, the CEO is not simply rewarded for a booming economy. Here we focus on true randomness that remains after all such filtering opportunities have been exhausted.

[2] Of course, more noise will induce the principal to adapt the incentive intensity due to the larger required risk premium. This paper does not consider the principal's problem and takes the incentive contract as given (cf. Section 2).

that the subject is unlucky and an amount of σ is subtracted. We use σ as a treatment variable to represent different amounts of noise and consider both a *stable environment* in which σ is low ($\sigma = 10$) and a *volatile environment* in which σ is high ($\sigma = 180$).[3] Expectancy theory predicts effort levels to be lower in the volatile environment while according to standard agency theory this will likely be the other way around.

We find that subjects make significantly more (correct) calculations in the volatile environment than in the stable environment. An increase in the amount of noise thus strengthens subjects' motivation to work. This is in line with agency theory and goes against one important ingredient of expectancy theory.

We are somewhat surprised by our own results. The idea that less control over performance leads to less motivation seems quite intuitive. In fact, the present experiment was inspired by an earlier one in which we also found little evidence that effort-performance expectancy plays a major role in work motivation (cf. Sloof and van Praag, 2008). In that experiment we used an arguably much more abstract and complicated effort *allocation* task, in order to introduce a monetary opportunity cost of effort. The observed effort allocations appeared largely independent of changes in the amount of noise, whereas expectancy theory would predict a shift away from the task with the noisier performance measure. Although these earlier results may indeed reflect a true lack of behavioral variance, we could not convincingly exclude that they were due to subjects' lack of understanding of the task. The present experiment corrects this by using a much simpler task and a cleaner design. Overall, the two experiments together cast serious doubt on the idea that more noise in the relationship between effort and performance will demotivate an employee to exert effort.

The paper proceeds as follows. The next section briefly presents the relevant formal predictions derived from standard agency theory and contrasts these with the (informal) predictions derived from expectancy theory. Given the extensive discussion of these two theories in our earlier paper, Sloof and van Praag (2008), this section is kept brief. Section 3 presents the details of our experimental design whereas Section 4 reports the results. The final section discusses our main findings and concludes.

2 Theory

We consider a situation in which measured performance p of the employee depends linearly on effort a and is subject to additive noise:

$$p = a + \varepsilon \tag{1}$$

Here ε is a random noise term with $E[\varepsilon] = 0$, variance σ^2 and distribution function F. The employee is paid on the basis of performance pay, where the

[3] These numbers are based on a pilot experiment in which subjects received a flat wage and were explicitly asked to put in sufficient effort. On average the subjects in the pilot made 176 correct calculations in the 40 minutes they were required to work.

compensation contract is assumed to be linear in performance:[4]

$$\pi(p) = s + b \cdot p \tag{2}$$

with s the fixed base salary and b the piece rate. We set aside the firm's problem of choosing the optimal compensation contract and simply take the above piece rate scheme (i.e. parameters s and b) as given.[5] The issue at hand is how the optimal effort level chosen by the employee varies with the amount of noise in the performance measure. Noisiness is captured in terms of *second order stochastic dominance* (SSD). Loosely put, if for two random payments S and V it holds that $S \succeq_{SSD} V$, S is stochastically less risky and preferred above V by every (strictly) risk-averse individual.[6] In our setup variations in noise follow from considering different distribution functions F. We are particularly interested in comparing a stable environment S with a more volatile environment V for which $F_S \succeq_{SSD} F_V$. In fact, in the experiment we focus on the special case in which the volatile environment differs from the stable environment by means of a (single) mean-preserving spread. This implies that $\sigma_S < \sigma_V$.[7]

The question of interest is in which environment the employee exerts more effort. This in general depends on the agent's preferences over money income and effort, as represented by his utility function $U(\pi, a)$. We assume throughout that the employee likes money and dislikes effort at an increasing rate, i.e. $U_\pi > 0$, $U_a < 0$ and $U_{aa} \leq 0$. Without further assumptions about $U(\pi, a)$ no unambiguous predictions can be made. But, under the assumptions that are typically made within the agency literature, it does follow that optimal

[4]Apart from assuming an additive noise term as in equation 1 and a linear pay scheme (equation 2), the linear agency model also assumes that the employee has CARA risk preferences $U(\pi, a) = -e^{-r(\pi - c(a))}$ and that noise ε is normally distributed (see Gibbons, 2005, for a concise discussion). The analysis below holds for more general utility functions and noise distributions. Key in equation (1) is that noise is additive; we could equally well assume that $p = \theta(a) + \varepsilon$, with $\theta(\cdot)$ an increasing function. By redefining effort as $\widetilde{a} = \theta(a)$, such that $p = \widetilde{a} + \varepsilon$ and $U(w, a) = U(w, \theta^{-1}(\widetilde{a})) = \widetilde{U}(w, \widetilde{a}))$, the analysis in the main text inmediately applies in terms of $\widetilde{a}$ and $\widetilde{U}(w, \widetilde{a}))$ (this is just a non-linear change in units).

[5]In the standard agency model the optimal incentive intensity that the principal chooses in equilibrium decreases with the amount of additive noise (i.e. σ^2). Under alternative assumptions – involving multiplicative noise – a positive correlation between risk and optimal incentive intensity may arise, for instance because the agent has ex ante pre-decision information about the noise that leads him to adapt his optimal effort choice (cf. Baiman et al., 1995; Zabojnik, 1996; Bushman et al., 2000; Prendergast, 2002: Baker and Jorgensen, 2003) or because the risk the agent bears is endogenous and depends on the effort choice itself (Godes, 2004). Because we take the incentive contract as given and solely focus on the agent's effort choice, our experiment does not have much to say about the *equilibrium* tradeoff between risk and incentives (see e.g. Ghosh and John, 2000, for an experiment that directly studies this incentives-insurance tradeoff).

[6]A (strict) risk averter has a utility function that is monotonically increasing and strictly concave in money income. See Chapter 4 in Wolfstetter (1999) for a discussion of stochastic dominance theory.

[7]In general second order stochastic dominance is not equivalent to simply comparing the variances of random variables (with equal means), see Rothschild and Stiglitz (1970) for a full discussion. The mean preserving spread, however, is an example of a SSD transformation (cf. Wolfstetter, 1999, p. 142). Proposition 1 below holds for all SSD transformations and is not restricted to mean preserving spreads.

effort levels are (weakly) higher in the volatile environment than in the stable environment.

Proposition 1 *Let $F_S \succeq_{SSD} F_V$. The employee will exert (weakly) more effort in the volatile environment if either one of the following three conditions holds:*

(i) *The costs of effort can be measured in money terms, such that $U(\pi, a) = U(\pi - c(a))$;*

(ii) *Utility is separable in money and effort, such that $U(\pi, a) = v(\pi) - c(a)$, with $v'' \leq 0$ and $v''' \geq 0$ (besides $v' > 0$, $c' > 0$ and $c'' \geq 0$);*

(iii) *Leisure is a normal good and the employee's level of absolute risk-aversion $A \equiv -\frac{U_{\pi\pi}}{U_\pi}$ is non-negative and non-increasing in effort a.*

The proof is relegated to Appendix A. The first part of the proposition incorporates the frequently used linear agency model. In this case where effort costs can be measured in money terms, the optimal effort level is actually independent of (changes in) the distribution of the noise term. A second often used version of the agency model is where utility is separable in money and effort. In that case there will be a positive relation between effort and noise if the agent is risk averse and his level of absolute risk aversion A does not increase too quickly with income.[8] This certainly holds when the agent has DARA (or CARA) preferences, as is typically assumed. The third set of conditions effectively makes slightly less stringent assumptions about the cross derivatives of U, but a somewhat more restrictive assumption about the employee's risk attitude. In words the latter entails that more labor supply implies a (weakly) lower level of risk aversion. Overall, taking all three cases together it follows that under reasonable assumptions about the employee's preferences, agency theory predicts that the employee will work harder when the performance measure is more noisy.[9]

Expectancy theory as developed by Vroom (1964) assumes that an employee's motivation to work depends on three factors: (i) the subjective probability that (increased) effort leads to (better) performance (*Expectancy*), (ii) the employee's expectation that better performance leads to higher rewards (*Instrumentality*) and (iii) the attractiveness of the rewards to the employee (*Valence*). To emphasize the interactive role these three factors play, expectancy theory is typically presented by means of the following "equation":[10]

$$\text{Motivation } = E \times I \times V \tag{3}$$

[8] With $A \equiv -\frac{U_{\pi\pi}}{U_\pi} = -\frac{v''}{v'}$ it follows that $\frac{\partial A}{\partial \pi} = \frac{-v' \cdot v''' + (v'')^2}{(v')^2} = \frac{-v'''}{v'} + A^2$. Hence $v''' \geq 0 \Longleftrightarrow \frac{\partial A}{\partial \pi} \leq A^2$.

[9] For the general case where no further assumptions about the agent's preferences are made, the effect can go either way. A risk loving agent, for instance, may want to reduce his effort and count more on luck when noise increases. A basic premise of agency theory, however, is that the agent is risk averse. Within the contract theory literature, typically either set of assumptions (i) or (ii) in Proposition 1 is made (cf. Bolton and Dewatripont (2005) and Laffont and Martimort (2002)).

[10] See for example the discussion of expectancy theory in modern textbooks on organizational behavior (e.g. French et al., 2008, pp.172-175) and management (e.g. Boddy, 2008, pp. 505-508 and Hitt et al., 2009, pp. 287-289).

Unlike agency theory, however, expectancy theory is not cast in a rigorous formal analytical model. This may explain why expectancy theory is hardly ever discussed within the economics literature. Yet within the fields of both management and organizational behavior it has received widespread acceptance. For example, in their textbook on the fundamentals of management, Robbins and DeCenzo (2008, p. 275) claim that:

> "The most comprehensive and widely accepted explanation of motivation to date is Victor Vroom's expectancy theory. Although the theory has its critics, most of the research evidence supports it."

Although we are less sanguine about the actual support given the existing empirical evidence,[11] the quote does nicely illustrate the importance attached to expectancy theory within the management literature.

From a conceptual level expectancy theory has much in common with agency theory (cf. Bonner and Sprinkle, 2002). Instrumentality corresponds to incentive intensity b while valence refers to the arguments π (money) and a (effort) in the agent's utility function and the particular form this function takes. The expectancy factor, however, is not fully captured in agency theory and may therefore lead to different predictions than in Proposition 1. To illustrate, suppose the noise term can only take two values, viz. $\varepsilon = \sigma > 0$ and $\varepsilon = -\sigma$, both with equal probabilities. Then if σ is low (*stable environment*), the overall effect of effort on performance is large relative to the effect of noise. The employee will have a sufficiently strong belief that he can improve his performance by exerting more effort. But if σ is high (*volatile environment*), performance is largely dependent on noise and the absolute effect of effort relative to the effect of noise is small. In that case the employee has far less control over the ultimate level of measured performance and therefore his effort-performance expectancy will be lower.[12] This will lower the employee's motivation compared to the case where σ is low. In contrast, agency theory focuses on the (expected) marginal benefits and marginal costs of exerting more effort. Variations in σ may have an impact here, because they affect the income levels at which these marginal benefits and costs are evaluated. This leads to the predictions in Proposition 1. But apart from that, the absolute size of the noise term does not play a direct role.

[11] Van Eerde and Thierry (1996) perform a meta-analysis based on 77 studies that measure correlations between the VIE-factors and various measures of work motivation. In general they find positive correlations for each factor in isolation, but results are mixed insofar some effects are significantly different from zero while others are not. This also holds for the expectancy factor that is of main interest here. Interaction effects between the three factors are often insignificant. As Van Eerde and Thierry (1996) point out, many studies suffer from severe measurement problems and they therefore recommend that experiments are conducted to overcome these. See also Sloof and Van Praag (2008) for a more elaborate discussion of the empirical issues involved.

[12] In the words of Peters (1977, p. 134): "...a low effort-performanc expectancy is one for which increasing amounts of effort are believed to result in either a constant or random level of performance."

The notion that the variance σ^2 and the absolute size of the noise term are closely related can be illustrated more generally. For any distribution $F(\varepsilon)$ with $E[\varepsilon] = 0$ and $var(\varepsilon) = \sigma^2$ it holds that:[13]

$$\sigma^2 = (E(|\varepsilon|))^2 + var(|\varepsilon|)$$

The first term on the r.h.s. can be interpreted as a measure of the *size* of noise (relative to effort) in overall performance p. The larger $E(|\varepsilon|)$ is, the more important luck becomes relative to the agent's own effort. The second term measures the *imprecision* of information about this size. The higher $var(|\varepsilon|)$, the less precise the agent's a priori information about the relative importance of noise for overall performance. Effort-performance expectancy seems to refer especially to the size of the noise term, predicting a negative relationship between size and effort incentives. In the experiment we therefore focus on the case where ε can take two values only, as in the example of the previous paragraph. Imprecision is therefore kept constant across treatments, with $var(|\varepsilon|) = 0$. In the volatile environment subjects can thus be sure that luck plays a major role while in the stable environment they are certain that it plays only a minor role.

Summing up, keeping the compensation contract fixed as in equation (2), standard agency theory predicts that employees work harder in the volatile environment whereas expectancy theory predicts them to work less.

3 Experimental design

In line with other experimental studies that examine effort incentives (cf. Arvey, 1972; Niederle and Vesterlund, 2007; Eriksson et al., 2008), our real effort task is arithmetic and consists of repeatingly adding up three two-digit numbers. In performing this task subjects were not allowed to use a calculator. Each calculation was presented on the computer screen in the following way:[14]

Add the values A, B and C

Value A	68
Value B	33
Value C	71

Your answer:

Press ok to confirm your answer and to proceed to the next calculation

[13] This follows because $var(\varepsilon) = E(\varepsilon^2) - (E(\varepsilon))^2 = E(\varepsilon^2) = E\left(|\varepsilon|^2\right) = var(|\varepsilon|) + (E(|\varepsilon|))^2$. Here the second equality follows from $E(\varepsilon) = 0$.

[14] We used the software package z-Tree as developed by Fischbacher (2007) to program the experiment.

After having submitted an answer, a new calculation problem appeared on the screen immediately. Subjects had 40 minutes to solve as many problems as they preferred. All subjects were confronted with exactly the same sequence of problems. Feedback on the number of (correct) answers was given only at the end of the assignment.

Performance pay was based on the number of (registered) correct calculations. Besides a base salary of 4500 points (and a fee of 5 euros for filling in an ex post questionnaire) subjects received a piece rate of 25 points per registered correct calculation. The conversion rate was 1 euro for 480 points. Each registered correct calculation thus yielded slightly more than 5 eurocents, on top of the base salary of 9.38 euros.

Before subjects started working, they were informed that the registration process of the number of correct calculations was not flawless and that at the end of their working time a given number σ (with $\sigma \in \{10, 180\}$) would either be added to or subtracted from the actual number of correct calculations (with equal probabilities).[15] The draw of the registration error was visualized on the screen by rotating a pointer over a two-colored disk, where each color filled 50% of the disk. The computer stopped the pointer at random. In case the pointer came to a stop in the white area, the registration error was to the subject's advantage and an amount of σ was added to his/her actual number of correct calculations. If the pointed stopped in the red area the registration error was to his/her disadvantage and an amount of σ was subtracted.

We considered two different treatments: the stable environment where the registration error is small and equal to $\sigma = 10$ and the volatile environment where the registration error is large and equal to $\sigma = 180$. We selected these numbers after conducting a pilot experiment with six subjects. These subjects received a flat wage and were explicitly asked to put in sufficient effort. On average they made 176 correct calculations in the 40 minutes available. The registration error in the volatile environment thus either nullifies or doubles the effect of putting in high effort for the average subject. This is not the case in the stable environment, where for reasonable effort levels the effect of noise is negligible. Clearly this also translates into sizable differences in terms of money. In the stable environment the registration error amounts to $\pm$ 52 eurocents whereas in the volatile environment it comes down to $\pm$ 9.38 euros.

The experiment was framed in a labor market context where subjects worked for a fictive company (cf. Appendix B). The registration error was motivated as resulting from the inaccurate performance measurement system the company used. In the instructions subjects were explicitly informed that the only measure used to assess their performance was the number of registered correct answers, as included in their remuneration schedule. Thus, the company had not activated any other incentives, monitoring activities or evaluation procedures. It was explicitly mentioned that, instead of doing the calculations, subjects could either read a book, a magazine, a newspaper, listen to their MP3 player or

[15] In order to keep the norm of honest communication to the subjects we also informed them of their actual number of correct calculations.

do something else they liked. We arranged (three different) newspapers and put these on each subject's desk. When recruiting the subjects by means of e-mail announcements, we also urged them to bring something interesting to read.[16] This message was repeated in the e-mail confirming participation in the experimental session they chose. By facilitating these leisure possibilities we thus explicitly created an opportunity cost to spending time on doing the calculations.

Subjects started with on screen instructions. Apart from explaining the experiment and the task, these instructions also presented two hypothetical examples to illustrate how the actual earnings are calculated. Once finished the instructions, subjects made three (non-paid) practice calculations to get familiar with the task at hand. Then the actual experiment started, in which subjects performed the calculations for 40 minutes (or did something else). After this, subjects were informed about the realized registration error and their registered (and actual) number of correct calculations. Subjects filled in an ex post questionnaire before they were (individually and discreetly) paid.

Overall 94 subjects participated, divided over six sessions (3 for each environment). Due to no shows the number of subjects was slightly unbalanced over the two treatments, with 50 subjects in the stable environment and 44 in the volatile environment. Subjects were recruited from the CREED database by means of e-mail notifications.[17] About 46% of the subjects were students in economics, 53% from other disciplines. On average subjects earned almost 23 euros in about 80 minutes, the duration of a complete session. Earnings varied considerably, with a minimum of 10.80 euros and a maximum of 37.30 euros.

4 Results

Our main measure of productive effort is the number of correct calculations individual subjects perform. Figure 1 below displays the frequency distributions (where the numbers of correct calculations are bunched into intervals of ten) for each of the two environments. As is evident from the figure, actual performance is typically higher for subjects in the volatile environment than for subjects in the stable environment. In the former group the median equals 173.5 whereas in the latter it equals 153.5. (These medians belong to the intervals that correspond to the mode of the two respective distributions.) A similar picture arises when we look at the total number of calculations made, i.e., whether correct or not.

[insert Figure 1]

[16] The exact phrase we used in the announcements was: "IMPORTANT. Because it may occur that you have substantial spare time during the experiment, we urge you TO BRING SOMETHING TO READ to the experiment (like a BOOK or a MAGAZINE), so that you can spend this spare time pleasantly."

[17] CREED refers to the Center for Experimental Economics in political Decision making at the University of Amsterdam.

To test the validity of these outcomes more formally, Table 1 presents the results from both ranksum tests and Kolmogorov-Smirnov tests. On average subjects solve about 18 problems more in the volatile environment and this difference appears significant. This quantity increase does not come at the expense of quality. The number of correct calculations is also significantly higher in the volatile environment whereas the precision rate (defined as the fraction of correct calculations of the overall number of calculations made by the subject) does not vary across environments. Overall these figures suggest a productivity increase of more than 10% when we move from the stable environment to the volatile environment.

Table 1: Total (correct) calculations and precision rate by environment

Environment	σ	n	Total calculations	Total correct calculations	Precision rate
Stable	10	50	169.2	162.0	0.958
Volatile	180	44	187.5	178.4	0.952
$\sigma = 10$ vs. $\sigma = 180$:					
p-value ranksum test			0.024	0.038	0.319
p-value Kol.-Smirnov test			0.024	0.067	0.791

We next verify that the observed differences between environments are not due to subject pool effects. Through the ex post questionnaire we administered exogenous background characteristics like age, gender, type of study, overall monthly income and the number of hours per week spent on a paid job. Following Holt and Laury (2002), we also confronted subjects with ten hypothetical choices between a safe lottery and a risky lottery, leading to an individual measure of risk aversion. A value of zero indicates risk neutrality whereas positive (negative) values indicate risk aversion (risk loving).[18] It must be noted though that this measure of risk aversion is suggestive at best, because lottery choices were only hypothetical and payments were not realized.[19] Finally, three questions were posed to subjectively measure subjects' arithmetic ability and their general motivation to perform calculations. To this end, we included statements to which subjects could react on a 7-point Likert scale, ranging from 'totally disagree' (= 0) to 'totally agree' (= 6). This yields scores on the variables 'Math skills', 'Difficulty', and 'Boring'.[20]

[18] A number of safe choices equal to 4 corresponds with risk neutrality. We therefore subtract this number of 4 from the actual number of safe choices to arrive at our normalized measure of risk aversion (which ranges from -4 to 6).

[19] Holt and Laury (2002) find that using hypothetical choices typically leads to underestimates of the actual level of risk aversion. To illustrate, Sloof et al. (2007) measure risk attitudes in an incentive compatible way and find an average level of risk aversion equal to 1.6, which is is indeed slightly higher as the one we observe here (1.37 on average over the two environments).

[20] The respective statements are: (i) "I am relatively good at mathematics", (ii) "I found the calculations hard" and (iii) "I found the calculations boring". Although the questions are

Table 2 below supports the condition that the two subject pools are comparable in terms of their background characteristics. On all accounts, none of the differences is significant (ranksum tests). The observed differences in outcomes between environments can thus not be attributed to differences in exogenous background characteristics like type of study, gender, income, ability etc. This reinforces our earlier conclusion that (ceteris paribus) higher noise levels lead to more effort.

Table 2: Background characteristics across environments

	Stable	Volatile	$p-$value
Age	23.90	23.61	0.586
Male	48.0%	50.0%	0.847
Economics	46.0%	45.5%	0.958
Job hours (per week)	11.28	9.86	0.635
Monthly income (euros)	682.10	708.98	0.646
Risk aversion (−4 to 6)	1.24	1.52	0.292
Math skills (0 to 6)	3.82	4.02	0.638
Difficulty (0 to 6)	0.88	1.09	0.353
Boring (0 to 6)	0.76	0.86	0.752

In order to verify the influence of background characteristics on outcomes, we performed OLS regressions to explain the number of (correct) calculations, including controls. Table 3 presents the results. The ordinal variables *math skills*, *difficulty* and *boring* are converted to 0/1-dummies, which equal 1 if the corresponding ordinal variable is above the median (and mode) of the distribution and 0 otherwise.[21] Because *math skill* significantly correlates with both studying *economics* and the *difficulty* variable (and these two regressors do not correlate with each other), for each effort outcome a second specification is estimated where *math skill* is left out.

answered ex post and thus influenced by the treatment subjects are in, we consider these variables exogeneous because the sequence of calculations was exactly the same across all subjects (and thus environments).

[21] This implies that *high math* equals one iff *math skill* is above 4, *high difficulty* equals one iff *difficulty* exceeds 1 and *high boring* equals one iff *boring* is above 4. Performing regressions where all possible answer categories are represented by dummies (except a reference catergory) leads to insignificant results due to the high number of regressors (24) relative to the sample size (94).

Table 3: Regressions of effort on background characteristics

	Total calculations		Total correct	
	(1)	(2)	(3)	(4)
Volatile environment	17.51**	17.19**	15.86*	15.53*
	(8.46)	(8.52)	(8.00)	(8.08)
Age	0.60	0.50	0.61	0.50
	(1.18)	(1.19)	(1.11)	(1.12)
Male	1.84	2.12	2.52	2.80
	(8.91)	(8.97)	(8.42)	(8.51)
Economics	12.59	15.76*	12.26	15.57*
	(8.85)	(8.65)	(8.37)	(8.21)
Job hours	−0.42	−0.53	−0.36	−0.47
	(0.55)	(0.55)	(0.52)	(0.52)
Monthly income	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Risk aversion	4.95*	4.90*	4.80*	4.75*
	(2.93)	(2.96)	(2.77)	(2.80)
High math	13.77		14.36	
	(9.24)		(8.74)	
High difficulty	−16.03	−18.05	−19.49*	−21.60**
	(11.50)	(11.50)	(10.87)	(10.90)
High boring	−10.41	−9.35	−11.55	−10.45
	(9.23)	(9.27)	(8.73)	(8.79)
constant	140.31***	145.59***	134.18***	139.68***
	(30.10)	(30.11)	(28.47)	(28.56)
R^2	0.20**	0.18**	0.22**	0.19**
adj. R^2	0.10	0.09	0.12	0.10
n	94	94	94	94

Standard errors in parentheses. ***/**/* significance at 1/5/10% level.

Reassuringly, the result keeps standing that a noisier performance measure is associated with more effort. The other significant coefficients in the table show that more able and more risk averse individuals perform more (correct) calculations. Ability is captured by the regressors *economics*, *high math* and *high difficulty*. Inclusion of all three regressors leads to mostly insignificant coefficients due to multicollinearity (as a result of a high positive correlation between being an economics student and scoring high on the subjective *math skills* measure). Subjects studying economics typically score high on the subjective math skills measure. Specifications (2) and (4) that omit the dummy *high math* show that economics students perform more (correct) calculations on average than other students. Subjects who found the calculations easy also performed better than those who found them relatively difficult. In line with Arvey (1972), ability has a sizable impact on arithmetic performance.

Subjects who are more risk averse also perform more (correct) calculations. Qualitatively this is in line with the predictions in Proposition 1. Under risk neutrality, agency theory predicts that effort levels are independent of the amount of (additive) noise. Risk averse agents typically work harder when noise increases. A given amount of noise is felt stronger by a more risk averse individual. It thus seems reasonable to expect that subjects who are more risk averse will exert more effort, in line with what we observe.[22]

We also included a number of behavioral questions in the questionnaire that shed some light on the motivational forces behind our main finding. Here we again used statements to which subjects could react on a 7-point Likert scale, as before ranging from 'totally disagree' (= 0) to 'totally agree' (= 6). Table 4 below lists these statements and the average scores within each environment. It also reports the $p-$values from ranksum tests comparing the two environments. On all but the first question subjects in the volatile environment score significantly higher.

The first two questions relate to the motivational impact of the salary structure. Subjects in the stable environment are motivated more by the given level of the piece rate than subjects in the volatile environment are. The latter attach a higher importance to the given level of the base salary. The next three statements concern the impact of noise. Subjects in the volatile environment are influenced stronger by the presence of noise. Here two opposing motivational forces appear at work. More noise not only leads to a stronger stimulus to work, but also has a demotivating impact. Together with the answers to the first two questions this suggests that, on the one hand, more noise weakens incentives because the impact of (additional) effort on compensation becomes smaller relative to the impact of noise. On the other hand, noise strengthens incentives because subjects are more motivated to attain a certain (minimum) level of income. The latter motivational force is in line with the answers to the last two behavioral questions that relate to loss aversion. Especially in the volatile environment subjects are concerned with minimizing the potential loss from a disadvantageous registration error.

Overall, the subjective behavioral questions seem to indicate that elements from both (standard) agency theory and expectancy theory play a role in work motivation. Our objective data on the amount of (correct) calculations made reveals that agency theory wins out. Subjects are significantly more productive in the volatile environment than in the stable environment.

[22] It must be noted that formally the comparative statics in risk aversion are ambiguous. To illustrate, let $U(\pi, a) = v(\pi) - c(a)$ as in part (ii) of Proposition 1. The first order condition for the optimal amount of effort a^* then equals:

$$v'(s + ba^* - b\sigma) + v'(s + ba^* + b\sigma) = \frac{2c'(a^*)}{b}$$

Higher levels of risk aversion imply that $v(\pi)$ is more curved, effectively increasing $v'(s + ba^* - b\sigma)$ and decreasing $v'(s + ba^* + b\sigma)$ (where we keep $v(s + ba^*)$ and $v'(+ba^*)$ constant). The former strengthens effort incentives while the latter leads to weakened incentives. More risk aversion may imply any combination of these two effects, so the impact on effort can in general go either way. The qualitative reasoning in the main text implicitly assumes that the first effect dominates and (like our measure of risk aversion itself) is suggestive at best.

Table 4: Behavioural questions across treatments

Motivational driver	Stable	Volatile	$p-$value
Incentive piece rate The variable salary of 25 points (0.052 euro) per correct answer motivated me to perform the calculations	4.50	3.59	0.01
Importance base salary I found the fixed salary of 4500 points (9.4 euro) important, because of the magnitude of the registration error	2.68	4.43	0.00
Influence of noise The amount of effort I exerted to perform the calculations was influenced by the magnitude of the registration error	1.38	2.84	0.00
Stimulus of noise The magnitude of the registration error stimulated me to work harder	1.66	3.07	0.00
Demotivation of noise I was demotivated to work, because of the magnitude of the registration error	0.76	1.95	0.00
Loss aversion I At the start of the experiment, my goal was to minimize the possible loss due to the registration error	2.24	4.07	0.00
Loss Aversion II Because of the possibility that [10 or 180] calculations would be subtracted, I worked hard to make sure that I would at least earn a certain minimum amount	2.70	4.68	0.00

5 Discussion

Our research interest lies in the effect of noise in a performance measure on work motivation. In the introduction we directly pitted agency theory against expectancy theory, because (under the assumptions on preferences commonly made) these theories lead to opposing predictions. From that perspective our experiment seems useful, because a priori it seems far from obvious what the outcome of the experiment would be. Our experiment should not be taken, however, as a direct test of either one of these encompassing theories per se. For instance, if we had found the opposite result that effort levels were significantly lower in the volatile environment, we would not have rejected agency theory. This would merely have indicated that the standard assumptions typically made within this theory do not capture an (arguably important) empirical phenomenon and therefore should be refined.

Similarly, our results neither indicate that effort-performance expectancy plays no role at all in work motivation. Clearly, one important element of effort-performance expectancy is how responsive performance is to effort. Suppose that performance is given by $p = \theta \cdot a + \varepsilon$, then the marginal productivity of effort $\frac{\partial p}{\partial a} = \theta$ will undoubtedly affect the employee's work motivation. The higher θ, the more effort the agent is likely to put in. Note that this part of expectancy theory is already fully incorporated in agency theory. Our interpretation of the expectancy factor goes beyond this, however, as we conjectured that the effect of effort on performance relative to the effect of noise plays an important role as well. Although intuitively the latter idea seems rather compelling, our results suggest otherwise. This may indicate that the expectancy factor is already fully captured by the marginal productivity of effort in generating performance.[23]

Finally, although our results are in line with standard agency theory, alternative theories that are not rooted in expected utility theory may explain them equally well. A prime candidate here is prospect theory as developed by Kahneman and Tversky (1979). As the answers to the behavioral questions suggest, subjects in the volatile environment were partly motivated to minimize the potential loss of a negative registration error by performing as many calculations as possible. They were also concerned with earning a certain minimum amount. This suggests that subjects have a reference income (aspiration level) in mind and experience earning below this reference as a loss (and more so than they consider earning above the reference income as a gain). Both reference-dependent

[23] The consideration of the marginal productivity of effort suggests an alternative operationalization of Vroom's expectancy construct, viz. to make θ stochastic such that the model contains multiplicative noise. It is easy to show that under the assumptions of the linear agency model, the agent's optimal effort level is then decreasing in $var(\theta)$, see Baker and Jorgensen (2003). Note that for this specification of the agency model (multiplicative) noise has a direct detrimental impact on work motivation, because the risk averse agent reduces his effort level in order to reduce the level of risk he faces. In an early field experiment Neider (1980) finds that employee participation in decisions about how they should perform their job clarifies the link between effort and performance and thereby boosts both effort levels and sales (see Smith et al., 2000, for a similar finding on participative decision making). In terms of the alternative specification considered here, worker participation lowers $var(\theta)$.

utility and loss aversion are important ingredients of prospect theory.

A number of recent empirical studies already suggested the importance of both a reference income level and loss aversion for labor supply. Camerer et al. (1997), for instance, find that inexperienced cabdrivers in New York keep working until they have reached a (loosely set) target income for that day. On high demand days they reach this target level quite quickly and therefore stop working early, yet on low demand days they work more hours to achieve their reference income. If these workers would substitute work hours from low demand to high demand days, they could increase their income up to around 10 percent. Chou (2002) similarly finds a significantly negative wage elasticity for the labor supply of Singapore taxi drivers. These conclusions have been challenged by Farber (2005, 2008), who employs different econometric methods and concludes that daily income targets play only a minor role in cabdrivers' labor supply. One of the general problems here is the non-experimental nature of the data, which makes it hard to separate the effect of an increase in wages due to demand increases from other effects like common supply shocks and selection effects that arise from the choice whether to work at all on a given day (see Goette et al., 2004, for discussion).

This problem is circumvented by the field experimental study of Fehr and Goette (2007). They study the effect of a temporary (i.e. month-long) 25% increase in the commission rate of bicycle messengers in Zurich. This wage increase appears to have two effects; bicycle messengers increase their overall labor supply by working more shifts during the treatment period, but at the same time decrease their effort level per shift (where effort is measured by revenue generated). They argue that these findings are best explained by employees having reference dependent, loss averse preferences. Our findings can also be loosely re-interpreted within this context. For a given fixed level of the piece rate, a smaller amount of noise effectively implies a higher piece rate in relative terms. This makes it easier in the stable environment to reach a particular earnings target, such that subjects put in less effort on performing the calculations.[24] Given that our experiment concerns one working period only, it does not yield results regarding the intertemporal substitution of labor supply. Our experiment therefore cannot differentiate between the standard neoclassical model of labor supply based on expected utility and alternative models assuming reference dependent preferences.

Although expectancy theory is a within-persons theory,[25] in this paper we opted for a between subjects design. In our earlier paper Sloof and Van Praag (2008) we followed Van Eerde and Thierry (1996)'s suggestion and used a within-

[24] Note that subjects signed up for our laboratory experiment without knowing what the experiment was about. Hence by design their participation decision (overall labor supply) could not be affected by the environment they were confronted with. In the ex post questionnaire we also included a question asking subjects about how much time they approximately spent on making the calculations. The answers did not differ significantly between the two environments (on average 37.16 in the stable and 36.89 in the volatile environment), providing weak evidence that overall labor supply was not lower in the stable environment.

[25] The motivational force to perform a task or to exert effort as given in equation (3) should be viewed as one that acts relative to other forces within the individual.

subjects setup. In that experiment subjects allocated their effort over a stable environment and a volatile environment they were simultaneously confronted with. Afterwards we felt that this may have contributed to the insignificant results we obtained. It is not unreasonable to conjecture that, realizing that they are paid on the basis of the different noisy environments anyway, subjects did not feel the need to differentiate their behavior over the different environments. They may have simplified the situation by focusing on the aggregate noise of the two environments together. A setup in which subjects go through the different environments sequentially may avoid this, but at the same time introduces other disadvantages as well (like e.g. confounding order, learning and concentration effects and a heightened sensitivity to different environments). The main disadvantage of between-subjects comparisons is that statistical variation across subjects may lead to spurious treatment effects. But with random assignment and a sufficient number of participants this disadvantage largely disappears. Our comparison of background characteristics across environments reveals that observed treatment differences cannot be attributed to subject pool effects. We are therefore confident that our results are not due to a supposedly inappropriate experimental design.

.

References

[1] Arvey, R.D., 1972, Task performance as a function of perceived effort-performance and performance-reward contingencies, *Organizational Behavior and Human Performance* 8, 423-433.

[2] Baiman, S., Larcker, D. and M. Rajan, 1995, Organizational design for business units, *Journal of Accounting Research* 33, 205-229.

[3] Baker, G., 2002, Distortion and risk in optimal incentive contracts, *Journal of Human Resources* 37, 728-751.

[4] Baker, G. and B. Jorgensen, 2003, Volatility, noise and incentives, Harvard Business School working paper.

[5] Boddy, D., 2008, *Management: An introduction (4^{th} ed.)*, FT Prentice Hall, Harlow.

[6] Bolton, P. and M. Dewatripont, 2005, *Contract theory*, The MIT Press, Cambridge.

[7] 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, 303-345.

[8] Bushman, R., Indjejikian and M. Penno, 2000, Private predecision information, performance measure congruity and the value of delegation, *Contemporary Accounting Research* 17, 561-587.

[9] Camerer, C., Babcock, L., Loewenstein, G. and R. Thaler, 1997, Labor supply of New York City cabdrivers : One day at the time, *Quarterly Journal of Economics* 112, 407-441.

[10] Chou, Y.K., 2002, Testing alternative models of labour supply: Evidence from taxi drivers in Singapore, *Singapore Economic Review* 47-17-47.

[11] Eriksson, T., Poulsen, A. and M.C. Villeval, 2008, Feedback and incentives: Experimental evidence, Working Paper.

[12] Farber, H.S., 2005, Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of Political Economy* 113, 46-82.

[13] Farber, H.S., 2008, Reference-dependent preferences and labor supply: The case of New York City taxi drivers. *American Economic Review* 98, 1069-1082.

[14] Fehr, E. and L. Goette, 2007, Do workers work more if wages are high? Evidence from a randomized field experiment, *American Economic Review* 97, 298-317.

[15] Fischbacher, U., 2007, z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* 10, 171-178.

[16] French, R., Rayner, Ch., Rees, G. and S. Rumbles, 2008, *Organizational behavior*, John Wiley and Sons Ltd, England.

[17] Ghosh, M. and G. John, 2000, Experimental evidence for agency models of salesforce compensation, *Marketing Science* 19, 348-365.

[18] Gibbons, R., 2005, Incentives between firms (and within), *Management Science* 51, 2-17.

[19] Godes, D., 2004, Contracting under endogenous risk, *Quantitative Marketing and Economics* 2, 321-345.

[20] Goette, L., Huffman, D. and E. Fehr, 2004, Loss aversion and labor supply, *Journal of the European Economic Association* 2, 216-228.

[21] Hitt, M.A., Black, J.S. and L.W. Porter, 2009, *Management (2^{nd} ed.)*, Pearson Prentice Hall, New Jersey.

[22] Holt, C.A. and S.K. Laury, 2002, Risk aversion and incentive effects, *American Economic Review* 92, 1644-1655.

[23] Kahneman, D. and A. Tversky, 1979, Prospect theory: An analysis of decisions under risk, *Econometrica* 47, 262-291.

[24] Laffont, J.-J. and D. Martimort, 2002, *The theory of incentives: The principal-agent model*, Princeton University Press, Princeton.

[25] Neider, L.L., 1980, An experimental field investigation utilizing an expectancy theory view of participation, *Organizational Behavior and Human Performance* 26, 425-442.

[26] Niederle, M. and L. Vesterlund, 2007, Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics* 122, 1067-1101.

[27] Peters, L.H., 1977, Cognitive models of motivation, expectancy theory and effort: An analysis and empirical test, *Organizational Behavior and Human Performance* 20, 129-148.

[28] Prendergast, C., 2002, The tenuous trade-off between risk and incentives, *Journal of Political Economy* 110, 1071-1102.

[29] Roberts, J., 2004, *The modern firm: Organizational design for performance and growth*, Oxford University Press, Oxford.

[30] Robbins, S.P. and D.A. DeCenzo, 2008, *Fundamentals of management: Essential concepts and applications*, Pearson Prentice Hall, New Jersey.

[31] Rothschild, M. and J.E. Stiglitz, 1970, Increasing risk: I. A definition, *Journal of Economic Theory* 2, 225-243.

[32] Sloof, R. and C.M. van Praag, 2008, Performance measurement, expectancy and agency theory: An experimental study, *Journal of Economic Behavior and Organization*, forthcoming.

[33] Sloof, R., Oosterbeek, H. and J. Sonnemans, 2007, On the importance of default breach remedies, *Journal of Institutional and Theoretical Economics* 163, 5-22.

[34] Smith, K., Jones, E. and E. Blair, 2000, Managing salesperson motivation in a territory realignment, *Journal of Personal Selling & Sales Management* 20, 215-226.

[35] Van Eerde, W. and H. Thierry, 1996, Vroom's expectancy models and work-related criteria: A meta-analysis, *Journal of Applied Psychology* 81, 575-586.

[36] Vroom, V., 1964, *Work and motivation*, Wiley, New York.

[37] Wolfstetter, E., 1999, *Topics in microeconomics*, Cambridge University Press, Cambridge.

[38] Zabojnik, J., 1996, Pay-performance sensitivity and production uncertainty, *Economics Letters* 53, 291-296.

Appendix A: Proof of Proposition 1

The proof is based on Wolfstetter (1999). Let $X = b \cdot \varepsilon$. Then monetary income in (2) can be directly written in terms of effort and noise as $\pi(a, X) = s + b \cdot a + X$. The employee's objective function becomes $V(a, X) \equiv U(\pi(a, X), a)$. Proposition 4.11 in Wolfstetter (1999, p. 146) states that if $V_{ax}(a^*, X) \leq 0$ and $V_{axx}(a^*, X) \geq 0$ everywhere, then the optimal effort level a^* weakly decreases after a SSD transformation of X (see also his Proposition 5.5). We show that this applies in each of the three cases considered. In our setup $V_{ax} = U_{a\pi} + b \cdot U_{\pi\pi}$ and $V_{axx} = U_{a\pi\pi} + b \cdot U_{\pi\pi\pi}$. Now, in case (i) $V_{ax} = (b - c'(a))U''$ and $V_{axx} = (b - c'(a))U'''$. By the first order condition $b = c'(a)$ for a^* in this case, we obtain $V_{ax} = V_{axx} = 0$. For case (ii) we get $V_{ax} = v'' \cdot b$ and $V_{axx} = v''' \cdot b$ and the claim follows immediately from the assumptions $v'' \leq 0$ and $v''' \geq 0$.

Finally, in regard to case (iii) note that $V_{ax} \leq 0$ is equivalent to leisure being a normal good. From the proof of Lemma 5.1 in Wolfstetter (1999, p. 153) it follows that $sign\{V_{axx} + A \cdot V_{ax}\} = -sign\left\{\frac{\partial A}{\partial a}\right\}$. With $V_{ax} \leq 0$ and $A \geq 0$ it follows that $V_{axx} \geq 0$ if $\frac{\partial A}{\partial a} \leq 0$. ■

Appendix B: experimental instructions (stable environment)

Explanation experiment

Thank you for participating in this experiment. You will be taking part in a study of the labor market. This experiment does not involve interaction with other participants. This means that your earnings only depend on your own results. During the experiment your earnings will be calculated in points. At the end of the experiment these points will be converted into euros at a rate of 480 points = 1 euro. On top of that you will receive 5 euros for filling in an exit questionnaire.

In this experiment you will work for a fictive company. Your job consists of (repeatedly) adding three two-digit numbers and to report the results. The length of your working day is 40 minutes and during your work you will not be monitored by your company. This means that you can choose your own work pace, so you can either work fast, slow or not at all. Thus, instead of working you may also read a newspaper (which you can find on your desk), a book, a magazine, listen to your MP3 player, or do something else you like. However, you are not allowed to use a calculator nor your mobile phone and the computer should only be used for the experiment itself. This means that you are not allowed to exit the experimental program, to use the internet etc.. Furthermore, you are not allowed to communicate with any other participant.

The compensation the company pays you for your work is partly based on your performance. Besides a fixed salary of 4500 points, you will earn performance pay based on the number of (registered) correct calculations. Your variable earnings are 25 points per correct calculation.

At the end of your working day the company assesses your productivity to determine your overall compensation. That is, it verifies the number of correct calculations you made. Unfortunately, the registration process the company uses to measure your productivity is not without error. In particular, the registration process will either add or subtract 10 correct calculations to/from the actual number of correct calculations you made. Both possibilities are equally likely. With 25 points per registered correct calculation, this implies that due to the registration error an amount of 250 is either added or substracted to/from your earnings.

In the experiment the error in the registration process is represented by the computer drawing a random number. This is visualized to you after your working day is over. You will then learn whether 10 correct calculations will be added to your total or whether 10 correct calculations will be subtracted. For now it is important to keep in mind that both outcomes have a probability of 50%.

If you have any question now or during the experiment, please raise your hand and one of the experimenters will come and help you.

Work conditions:

Your fixed salary (points)	4500
Your variable salary per registered correct calculation (points)	25
The common error (addition or subtraction of correct calculations) in the registration process	10
The common error (addition or subtraction of points) in your earnings	250
Length of working day (minutes)	40

If you have read the instuctions you can press Continue to **proceed to two hypothetical examples.**

Continue

Figure 1: Frequency distributions of the number of correct calculations