UvA-DARE (Digital Academic Repository)

Performance measurement, expectancy and agency theory: An experimental study

Sloof, R.; van Praag, C.M.

Link to publication

Citation for published version (APA):
Sloof, R., & van Praag, C. M. (2005). *Performance measurement, expectancy and agency theory: An experimental study*. (Scholar Working Papers Series; No. WP 52/05). University of Amsterdam/Tinbergen Institute.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Download date: 26 Jan 2021

TI 2005-026/1
Tinbergen Institute Discussion Paper

Performance Measurement, Expectancy and Agency Theory

Randolph Sloof
Mirjam van Praag

Faculty of Economics and Econometrics, Universiteit van 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

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Performance measurement, expectancy and agency theory: An experimental study*

Randolph Sloof and Mirjam van Praag
University of Amsterdam and Tinbergen Institute

February 25, 2005

Abstract

Theoretical analyses of (optimal) performance measures are typically performed within the realm of the linear agency model. An important implication of this model is that, for a given compensation scheme, the agent's optimal effort choice is unrelated to the amount of noise in the performance measure. In contrast, expectancy theory as developed by psychologists predicts that effort levels are increasing in the signal-to-noise ratio. We conduct a real effort laboratory experiment to assess the relevance of this prediction in a setting where all key assumptions of the linear agency model are met. Moreover, our experimental design allows us to control expectancy exactly as in Vroom's (1964) original expectancy model. In this setting, we find that effort levels are invariant to changes in the distribution of the noise term, i.e. to expectancy. Our results thus confirm standard agency theory and reject this particular aspect of expectancy theory.

1 Introduction

According to standard principal-agent theory the design of a compensation scheme is guided by a trade-off between risk and incentives. The optimal compensation contract must strike a balance between insuring the risk-averse agent against (uncontrollable) risk and providing him with sufficient incentives to exert effort. Within this theory an important characteristic of the performance measure on which compensation is based is the so-called signal-to-noise ratio. This ratio reflects the marginal effect of effort on measured performance relative to the amount of uncontrollable noise. Theory predicts that, the higher the signal-to-noise ratio, the stronger the optimal responsiveness of compensation to performance, i.e. the more high powered incentives will be.

*First version: July 2004. This paper benefitted from the helpful comments by Robert Dur, Sandra Maximiano, Hessel Oosterbeek, Arno Riedl, Dirk Sliwka and Joep Sonnemans.

The standard agency model has been critically re-evaluated over the past decade (see Gibbons 2005 for an overview). It has been argued that the riskiness of a performance measure is not a sufficient criterion for the selection of this measure into a remuneration contract; in reality many measures that are essentially riskless go unused (cf. Prendergast 2002). Apart from the signal-to-noise ratio, at least two other aspects play an important role in the selection of performance measures. The first one concerns the phenomenon of *distortion* and was first discussed in Steven Kerr's seminal publication "On the folly of rewarding A, while hoping for B" (Kerr 1975). It has been formally incorporated into agency theory by Holmstrom and Milgrom (1991), Baker (1992, 2002), and Feltham and Xie (1994). Distortion becomes relevant whenever the objective of the principal is not contractible.[1] The alternative performance measure that is then used, is distorted whenever the marginal effect of an agent's action on the performance measure is not aligned with the marginal effect on the objective of the principal. A distorted performance measure thus gives incentives to agents to pursue suboptimal tasks or a suboptimal combination of tasks. For example, managers who are remunerated based on the returns to their investments (ROI) not only get incentives to increase returns, but also to decrease their levels of investment. In many cases the latter behavior will not be aligned with the company's objective. In general it holds that compensation schemes should put a lower weight on more distorted performance measures (cf. Baker 2002).

The second important aspect concerns the amount of pre-decision information an agent has before acting.[2] Baker and Jorgensen (2003) make a useful distinction between noise and volatility. Noise is a type of uncertainty to which an agent should not react by changing his actions, while volatility is a type of uncertainty to which an agent should adapt his optimal action choice. They show that, under certain circumstances, noise reduces the optimal incentive intensity whereas volatility increases the optimal incentive intensity.

The objective of this paper is to introduce and test the relevance of a third aspect of performance measures which is not considered in the standard agency model, viz. *expectancy*. We introduce this aspect by means of the following example. Consider a sales representative who is responsible for selling a company's product in a particular region. Her overall compensation w equals the sum of a fixed salary s and a given fraction b of overall sales realized. The state of the economy in the region is quite stable, such that the demand for the company's product is not very noisy and mainly dependent on the sales rep's selling effort. For concreteness, let sales equal $y = a \pm 50$, where a reflects the non-contractible effort level of the sales rep (measured in e.g. hours per year). Here the noise term equals either $\varepsilon = -50$ or $\varepsilon = 50$, with equal probabilities. With compensation scheme $w = s + by$ the expected marginal (\$) benefit of effort is b. The sales representative is expected to choose effort such that $c'(a) = b$, where $c(a)$

[1] This lack of contractibility can be due to its non-verifiability or, as is often the case, due to the inefficiently high risk premium that it would require to meet the agent's participation constraint.

[2] This is studied within the context of agency theory in e.g. Baiman et al. (1995), Bushman et al. (2000) and Baker and Jorgensen (2003).

denotes the costs of effort measured in money terms. Note that this *incentive compatibility constraint* is independent of the distribution of the noise term ε.[3] The effort level chosen is thus invariant to changes in the signal-to-noise ratio.

For some reason, a year later the sales rep is assigned to a different region (or a different product line) in which demand is much more noisy. In particular, overall sales now equal $y = a \pm 1000$. To account for the higher risk involved, the new compensation scheme $w = z + by$ pays a higher fixed wage $(z > s)$. The commission rate b is kept the same though.[4] An interesting question now is whether the sales rep will exert the same level of effort as last year. According to standard agency theory she would; the effort level actually chosen is determined by exactly the same incentive constraint $c'(a) = b$. However, because the signal-to-noise ratio in the 2^{nd} year's sales is much smaller, intuitively it seems likely that the sales rep will put in less effort in the second year. Here the idea is that she will not only take the marginal effect of effort on sales into account, but also its absolute effect as compared to the effect of noise (i.e. demand fluctuations).

The above intuition is an important element of expectancy theory, a psychological theory of motivation first introduced by Vroom (1964).

> "The specific outcomes attained by a person are dependent not only on the choices that he makes but also on events that are beyond his control. For example, a person who elects to buy a ticket in a lottery is not certain of winning the desired prize. Whether he does so is a function of many chance events. Similarly, the student who enrolls in medical school is seldom certain that he will successfully complete the program. ... Most decision-making situations involve some element of risk, and theories of choice behavior must come to grips with the role of these risks in determining the choices that people do make. ...Whenever an individual chooses between alternatives that involve uncertain outcomes, it seems clear that his behavior is affected not only by his preferences among these outcomes but also by the degree to which he believes these outcomes to be probable. Psychologists have referred to these beliefs as expectancies... An expectancy is defined as a momentary belief concerning the likelihood that a particular act will be followed by a particular outcome. .. Expectancy is an action-outcome association." (Vroom, 1964, p. 20)

Expectancy theory thus emphasizes the importance of the *perceived* relationship between effort and good outcomes (i.e. performance) for effort incentives. In particular, it predicts that the incentive to exert effort will be stronger, the stronger this perceived relationship between effort and performance. Within our earlier example this implies that, for a given level of incentive intensity b in

[3] This holds because the example given satisfies the assumptions of the so-called *linear* agency model. In Section 2 this will be elaborated upon.

[4] Here we assume that the contract satisfies the participation constraint, i.e. the employee will not be tempted to terminate the contract. In other words, sorting based on risk attitude is assumed to play no role.

the compensation contract, a performance measure with a higher signal-to-noise ratio will give stronger incentives to exert effort. The underlying idea is that agents will be demotivated to exert effort whenever the size of the (marginal) effect of their effort on performance is small relative to the size of the effect of noise. For example, when sales equal $y = a \pm 1000$ a "good" outcome (i.e. high sales) can only occur when the noise term equals $\varepsilon = 1000$ and is thus largely independent of effort level a. This will demotivate the agent to exert effort in the first place.

Although expectancy theory is only rarely discussed within the economics literature, it has received widespread acceptance amongst psychologists:

> "Vroom's writings (1964) have had a substantial impact on the field of organizational psychology and expectancy theory remains as one of the two or three most heavily researched theories of motivation." (Mitchell 1982)

Furthermore, expectancy theory has been subjected to ample empirical testing. As the meta-analysis of van Eerde and Thierry (1996, p. 581) reveals, the particular relationship between expectancy and effort has received considerable empirical support. However, as van Eerde and Thierry argue, many of the empirical studies suffer from severe measurement problems. They therefore recommend that experiments are undertaken to overcome the measurement problems identified.

In this paper we follow this suggestion and test the relevance of effort-performance expectancy for incentive contracts by means of a controlled laboratory experiment. Our design is inspired by the real effort experiment of van Dijk et al. (2001) and closely follows the example of the sales representative given above. Just as in the example we vary the distribution of the noise component while keeping everything else (e.g. incentive intensity b) constant. In this way our design allows us to control expectancy in a specific manner such that it resembles Vroom's original idea much more closely than most tests of expectancy theory so far (cf. van Eerde and Thierry 1996).

In contrast to the evidence gathered in the psychological literature, we find that effort-performance expectancy is not an important determinant of the level of effort exerted by subjects. Their effort choices appear largely independent of the noise in the performance measure. We thus obtain no evidence that the standard (linear) agency model should be adapted to incorporate expectancy in order to better reflect real life behavior of agents.

This paper proceeds as follows. The next section presents the linear agency model and formally derives that in this model the incentive constraint is independent of the signal-to-noise ratio. Section 3 discusses expectancy theory in more depth and relates this theory to the agency model. In particular, it shows why the latter model is insufficient if expectancy would be relevant. This section also discusses some measurement problems that have hampered earlier empirical tests of expectancy theory. Section 4 deals with our experimental design. Section 5 presents the results and Section 6 concludes.

2 The linear principal-agent model

The principal-agent model concerns a situation in which an agent takes an action a to (stochastically) increase output y. The agent's action itself is non-verifiable, only output can be contracted upon. This output is initially owned by the principal, but she might share it with the agent by paying him a wage contingent on output $w(y)$. The agent's utility over wage and effort is given by $U(w, a)$, with $\frac{\partial U}{\partial w} > 0$ and $\frac{\partial U}{\partial a} < 0$. The agent is thus action-averse. Output is assumed to depend stochastically on a, such that ex ante it is uncertain how much the agent will produce. The timing of events is as follows:

1. The principal offers a compensation contract $w(y)$ to the agent;
2. The agent either accepts or rejects the compensation contract. Rejection yields him reservation utility $\overline{U}$ (which we normalize to 0);
3. If the agent accepts, he chooses an action a at private costs of $c(a)$;
4. Uncertainty is resolved and output y becomes known;
5. The agent and the principal receive payoffs according to the contract agreed upon.

The agency model in fact comes in various forms, depending on the exact assumptions made about the agent's preferences $U(w, a)$ and how effort a stochastically maps into output y.[5] Inspired by the work of Holmstrom and Milgrom (1987) on the optimality of linear compensation schemes in some complex contracting environments, the performance measurement literature typically focuses on the so-called *linear* agency model.[6] Key assumptions of this specification are (cf. Wolfstetter 1999, pp. 283-285, Bolton and Dewatripont 2005, pp. 137-139):

A.1 The costs of effort $c(a)$ can be measured in money terms, such that the agent's utility function can be written as $U(w, a) = U(w - c(a))$ (with $U' > 0$ and $U'' \leq 0$);

A.2 The noise ε in the production function is additive: $y = \theta(a) + \varepsilon$, with $\theta(\cdot)$ increasing and concave and $E[\varepsilon] = 0$;

A.3 The compensation contract is linear in output: $w(y) = s + b \cdot y$.

An important implication of Assumptions A.1 through A.3 is that *effort incentives* are independent of the distribution of uncertainty. Put differently, the incentive compatibility constraint, and thereby the agent's optimal action

[5] See Laffont and Martimort (2002) for a book length treatment of the principal-agent model.

[6] See e.g. Baker (2002), Baker and Jorgensen (2003), Bushman et al. (2000), Datar et al. (2001), Feltham and Xie (1994) and Holmstrom and Milgrom (1991).

choice, is unrelated to the amount of noise in the performance measure. To illustrate this, from Assumption A.1 the agent's incentive compatibility constraint can be written as:

$$a = \arg\max_{\widetilde{a}} \int U(w - c(\widetilde{a}))\, dF(\varepsilon),$$

where $F(\cdot)$ reflects the distribution of the noise term ε. Differentiating this expression with respect to $\widetilde{a}$ we obtain the first order condition:

$$\int U'(w - c(\widetilde{a})) \cdot \left[\frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial \widetilde{a}} - c'(\widetilde{a})\right] \; dF(\varepsilon) = 0.$$

Assumption A.2 implies that $\frac{\partial y}{\partial \widetilde{a}} = \theta'(\widetilde{a})$ is independent of ε. Moreover, Assumption A.3 entails that $\frac{\partial w}{\partial y} = b$ is independent of y, and thus, of ε. Hence the first order condition can be rewritten as:

$$\left[b \cdot \theta'(\widetilde{a}) - c'(\widetilde{a})\right] \cdot \int U'(w - c(\widetilde{a}))\, dF(\varepsilon) = 0.$$

Given $U' > 0$ the incentive constraint reduces to:

$$b \cdot \theta'(\widetilde{a}) = c'(\widetilde{a}) \tag{1}$$

It immediately follows that the equilibrium effort level that solves (1) is independent of the distribution $F(\cdot)$. In particular, the variance in the noise term σ_ε^2 is predicted to have no effect on effort incentives.

One remark is in order. The optimal incentive intensity b^* the principal chooses in the first stage does depend on σ_ε^2. This follows because the amount of noise affects the participation constraint of the agent, which equals:

$$\int U(w - c(a))\, dF(\varepsilon) \geq \overline{U} = 0. \tag{2}$$

The optimization problem for a risk-neutral principal thus equals the following program:

$$\max_{\{s,b\}} E[y - w(y)] = (1 - b) \cdot \theta(a) - s$$
$$\text{s.t. constraints (1) and (2)}.$$

In general, the solution b^* (and s^*) to this program will depend on σ_ε^2. For example, when the noise term ε is normally distributed and the agent has CARA preferences,[7] and assuming further that productivity is linear ($\theta(a) = \theta \cdot a$) and costs of effort are quadratic ($c(a) = \frac{k}{2}a^2$), the optimal incentive intensity equals

$$b^* = \frac{\theta^2}{\theta^2 + r \cdot k \cdot \sigma_\varepsilon^2} = \frac{S}{S + r \cdot k},$$

[7] This means that the agent has constant absolute risk aversion and his utility function equals $U(w, a) = -\exp(-r \cdot [w - c(a)])$, where $r > 0$ is the risk aversion coefficient (for $r = 0$ we have $U(w, a) = w - c(a)$). CARA-preferences imply that there are no wealth effects. The principal can thus use s as an instrument to ensure participation and, independently, b to induce the agent to exert effort (cf. Laffont and Martimort 2002, p. 232).

with $S \equiv \frac{\theta^2}{\sigma_\varepsilon^2}$ reflecting the signal-to-noise ratio and $r \geq 0$ the agent's aversion to risk. The optimal incentive intensity follows from a trade-off between having to pay the agent a risk premium to participate and providing him with incentives to put in high effort. More noise (i.e. higher σ_ε^2) then lowers the optimal incentive intensity b^*. But the important thing to note is that, for a given value of b, the incentive compatibility constraint (1) is unaffected by the noise term. In the particular example, the agent chooses effort level $a^* = \frac{b\theta}{k}$, which is independent of S and σ_ε^2.

The main prediction of standard agency theory that we would like to test experimentally can thus be summarized as follows:

(AT) Keeping the incentive intensity b fixed, the agent's effort level is independent of the noise in the performance measure as measured by σ_ε^2.

This prediction will be tested against the backdrop of the expectancy model from psychology, which suggests that σ_ε^2 (and thus the signal-to-noise ratio S) will be a fundamental determinant of an agent's effort level. This will be explained in the next section.

3 Expectancy theory

3.1 Theoretical matters

Motivational models developed within the organizational psychology literature are commonly divided into two categories: one focuses on an individual's internal attributes (*content theories*) and the other focuses on the individual's interactions with the environment (*process theories*). Expectancy theory, as first developed by Vroom (1964), is a process theory of motivation. It has held a major position in the study of work motivation (cf. van Eerde and Thierry 1996) and has served as a rich source for theoretical innovations in various domains, such as organizational behavior and compensation (Lawler 1971).

Expectancy theory identifies three factors, which play an interactive role in motivation. The first of these factors, effort-performance (E-P) *expectancy*, concerns the individual's perception that effort is positively correlated with performance. The higher this E-P expectancy is, the more motivated the individual will be to exert effort. To be more precise, Vroom (1964) defines E-P expectancy as the subjective probability that an action or effort (E) will lead to an outcome or performance (P). As we will discuss below, it is this factor that distinguishes expectancy theory from agency theory and is therefore the focus of this study.

The second factor is the so-called performance-outcome (P-O) expectancy, also referred to as *instrumentality*. It concerns a person's expectation that his remuneration is closely tied to his level of performance. This factor has also a positive effect on motivation to exert effort.

The third factor is called *valence*, and is a measure of the degree to which an individual values a particular reward. Again, the higher this factor is, the more motivated the individual will be. Figure 1, taken from Fudge and Schlacter (1999), depicts the basic expectancy theory model constructed from E-P expectancy, instrumentality and valence.

[Figure 1]

Expectancy theory thus points at three instruments that employers can use to increase an employee's motivation: (i) by increasing the subjective expectations that greater effort will lead to higher levels of performance (E), (ii) by strengthening the perceived link between performance and rewards (I), and (iii) by ensuring that employees value the rewards given for high performance (V). These three factors are called the VIE factors.

Notwithstanding the substantial looseness of the definitions within the expectancy model (cf. Harder 1991), the second and third factor of the model can be nicely translated into the agency framework of Section 2. Instrumentality is given by the incentive intensity b, while valence refers to the arguments that appear in the agent's utility function (here wage w and effort a, but other arguments can easily be incorporated). The similarities between the two models might explain why, in the economics and management literature, expectancy theory and agency theory are usually taken together and opposed to still some other behavioral models (see, e.g. Harvey et al. 2001, and Pennings 1993). The first factor, however, has not yet been considered for inclusion in agency theory. It is exactly this expectancy factor that may cause agents not only to react to the pay-performance sensitivity b when they select their level of effort, but also to the amount of noise in the performance measure reflected by σ_ε^2. In this study we focus on the effect of the single factor expectancy (E) on the worker's effort level in isolation.[8]

Recall the example of the sales representative in the Introduction. Expectancy theory predicts that an agent will exert less effort when the expectancy, i.e. the perceived effect of his effort on ultimate performance, is low. Hence, in the example the sales rep will exert less effort when demand is more noisy. This holds because in that case the relative impact of effort on overall sales (as compared to noise) is much smaller. More generally, expectancy theory predicts a negative relationship between effort incentives and the amount of noise in the performance measure, because the agent's (relative) impact on the final outcome becomes smaller. In terms of the model of Section 2 this prediction can be formulated as follows:

(ET) Keeping the incentive intensity b fixed, the agent's effort level decreases with the level of noise in the performance measure as measured by σ_ε^2.

[8] Hence we do not investigate how the various VIE factors combine into worker motivation. More generally, the effects of the VIE factors on motivation can be assessed for each of the factors separately, as well as for combinations of them. In the latter case one can address question like: do the separate effects of each of the factors add up or is the effect of these factors on motivation multiplicative? Or, put in economic terms: are the three factor substitutes or complements in generating motivation?

Prediction ET suggests that according to expectancy theory σ_ε^2 – and thus the signal-to-noise ratio S – is a sufficient statistic for characterizing a performance measure. However, the example of the sales rep seems a bit peculiar in one respect. The variance σ_ε^2 of a symmetric two-point distribution $\Pr(\varepsilon = -l) = \Pr(\varepsilon = l) = \frac{1}{2}$ equals l^2. This suggests that a higher l reflects more uncertainty. This is somewhat misleading though, because for every level of l the agent has the same amount of information about what the value of the noise term will be (either $-l$ or l with equal probability). In the example it is actually the increase in the (relative) size of the noise term $|l|$ that increases σ_ε^2, rather than changes in its variability. We could alternatively increase the variability of the noise term ε, while keeping its relative size intact. But for this we need to move beyond symmetric two-point distributions (which we do in our experiment, see Section 4).

For any distribution $F(\varepsilon)$ with $E[\varepsilon] = 0$ a sharp distinction between (relative) size and variability of the noise term can be made by decomposing total variance σ_ε^2 in the following way:

$$\sigma_\varepsilon^2 = (E(|\varepsilon|))^2 + var(|\varepsilon|) \tag{3}$$

The first term on the r.h.s. can be interpreted as a measure of the relative *size* of the noise term in overall output y. The larger $E(|\varepsilon|)$, the more important luck becomes relative to the agent's own effort. The second term measures the imprecision of information about this relative size. The higher $var(|\varepsilon|)$, the less precise the agent's a priori information about the relative importance of noise for overall output.

Now expectancy theory especially predicts a negative relationship between the size of the noise term and effort incentives: keeping $var(|\varepsilon|)$ fixed (as in the example given), an increase in $E(|\varepsilon|)$ decreases effort. The effect of increasing imprecision while keeping relative size fixed (i.e. a mean-preserving spread of $|\varepsilon|$) is a priori ambiguous though. On the one hand, one may expect that a larger imprecision boosts effort incentives. This follows from the idea that when $var(|\varepsilon|)$ increases while keeping $E(|\varepsilon|)$ fixed, some low values of ε (i.e. below $E(|\varepsilon|)$) become more likely. And with a larger probability that effort can have a relatively sizable impact, agents are more willing to put in costly effort. On the other hand, some high values of ε (i.e. above $E(|\varepsilon|)$) also become more likely. This seems to weaken effort incentives. A priori the net effect is thus unclear. Our empirical investigation allows for a separate assessment of the importance of relative size and imprecision of the noise term.

In the next section we discuss the design of an experiment meant to test prediction AT of standard agency theory against the prediction ET of expectancy theory. However, before doing so, we first discuss some empirical issues resulting from the application of expectancy theory in empirical research.

3.2 Empirical matters

Because expectancy theory holds a major position in the study of work motivation, it has been subjected to ample empirical testing. In a meta-analysis, van Eerde and Thierry (1996) review the results from 77 studies that measure correlations between the VIE factors expectancy, instrumentality or valence (or combinations of these), on the one hand, and five measures of work motivation, viz. effort, performance, preferences, choices or intentions, on the other hand. In general they find significantly positive correlations for each factor in isolation. Most interesting for our purposes, the higher the expectancy is that effort will lead to performance, the more effort a worker will put forth.[9] Combining the VIE factors in various ways does not lead to higher effect sizes though. Because expectancy theory explicitly predicts that the VIE factors play an interactive role in motivation, this may be taken as evidence that the VIE-model lacks validity. As van Eerde and Thierry (1996) point out, however, many of the studies they have reviewed have either used the various concepts in another way than Vroom's (1964) theory originally implied, or have measured them in such a way that the results they produce might contain serious biases. They emphasize three aspects of measurement that can be improved in future studies: (i) the (subjective versus objective) measurement of both the VIE-factors and of work motivation, (ii) within-subjects analysis versus between-subjects analysis, and (iii) the measurement of correlations between the various VIE-factors and work motivation versus the measurement of causal effects. We will briefly discuss these three issues in turn.

The first issue concerns the measurement of the VIE-factors and of work motivation. With respect to the expectancy factor that is of main interest to us, Vroom (1964, pp. 28-30) proposes three approaches. The first one is to measure expectancies through (verbal) reports from individuals about the probability of outcomes, i.e. subjectively. This approach has the lion's share in the studies reviewed by van Eerde and Thierry (1996). They critize this approach because, in case measures of work motivation are also measured through self-reports, there is a risk that the relationship between expectancy and work motivation is spuriously inflated by common method bias and shared measurement error. The second and third approaches proposed by Vroom are hardly used and are most applicable in experimental settings:

> "One approach might be to assume that expectancies correspond perfectly with the *objective* probabilities. ... This assumption shifts the problem of measuring expectancies to one of measuring objective probabilities. ... How legitimate is it to assume that people can accurately estimate the actual probability of events? Under some conditions, such an assumption may be justified. If a person has had a considerable amount of experience in the situation attempting different courses of action and if he has been provided with prompt feedback following these actions, it might be appropriate to assume that his expectancies approximate actual probabilities. ... Alternatively we might assume that expectancies are identical with *com-*

[9] Here effort is either measured by objective measures of effort expenditure on a task such as time spent, or by more subjective measures like effort ratings by supervisors, self-reports of effort spent on a task, and intended effort.

> *municated probabilities.* ... The assumption that expectancies are completely determined by communicated probabilities seems tenable when subjects have little additional basis for judging probabilities and when they have not previously been deceived in experimentation." (Vroom, 1964, p. 29-30)

Hence these approaches might be suitable under certain (laboratory) conditions.

The second issue raised by van Eerde and Thierry (1996) is that measures of work motivation and VIE-factors are typically correlated according to a between-subjects analysis. They criticize this approach in that:

> "It is important to note that this is at variance with Vroom's (1964) idea of the model. Vroom referred to an individual's force as one which acts relative to other forces within the individual. As such, a relation between VIE variables and a criterion [measure of work motivation] should be performed according to a within-subjects analysis. ... It is unclear why so many empirical studies have used the inappropriate between-subjects methodology, although the cumbersomeness of a within-subject test may have contributed to this." (van Eerde and Thierry 1996, p. 577)

The meta-analysis demonstrates that the within-subjects correlations are significantly higher than the between-subjects correlations when effort is used as a measure of work motivation. As van Eerde and Thierry (1996, p. 582) note in this regard: "Unfortunately, there are few within-subjects correlations within our set of studies, and virtually all are based on self-reported criterion [i.e. work motivation] measures that were simultaneously taken with the VIE variables. Thus it is possible that these correlations are distorted by response bias."

The third issue that van Eerde and Thierry (1996) raise is perhaps the most fundamental. They argue that a limitation of their meta-analysis is that the direction of the effects cannot be established because the effect sizes they use in their analysis are correlations. It is thus unclear whether higher work motivation (effort) leads to higher expectancies, or whether the relationship is the other way around as expectancy theory predicts.

A controlled laboratory experiment may overcome all the problems mentioned above. First of all, in the laboratory expectancy can be controlled by means of communicated actual probabilities. As Vroom notes, the underlying assumption that expectancies are fully determined by these communicated probabilities is tenable when subjects believe that they are not being deceived. Within experimental economics, no deception is the norm (cf. Friedman and Sunder 1994, pp. 17-18). Moreover, subjects can accurately estimate actual probabilities due to the experience with the decision task gained in experimental practice rounds and to the prompt feedback following their actions. Another advantage of a laboratory experiment is that effort (as measure of work motivation) can be measured in an objective fashion, see the next section.[10]

Second, in a laboratory experiment it is straightforward to confront the same subjects with various different treatments such that the data can be analyzed on a within-subjects basis. Third, an experimental setting is ideal for assessing the direction of the effects. We can systematically vary expectancies (i.e. communicated actual probabilities) and look what the impact is on observed effort levels. Finally, as noted above empirical tests have proven it difficult to assess the interactivity of the effects of the various VIE-factors. Therefore, the mere effect of expectancy is best measured when holding the other two factors, i.e. instrumentality and valence, constant. In an experiment this is easily accomplished.

4 Experimental design

In testing prediction AT we need to keep as close as possible to Assumptions A.1 through A.3 made in Section 2. The easiest way to satisfy the requirement that effort costs can be measured in money terms would be to set up an experiment in which subjects choose a "decision number" a, where the costs of each possible choice of a are pre-specified in an increasing cost schedule $c(a)$. A higher number chosen then corresponds to more effort being exerted.[11] The disadvantage of such a setup is that effort is measured in a highly abstract and artificial way. The external validity of the results would then be strongly in demand. van Dijk et al. (2001), for instance, argue that there is a clear difference between allocating budgets and allocation real effort. They therefore use a real effort experiment to evaluate different incentive systems. Another disadvantage of identifying effort with the choice of a decision number is the lack of commitment and/or intrinsic motivation for performing the task. As these aspects may be important for the behavioral effect of effort-performance expectancy, we chose to conduct a real effort experiment that as closely as possible satisfies Assumption A.1.

In particular, we followed the real effort setup of van Dijk et al. (2001). Individual subjects had to divide their effort over two different optimization tasks, viz. task A and task B. Each task consisted of searching for high values in a two-dimensional grid through trial and error.[12] Subjects did so by taking

[10] The observability of effort may appear inconsistent with the assumption made in standard agency theory that the principal cannot observe the agent's effort. The main issue here, however, is that effort is not contractible. This may be due to effort being unobservable to the principal, or alternatively, effort being unverifiable to a third party like a court.

[11] Such a setup is for instance used by Bull et al. (1987) in their experimental evaluation of piece rate incentive systems, in the examination of tournament pay by Orrison et al. (2004), and in Fehr et al. (1993)'s experimental analysis of gift exchange in labor markets.

[12] This task originates from the ergonomic literature (see e.g. Bridger and Long 1984) and is also used in Montmarquette et al. (2004) and Pingle (1995, 1997). In other economic experiments real effort takes the form of solving mazes on the computer (Gneezy 2003), putting letters into envelopes (Falk and Ichino 2003), cracking walnuts (Fahr and Irlenbusch 2000), or answering GMAT-type of test questions (Gneezy and Rustichini 2000).

successive steps in the grid (this will be explained in more detail below). Unlike Van Dijk et al., however, we fixed the overall number of steps subjects should take. The decision problem thus boiled down to allocating a fixed number of steps over the two different tasks. The number of steps taken in each task can be interpreted as the amount of effort put forward in that task. By fixing the total number of steps, we intended to keep the overall level of effort exerted constant and we focused on a effort allocation decision instead. The (opportunity) costs of taking a step in one grid is that one step less can be taken in the other grid. The costs of putting more effort into one task can thus be measured by its opportunity costs, i.e. in money terms. This is what Assumption A.1 requires.

In the experiment overall performance in each task equalled the sum of the value found in the grid and a random noise term. With additive noise, the essential part of Assumption A.2 is satisfied. However, A.2 also assumes that the marginal productivity of effort is known to the agent. In the experiment this would come down to informing the subjects about the exact functional forms that underlie the value functions in the two grids. Clearly this would completely change the nature of the tasks and reduce them to formal optimization problems (rather than searching through trial and error). Most importantly, the amount of effort put in to solve these problems would not be easily measurable. We therefore did not inform subjects about the exact functional forms. Rather, we secured that the marginal return to effort is comparable across the two tasks. Hence, from a benefits perspective subjects had no reason at all to favor one of the tasks above the other.

Finally, Assumption A.3 is met in our experiment because subjects received a payment which is linear in overall performance. We next describe the experiment in detail.

Framing of the tasks and earnings To improve external validity we framed the experiment in a labor market context (cf. Loewenstein 1999). Subjects were requested to take the role of a sales representative of a particular company. Sales representatives are responsible for selling the company's product in two different regions, labelled A and B. The career of sales reps lasts for 30 years (rounds). In each year, their decision problem is how to divide overall selling effort over the two regions. In particular, in every year there are n working weeks available, which have to be allocated over the two regions. The number of weeks/steps n is used as a treatment variable and equals either $n = 25$ or $n = 40$. The individual allocation problem is represented by taking steps in two different two-dimensional grids. Each region is represented by one grid, see Figure 2 below. To each position (H, V) in a grid corresponds a particular function value $R = h(H, V)$, representing the effect of effort on overall sales in that region. Subjects are asked to search for high values in the grids by taking (horizontal or vertical) steps.

[Figure 2 about here]

The underlying function $h(H, V)$ that determines R is unknown to the subjects and differs across regions and over the years. The following general functional form is used:

$$R = h(H, V) = 100 - [(a_1 H - b_1)^2 + (a_2 V - b_2)^2 + c(a_1 H - b_1)(a_2 V - b_2)]^{3/4}$$

The maximum of this function is 100, which is reached for $H = \frac{b_1}{a_1}$ and $V = \frac{b_2}{a_2}$. Variations over regions and years are accomplished by varying the parameters a_1, a_2, b_1, b_2 and c. These parameters were constrained in such a way that in all cases $R = 0$ in the origin $(H, V) = (0, 0)$ where search started, the shortest route to the maximum always consisted of exactly 25 steps, and that the function h was a single peaked mountain.

To ascertain that the marginal return to effort is comparable across tasks, within each year the parameters of the two functions were chosen such that the function in one region just equalled the function in the other regions up to a rotation of either $90, 180, 270$ or 360 degrees.[13] The degree of rotation varied randomly over the years.

At the start of every year sales representatives choose in which region they want to start searching. During the year they can switch from one region to the other whenever they like and return at their last position in the region. Subjects see on their screens, see Figure 2, to what result their last step has amounted to in the region in which they were active. In the instructions subjects are explicitly informed about the fact that there are diminishing returns to effort in each region. The year ends when n steps have been made in the two regions together.[14]

The subject's actual overall sales in region J in year t equals the sum of the value reached in that region, R_{Jt}, and a noise term, ε_{Jt} (with $J \in \{A, B\}$ and $t \in \{1, .., 30\}$). These noise terms are meant to reflect the state of the economy and are drawn after a subject has finished his/her search in that year. The probability distributions may differ between the two regions. In Figure 2 they are reflected below the regions by means of little red blocks. The number of red blocks above a particular value represents the probability of that value. After a year is over, for each region separately one little red block is picked at random and this gives the outcome of the random factor ε_{Jt}. This procedure is simulated visually on the screen.

At the end of every year sales representatives are compensated based on performance pay:

[13] This comes down to selecting equal parameter values in both regions except for a_1 and a_2, and to restrict these two parameters across regions A and B in the following sense: $|a_{1A}| = |a_{1B}|$ and $|a_{2A}| = |a_{2B}|$. Thus, the optimal path and the location of the maximum are equal in both regions, except that they may be mirrored in different quadrants of the two-dimensional grid.

[14] There was a time limit of four minutes per year in which the subjects had to take all the steps. This appeared to be well enough; within this time span none of the subjects ever made less then the overall number of steps and in most cases they finished well in advance. In between steps subjects had to wait one second before they could take another step.

$$W_t = 25 + 0.5 \cdot (R_{At} + \varepsilon_{At}) + 0.5 \cdot (R_{Bt} + \varepsilon_{Bt}) \tag{4}$$

Here W_t represents the earnings of a sales representative in year t. Representatives obtain a fixed wage of 25 points and a share of 50% of the overall sales in each of the two regions. Thus, in terms of the model of Section 2, we fix the incentive intensity b at 0.5 (and $s = 25$).

Subjects' overall earnings equalled the sum of their yearly earnings. (There was no show up fee.) The conversion rate was such that 150 points in the experiment corresponded with 1 euro in money.

Distribution of the noise terms Our main treatment variable is the distribution of the noise terms ε_{At} and ε_{Bt} belonging to region A and B respectively. The 30 years that a sales rep works for a firm are divided into 6 contracting periods of five years. Within each contracting period, the distributions belonging to ε_{At} and ε_{Bt} are kept fixed. Between contracting periods these distributions vary. Before the start of a new year, subjects are informed about the probability distributions belonging to the two regions (in a similar visual way as in Figure 2). They can thus use this information for their decision with which region to begin.

Six different distributions were chosen that varied systematically in the overall variance σ_ε^2, the size $(E(|\varepsilon|))^2$ and the imprecision $var(|\varepsilon|)$ of the distribution (cf. expression (3)). Table 3 lists the different distributions used, ordered from low to high variance.[15] For ease of reference they are labelled after their main features of interest, i.e. using the two-tuple $\left\langle \frac{(E(|\varepsilon|))^2}{100}, \frac{var(|\varepsilon|)}{100} \right\rangle$. Every distribution is a (possibly degenerate) symmetric three-point distribution, which can only take the values $-l$, 0, and l.[16] For example, under distribution $\langle 64, 32 \rangle$ each of the three values -120, 0 and 120 are equally likely (cf. Table 3).

In each contracting period we compare two distributions. We use α to represent the distribution with the (weakly) lower variance and β the one with the higher variance. In comparing two distributions, we always either keep the size, the imprecision, or the variance fixed. Table **??** provides an overview of the comparisons made.

In the third contracting period (i.e. years 11 to 15), we have $\alpha = \beta = \langle 4, 0 \rangle$. The noise terms belonging to the two different regions A and B then have exactly the same distribution. Even in that case a subject might, for whatever reason, put more effort into any one of the two regions. For example, subjects who are

[15] In the three practice rounds we used two different distrubtions not listed in Table 3, viz. distribution $\langle 4, 12 \rangle$ with $Pr(\varepsilon = 0) = \frac{3}{4}$ and $l = 80$, and distribution $\langle 36, 12 \rangle$ with $Pr(\varepsilon = 0) = \frac{1}{4}$ and $l = 80$.

[16] The distributions thus have the following general form: $\Pr(\varepsilon = 0) = (1 - p)$ and $\Pr(\varepsilon = -l) = \Pr(\varepsilon = l) = \frac{1}{2}p$. We immediately obtain that $\sigma_\varepsilon^2 = p \cdot l^2$, $(E(|\varepsilon|))^2 = p^2 \cdot l^2$ and $var(|\varepsilon|) = p(1-p) \cdot l^2$. An increase in parameter l increases both imprecision and relative size. By varying p (for $p > \frac{1}{2}$) we can vary imprecision and size in opposite directions. Our choice of distributions is based on these observations. When ε is normally distributed relative size and imprecision always vary in the same direction.

Table 1: Distributions of the noise terms

	size $(E(\|\varepsilon\|))^2$	imprecision $var(\|\varepsilon\|)$	variance σ_ε^2	$Pr(\varepsilon = 0)$	$Pr(\varepsilon = -l)$ $= Pr(\varepsilon = l)$	l
$\langle 4, 0\rangle$	400	0	400	0	$\frac{1}{2}$	20
$\langle 16, 48\rangle$	1600	4800	6400	$\frac{3}{4}$	$\frac{1}{8}$	160
$\langle 64, 0\rangle$	6400	0	6400	0	$\frac{1}{2}$	80
$\langle 64, 32\rangle$	6400	3200	9600	$\frac{1}{3}$	$\frac{1}{3}$	120
$\langle 144, 0\rangle$	14400	0	14400	0	$\frac{1}{2}$	120
$\langle 144, 72\rangle$	14400	7200	14400	$\frac{1}{3}$	$\frac{1}{3}$	180

more risk averse might not want to switch after $\frac{1}{2}n$ steps to the other region because that region will give higher (and possibly negative) incremental values upon each step taken just after the start. To account for such biases caused by unobserved individual preferences, all comparisons are made relative to the observed differences in contracting period 3. In other words, this contracting period serves as a *benchmark*.

The other 5 contracting periods focus on three different types of comparisons. First, in the first and the final contracting period the higher variance of distribution β is due purely to the size effect; both α and β have the same imprecision $var(|\varepsilon|)$. This enables us to identify the pure effect of increasing relative size. Second, in the second and fifth contracting period the higher variance of β is solely the consequence of a higher imprecision; there the two distributions have the same size $(E(|\varepsilon|))^2$. Third, in contracting period 4 α and β have equal variance, but the size of distribution α is higher. This allows a test of whether the sheer composition of the overall variance matters.

The two size comparisons are chosen with special care. In the first contracting period ε_α can take the values of -20 and 20, whereas ε_β equals either -120 or 120. Because in every region the maximum function value equals $R^{\max} = 100$, under distribution β bad luck (i.e. $\varepsilon_\beta = -120$) then can never be compensated with search effort to obtain non-negative overall sales. This may especially weaken effort incentives. In contrast, in the last contracting period the β-distribution takes the values -80 and 80 with equal probabilities. Then it is possible to compensate bad luck with high effort. Although both comparisons thus focus on an increase in size, they are somewhat different in nature.

Sessions Overall we conducted four sessions, which all consisted of the six contracting periods described in Table **??**. Sessions differed along two dimensions. First, a priori we were afraid that our results might be sensitive to the number of steps n subjects should take in each year. We therefore conducted two sessions in which the number of steps is relatively low ($n = 25$) and two other sessions in which n is relatively high ($n = 40$). Second, we varied the iden-

Table 2: Comparisons between noise distributions

Contracting period	Years	α versus β $\sigma^2_\varepsilon(\alpha) \leq \sigma^2_\varepsilon(\beta)$	Type of comparison
1	1 to 5	$\langle 4,0\rangle$ versus $\langle 144,0\rangle$	size
2	6 to 10	$\langle 64,0\rangle$ versus $\langle 64,32\rangle$	imprecision
3	11 to 15	$\langle 4,0\rangle$ versus $\langle 4,0\rangle$	*Benchmark*
4	16 to 20	$\langle 64,0\rangle$ versus $\langle 16,48\rangle$	composition
5	21 to 25	$\langle 144,0\rangle$ versus $\langle 144,72\rangle$	imprecision
6	26 to 30	$\langle 4,0\rangle$ versus $\langle 64,0\rangle$	size

tity of the region with the low variance distribution α. On the computer screen region A always appeared on the left hand side and region B always appeared on the right. Clearly, individual subjects may have a bias for either side. For instance, subjects might tend to start working at the left hand side, and this could affect the total number of weeks (steps) dedicated to that particular region.[17] This would then bias the results. For both values of n we therefore conducted one session in which region A always has the low variance α-distribution, and another session in which region A always has the high variance β-distribution. The former is referred to as an (α, β) session, the latter as a (β, α) session.

The experiment was conducted in the CREED laboratory of the University of Amsterdam in April 2004. Subjects were recruited by e-mail announcements. Overall 74 subjects participated, with respectively 17, 23, 20 and 14 subjects in the four different sessions.[18] Each subject participated in one session only. Most of the subjects were (under-)graduate students, the majority of them majoring in economics. Sessions took around one and a half hours. Each session started with an identical oral introduction, read aloud to ascertain uniformity. After that subjects were randomly allocated to positions in the computer room, where they could start reading the instructions on their screen in the language of their choice (either English or Dutch; see the Appendix). After three practice rounds, the actual experiment started. Because this is an individual experiment, subjects could work at their own pace. Subjects got paid (at their desk) immediately upon finishing the experiment and left the laboratory subsequently. On average subjects earned almost 18 euros. Earnings varied considerably though, with a minimum of 7.90 euros and a maximum of 26 euros.

[17] van Dijk et al. (2001) find evidence that is consistent with such a bias, see their Table 4.

[18] We recruited the same number of subjects (22) for each of the four sessions. The differences in actual participation are due to no shows and spontaneous show-ups.

5 Results

Standard agency theory predicts that the way in which subjects allocate their effort over the two different sales regions is independent of the distribution of the noise terms (cf. Hypothesis AT). Expectancy theory, on the other hand, suggests that subjects have an inclination to bias their effort allocation towards the region with the lower variance, i.e. distribution α.[19] Effort put into a region is measured in two different ways. The most important one is the number of steps taken in (i.e. weeks devoted to) that region. The other one is the percentage of overall sales coming from that region, excluding the noise terms. For region A this measure thus equals $100 \cdot \frac{R_{At}}{R_{At}+R_{Bt}}$ in year t. Apart from that, we also consider in which region a subject starts working.

We first look at the benchmark contracting period 3 in which the noise terms in the regions have the same distribution. Table 3 provides an overview of the average amount of effort put into region A and the probability that a subject starts with this region. Both the number of steps and the relative sales appear evenly divided over the two regions. However, subjects have a tendency to start with region A which appears on the left hand side of the screen. These impressions are confirmed by statistical tests. We compare the (5 years) average amount of effort a subject devotes to region A with the average amount of effort s/he puts forward in region B by means of a sign rank test for matched pairs. Only in the fourth session we find a significant difference in the number of steps made. Both in this session and in the second session (and overall) subjects do start significantly more often with region A. There thus appears a slight tendency to start on the left.

Figure 3 depicts the descriptive statistics for each of the six contracting periods separately. The first panel concerns the number of steps in the low variance region α, the second panel the relative sales in region α and the final panel reports the propensity to start searching in region α. The figure suggests, if anything, that subjects put slightly less effort in the low variance region than they put in the higher variance region. However, these impressions from Figure 3 are not derived from a rigorous within-subjects analysis.

[Figure 3 about here]

We next test for all contracting periods but period 3 (the benchmark) whether the amount of effort put in the region with the low variance distribution α differs significantly from the corresponding region in the benchmark contracting period

[19] These predictions with respect to allocative decisions bear a resemblance with Bosman and van Winden (2001). They conduct an investment experiment in which subjects have to allocate their capital over a safe and a risky project, respectively. Irrespective of the allocation chosen, there is a ("global") risk that they lose all the returns to their allocation decision. Bosman and van Winden find that subjects allocate less to the risky project when the latter probability of global risk increases. This contradicts standard expected utility theory (which predicts no effect), but supports psychological theories that predict that anxiety motivate subjects to take less risk.

Table 3: Average outcomes in the benchmark ($A = B$ as $\alpha = \beta$)

Session	# of subjects	# of steps in A	Relative sales in A	Prob. start in A
I: $n = 40$; $(A, B) = (\alpha, \beta)$	17	20.05	50.24	0.59
II: $n = 40$; $(A, B) = (\beta, \alpha)$	23	20.72	48.80	0.78^{***}
III: $n = 25$; $(A, B) = (\alpha, \beta)$	14	12.51	50.49	0.6
IV: $n = 25$; $(A, B) = (\beta, \alpha)$	20	13.41^{**}	52.57	0.59^{**}
all sessions	74	17.04	50.47	0.65^{***}

*** $(^{**})$ significantly different from region B at 1% (5%) according to a sign rank test

3. In particular, in sessions I and III region A of contracting period 3 serves as a benchmark, whereas in sessions II and IV region B of contracting period 3 does so. Statistical tests are thus based on a within-subjects basis. In that way our conclusions are not biased due to differences in ability or risk attitude between subjects. Because for every individual subject we create an individual benchmark (viz. his/her behavior in contracting period 3) the statistical tests can be based on the overall pool of 74 subjects. Table 4 reports the $p-$values obtained from the sign rank tests.

Out of 15 comparisons, only two appear significant. In contracting period 2 the realized relative sales in the region with low variance distribution α are significantly lower than those in the corresponding region α in the benchmark period. This suggests that subjects allocate relatively more effort to distributions with higher imprecision and variance, in contrast to hypothesis ET. The second significant difference is that in contracting period 6 subjects are less likely to start with region α than in the benchmark period. This indicates that subjects prefer to start with the high size (and variance) distribution. Overall, however, with 13 (out of 15) insignificant differences, the data support hypothesis AT. This holds true especially for our most important measure of effort, i.e. the number of steps taken in a region.

The same conclusions are obtained when we perform our tests at a less aggregate level. In particular, when we consider sessions I and II (with $n = 40$), and sessions III and IV (with $n = 25$) in isolation, exactly the same test results are obtained. This indicates that the number n of total steps allowed plays no role. For the (α, β)-sessions I and III in isolation we get that, apart from contracting period 6, now also in contracting periods 1 and 2 subjects tend to start with the β-region more often than with the corresponding region B in contracting period 3. We thus obtain some evidence that subjects tend to start with the high variance region. But, their ultimate allocation of effort between the two regions is independent of the distributions of the two noise terms.

In sum, our results confirm hypothesis AT and reject hypothesis ET; subjects

Table 4: Comparisons across distributions

Contracting period	Type of comparison	# of steps in α	Relative sales in α	Prob. start in α
1	size	0.201	0.147	0.258
2	imprecision	0.187	0.004	0.105
3		*Benchmark*		
4	composition	0.690	0.698	0.650
5	imprecision	0.728	0.730	0.191
6	size	0.470	0.261	0.001

allocate their effort independently of the amount of noise in the performance measure. Moreover, because subjects do behave in accordance with standard agency theory, we do not find evidence that the signal-to-noise ratio is an insufficient statistic to characterize performance measures. There seems to be no need to make the subtle distinction between size and imprecision of the noise component.

6 Conclusion

The leading analytical model within the performance measurement literature is the linear agency model. One implication of this model is that the agent's incentive constraint is independent of the amount of noise in the performance measure. The effort level that a given compensation scheme induces is thus predicted to be independent of the distribution of the noise term. A different prediction is obtained from expectancy theory, which suggests that a higher signal-to-noise ratio will motivate the agent to exert more effort. In this paper we present the results of a laboratory experiment designed to test these opposing predictions. Subjects' effort choices appear invariant to changes in the distribution of the noise term. Our results thus validate the linear agency model in this respect and cast doubt on the relevance of this particular aspect of expectancy theory.

In a meta-analysis of a large number of different studies van Eerde and Thierry (1996) find empirical support for the particular relationship between the expectancy factor and effort. However, they also point to three empirical drawbacks pertaining to most tests of the rather loosely defined expectancy theory: (i) the use of subjective instead of objective measures of the relevant concepts, (ii) the use of between-subjects analyses instead of the theoretically correct within-subject analyses and, (iii) the measurement of correlations instead of causal effects. We have tried as much as possible to repair these empirical shortcomings in our laboratory experiment. In that sense, our study is unique.

There are (at least) three possible explanations for the discrepancy between our experimental results and previous findings. First, the empirical shortcomings of most previous empirical tests identified by van Eerde and Thierry (1996) may have caused the expectancy-effort relationship to be positive instead of insignificant. Second, although our study is unique in the sense that we have measured quite precisely the concepts related to expectancy theory, we may have failed to measure them in an externally valid way. Third, we purposely incorporated a real effort task that has been used successfully by others to study compensation issues (see e.g. van Dijk et al. 2001). Yet it might still be the case that the kind of effort subjects exert in our experiment is not representative of effort in real employment relationships. For instance, searching for a high value in a two-dimensional grid may not provide the same level of commitment and/or intrinsic motivation as actual sales activities of sales reps in reality do. If this is indeed the case our experiment fails to capture some elements that are important for the behavioral effect of effort-performance expectancy. Future experiments therefore might want to make use of more realistic/representative tasks, while keeping the same level of control. The latter may prove difficult, because especially the costs of effort are hard to measure and control.

Appendix

At the start of the experiment subjects received on-screen instructions. Subjects could choose between instructions either in Dutch or in English. Below the latter are reproduced (for the sessions in which $n = 25$).

Introduction 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 150 points = 1 euro.

We would like you to read these instructions carefully. When you have completed reading the instructions, there will be 3 practice rounds. In these practice rounds you can get familiar with the decision situation and with the various computer screens. Your points earned during the practice rounds will not be included in your final earnings. After you have completed the practice rounds, the actual experiment starts. The actual experiment consists of 30 rounds.

Your role as a sales representative In the experiment you take the role of a sales representative of a particular company. As a sales representative you are responsible for selling the company's products in two different regions. These regions are labeled region A and region B respectively.

Overall you will work for 30 years for this company. In every year, your compensation is based on performance pay. This implies that, besides a fixed wage of 25 points per year, you receive 50% of actual overall sales in regions A and B in every year. Hence your yearly earnings equal (in points):

Yearly earnings = 50% of overall sales in region A + 50% of overall sales in region B + 25

The actual overall sales within a region depend on both the amount of effort you put in as a representative in trying to sell the company's products and on the state of the economy (i.e. the business cycle). In the experiment the influence of the state of the economy is reflected by random factors that affect overall sales per region and that are beyond your control.

However, you can control, within certain limits, the impact of effort on overall sales. In every year there are 25 working weeks available. You can spend each week either on sales activities in region A, or on sales activities in region B. Your task in this experiment is to allocate the 25 available weeks over the two different regions. How this is represented in the experiment is explained on the following page.

Allocating your 25 weeks over the two regions In the experiment each region is represented by a two-dimensional grid, see the figure below. Within each grid (region) you can optimize the effect of effort on overall sales by searching for high values. Every position in the grid corresponds with a different value. You can search for high values by taking steps (i.e. putting effort into sales activities), either horizontally (left or right) or vertically (upwards or downwards). The maximum value you can reach is 100, both in region A and in region B. You can visualize this maximum value as the top of a single-peaked mountain. The mountain becomes less steep, the closer you get to its top. In other words, the additional sales returns to effort decrease within each region.

[Figure 2 appeared here]

Within each of the grids you start your search at the origin, where both the horizontal and vertical position equals 0. The corresponding value of this position is 0 too. By taking steps, one at a time, you can search for a higher value. A step means moving one position upward, downward, left or right. The maximum number of steps you can make in either direction is 25 steps from the origin. By taking a step you do not necessarily increase the value: you may get lower values and even negative values. The location of the maximum value always falls within the grid, but differs per region and also per round (year).

Overall you must take 25 steps (representing the number of available working weeks per year). You have to allocate these 25 steps over regions A and B. At the start of a year you choose with which region you would like to begin. During a year you can switch from one region to the other whenever you like. For each region the screen indicates the change in value that results from your last step in this region (see the numbers behind "Change" in the figure).

You can always go back, but this is counted as another step. For example, you take one step to the left in region B and the value in this region changes from 15 to 10. By going back, i.e. by next taking one step to the right, the value changes from 10 to its original value of 15. Although your position in the grid (and thus the value) has not changed after these two moves, the actual movements made are counted and recorded as two steps.

In each year (round) you have 4 minutes in total to make your 25 steps. The remaining time left within a year is always depicted on your screen (see the figure). As soon as you have made your 25 steps, the clock jumps ahead of time such that there are only a few seconds left. This ensures that there is no unnecessary delay after you made your 25 steps. However, in between steps there is always a delay of one second before you can make another step. In general, 4 minutes will be more than sufficient to make your 25 steps. But, in case you do not make 25 steps within 4 minutes, any remaining steps are lost. During the practice rounds you can get familiar with the decision situation and the time span per round (year).

Overall sales For both regions, the value you have reached at the end of a round (year) determines your actual overall sales for that region and year only in part. In particular, overall sales in a region equal the sum of the value reached and the random factor:

Overall sales in region A = value reached in region A + random factor region A

Overall sales in region B = value reached in region B + random factor region B

At the end of the year the random factors are determined. These are drawn from probability distributions that (may) differ between the two regions and over the years. The probability distribution pertaining to a region during a specific year is always depicted below the grid of that region; see the figure again. As you can see, in each region (and year) the outcome of the random factor can either be positive or negative (or zero), reflecting that you can either be lucky or unlucky. The expected value of the random factors is always 0. However, between regions (and over the years, see below) the random factors (may) differ in the possible outcomes that can be obtained. The random factor typically allows more extreme values in one region than in the other region. The importance of luck, i.e. the impact of the state of the economy on overall sales, thus varies over the regions.

[Figure 2 appeared here for a 2^{nd} time]

The probability that a particular value of the random factor results is represented by the amount of little red blocks above it. For each region separately the computer draws the outcome of the random factor by selecting one of the red blocks at random. Hence the more red blocks there are for a particular value of the random factor, the higher is the chance that this value will result. In the 3 practice years you can get familiar with this procedure.

In the experiment the overall sales in a region can become negative. (This reflects a situation where you did not reach a minimum sales target.) In that case you do share in the losses. Your earnings may thus be negative in some years.

Contracting periods The 30 years you are working for the company are subdivided in 6 contracting periods of 5 years each. During these five years the regions you work in as a sales representative remain unchanged. As a consequence, the probability distributions of the random factors in region A and B remain the same as well for these five years. When a new contracting period starts, you will be assigned to two other regions (although we continue labeling them A and B), such that the probability distributions belonging to the two regions change as well. At the start of each contracting period and, repetitively, at the start of each year you are informed about the probability distributions that apply in the two regions.

On your screen you will also find your total earnings from the years that already have been completed. Your overall earnings in the experiment equal the sum of your yearly earnings over your entire career of 30 years.

End of the instructions You have reached the end of the instructions. If you wish, you can turn back to some earlier pages to read some parts of the instructions again. When you click on "Ready" you will start with the three practice years. In these practice years you cannot earn any money, but you can use them to get familiar with the decision task. After you have completed the three practice years, the actual experiment starts. In case you have any questions (now or during the experiment), please raise your hand and one of us will come to your table to answer your question.

References

Baiman, S., D. Larcker, and M. Rajan (1995). Organizational design for business units. *Journal of Accounting Research 33*, 205–229.

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

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources 37*, 728–751.

Baker, G. and B. Jorgensen (2003). Volatility, noise and incentives. working paper.

Bolton, P. and M. Dewatripont (2005). *Contract theory*. The MIT Press.

Bosman, R. and F. van Winden (2001). Global risk, effort, and emotions in an investment experiment. Working Paper.

Bridger, R. and L. Long (1984). Some cognitive aspects of interface design in a two-variable optimization task. *International Journal of Man-Machine Studies 21*, 521–539.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy 95*, 1–33.

Bushman, R., R. Indjejikian, and M. Penno (2000). Private predecision information, performance measure congruity and the value of delegation. *Contemporary Accounting Research 17*, 561–587.

Datar, S., S. C. Kulp, and R. Lambert (2001). Balancing performance measures. *Journal of Accounting Research 39*, 75–92.

Fahr, R. and B. Irlenbusch (2000). Fairness as a constraint on trust in reciprocity - earned property rights in a reciprocal exchange experiment. *Economics Letters 66*, 275–282.

Falk, A. and A. Ichino (2003). Clean evidence on peer effects. IZA Discussion Paper Series No. 732.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics 108*, 437–459.

Feltham, G. and J. Xie (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review 69*, 429–453.

Friedman, D. and S. Sunder (1994). *Experimental methods. A primer for economists*. Cambridge University Press.

Fudge, R. and J. Schlacter (1999). Motivating employees to act ethically: An expectancy theory approach. *Journal of Business Ethics 18*, 295–304.

Gibbons, R. (2005). Incentives between firms (and within). *Management Science 51*, 2–17.

Gneezy, U. (2003). Do high wages lead to high profits? An experimental study of reciprocity using real effort. Working Paper.

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

Harder, J. (1991). Equity theory versus expectancy theory: The case of major league baseball free agents. *Journal of Applied Psychology 76*, 458–464.

Harvey, M., C. Speier, and M. Novecevic (2001). A theory-based framework for strategic global human resource staffing policies and practices. *International Journal of Human Resource Management 12*, 898–915.

Holmstrom, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica 55*, 303–328.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentives contracts, asset ownership, and job design. *Journal of Law, Economics and Organization 7*, 24–52.

Kerr, S. (1975). On the folly of rewarding A while hoping for B. *Academy of Management Journal 18*, 769–783.

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

Lawler, E. (1971). *Pay and organizational effectiveness*. New York: McGraw-Hill.

Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioral economics. *Economic Journal 109*, F25–F34.

Mitchell, T. (1982). Expectancy-value models in organizational psychology. In N. Feather (Ed.), *Expectations and Actions: Expectancy-value models in psychology*, Chapter 10. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Montmarquette, C., J.-L. Rulliere, M.-C. Villeval, and R. Zeiliger (2004). Redesigning teams and incentives in a merger: A real effort experiment with managers and students. *Management Science 50*, 1379–1389.

Orrison, A., A. Schotter, and K. Weigelt (2004). Multiperson tournaments: An experimental examination. *Management Science 50*, 268–279.

Pennings, J. (1993). Executive reward systems: A cross-national comparison. *Journal of Management Studies 30*, 261–280.

Pingle, M. (1995). Imitation versus rationality: An experimental perspective on decision making. *The Journal of Socio-Economics 24*, 281–315.

Pingle, M. (1997). Submitting to authority: Its effect on decision making. *Journal of Economic Psychology 18*, 45–68.

Prendergast, C. (2002). The tenuous trade-off between risk and incentives. *Journal of Political Economy 110*, 1071–1102.

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

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.

Vroom, V. (1964). *Work and motivation*. New York: Wiley.

Wolfstetter, E. (1999). *Topics in Microeconomics. Industrial Organization, Auctions, and Incentives*. Cambridge University Press.

Figure 1 An expectancy theory model

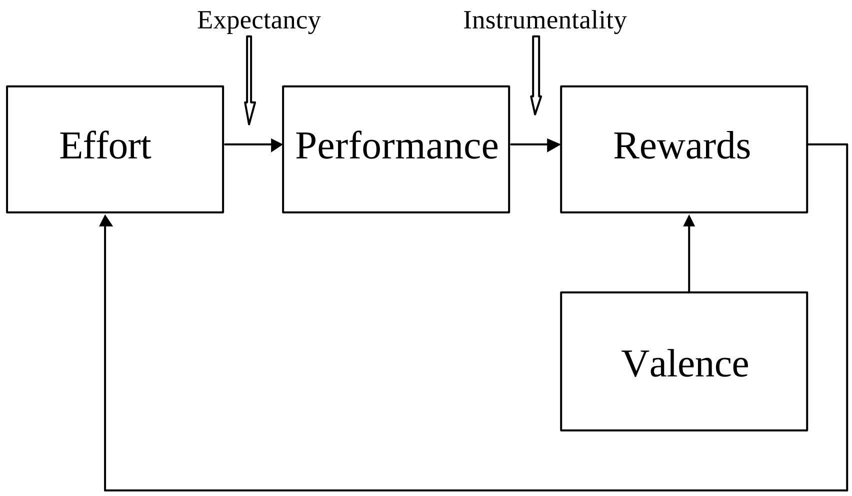


Figure 2: Computer screen reflecting the subjects' allocation problem

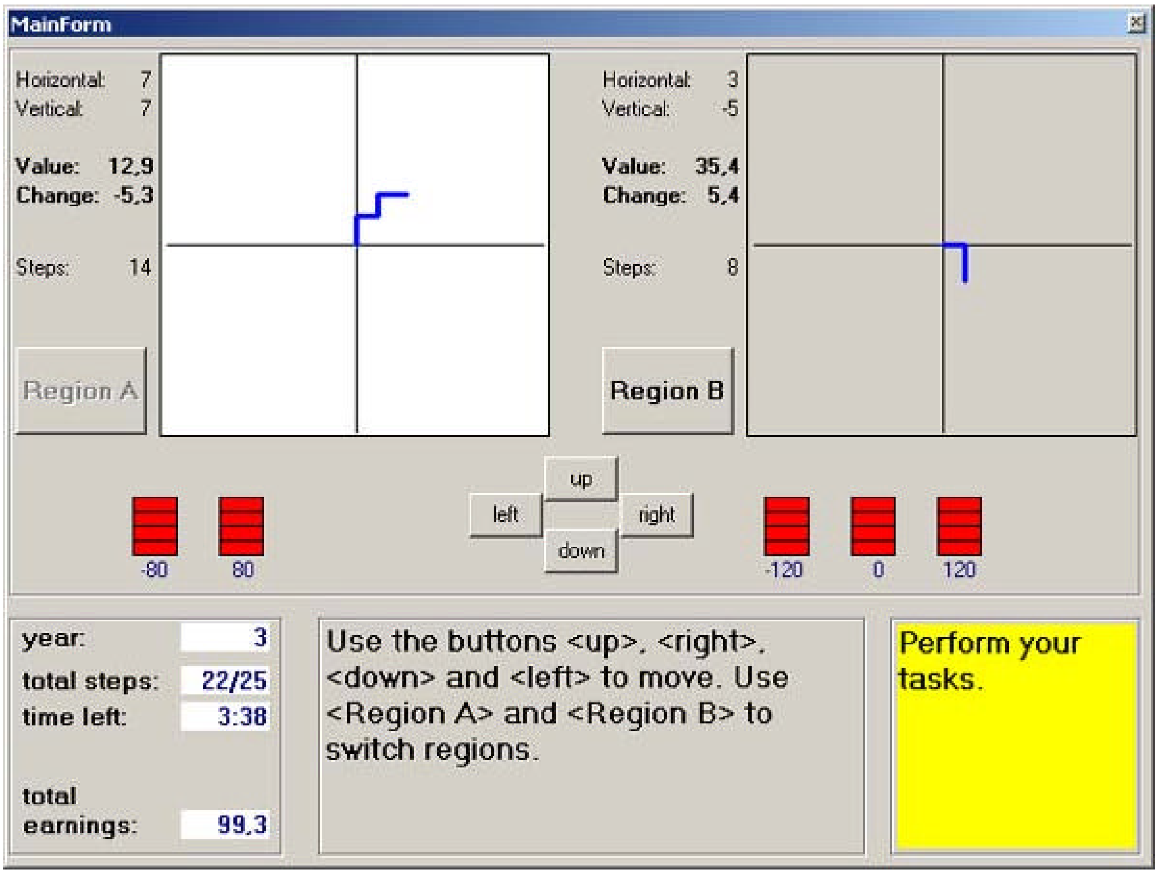
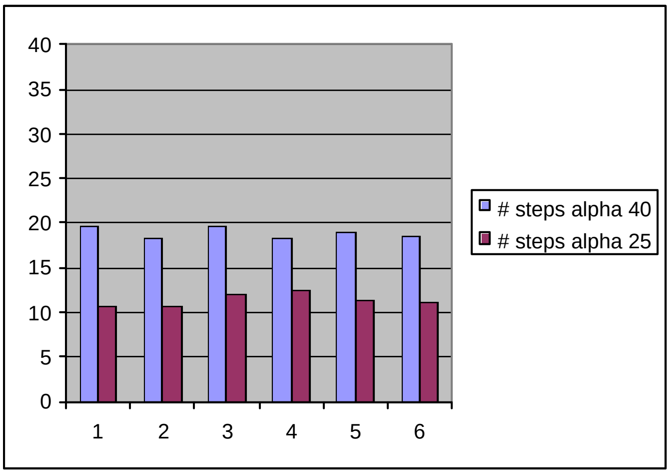
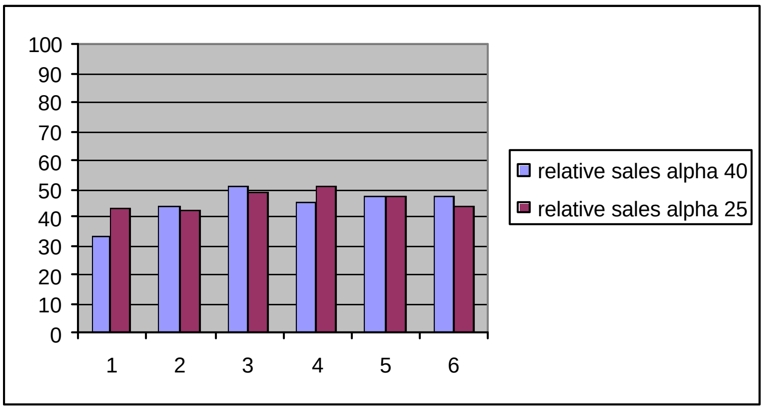


Figure 3 Descriptive statistics for each contracting period

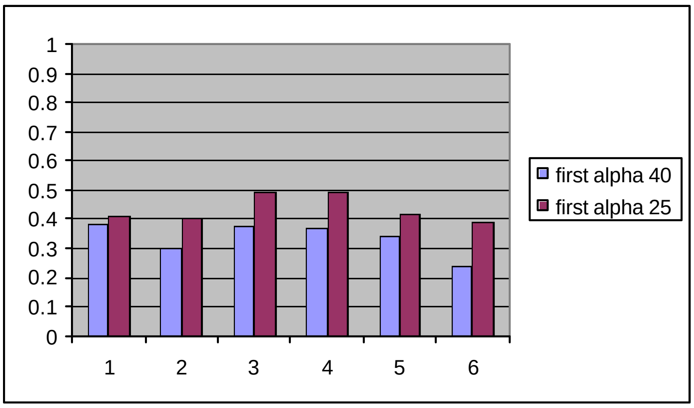
Panel A: average number of steps in region α



Panel B: average relative earnings in region α



Panel C: average starting propensity in region α



Remarks: Region α represents the region with the lower variance. Panel A depicts the number of steps taken in this region, panel B the relative sales derived from region α (excluding the noise terms), and panel C the proportion of rounds in which subjects started searching in region α.

In each of the panels averages are shown for every contracting period separately (the values 1 to 6 on the horizontal axis). In each of the panels and for each of the separate contracting periods, the first column shows the average values for the first two sessions (in which the total number of steps was 40), whereas the second column shows the averages for sessions 3 and 4 (with n=25).