Performance Measurement, Expectancy and Agency Theory

Randolph Sloof
Mirjam van Praag

Faculty of Economics and Econometrics, Universiteit van Amsterdam, and Tinbergen Institute.

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

Randolph Sloof[†]and Mirjam van Praag

University of Amsterdam and Tinbergen Institute

November 2, 2006

Abstract

Theoretical analyses of (optimal) performance measures are typically performed within the realm of the linear agency model. This model implies that, for a given compensation scheme, the agent's optimal effort is unrelated to the amount of noise in the performance measure. In contrast, expectancy theory as developed by psychologists predicts lower effort levels for noisier performance measures. We conduct a real effort laboratory experiment and find that effort levels are invariant to changes in the distribution of the noise term, i.e. to expectancy. This suggests that enriching the economic (linear agency) model commonly applied within this area by including an expectancy parameter is not needed.

Keywords: Performance measurement, expectancy theory, real effort experiments, agency theory, personnel economics.

JEL-codes: C91, J33

1 Introduction

Consider a sales representative who is responsible for selling a company's product in a particular geographic region. Her overall compensation w equals the sum of a fixed salary s and a given fraction b of overall sales. 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

---

[†]Corresponding author: University of Amsterdam, School of Economics, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. tel: +31 205255241, fax: +31 205254310, e-mail: r.sloof@uva.nl.

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.

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 volatile. 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 ($z > s$) such that the employee will not be tempted to terminate the contract. The commission rate b is kept the same though. The question is: will the sales rep exert the same level of effort as last year?

Economic theories of performance measurement and rewards typically make use of a particular linear version of agency theory. Using this standard model economists would answer the question affirmatively: the effort level actually chosen in both regions is determined by exactly the same incentive constraint $c'(a) = b$, where $c(a)$ denotes the costs of effort measured in money terms. This is independent of the distribution of the noise term ε.[1] The main driving force behind this prediction is that the sales rep's effort level has no impact on the amount of risk she faces. That is, she cannot affect the variability in her compensation by changing the amount of effort she puts in.

Notwithstanding the above prediction, intuitively it seems likely that the sales rep will put in *less* effort in the second year. The idea is that she will not only take the marginal effect of effort on sales into account, but also its absolute effect relative to the effect of noise (i.e. the effect of demand fluctuations). This intuition is an important element of expectancy theory, an influential psychological theory of motivation first introduced by Vroom (1964).

> "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)

One of the predictions derived from expectancy theory is that the incentive to exert effort will be stronger, the stronger the *perceived* relationship between effort and performance. In our earlier example this implies that, for a given level of incentive intensity b, a performance measure with less noise 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

---

[1] The standard economic model *does* predict that a noisier performance measure leads to a lower optimal commission rate b if the agent is risk averse, because high powered incentives based on risky measures require higher risk premia (i.e. a higher base salary). This is the well-known tradeoff between risk and incentives. However, *given the commission rate b,* the agent would not adapt her effort level to changes in the noise of the performance measure (see Section 2 for a full discussion).

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 the 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 in the field of organizational behavior. Within this field, it is one of the most widely applied theories in studies of incentive compensation. As a modern textbook on organizational behavior puts it:

> "Because of its profound organizational implications, expectancy theory is one of the most popular theories of work motivation." (George and Jones 2005, p. 185)[2]

Furthermore, expectancy theory has been subjected to ample empirical testing by psychologists. The meta-analysis of van Eerde and Thierry (1996, p. 581) reveals that the relationship between expectancy and effort has received mixed 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 conducted 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).

We find no evidence that effort-performance expectancy is an important determinant of the level of effort exerted by our subjects. Their effort choices appear largely independent of the noise in the performance measure. We thus obtain no evidence that the standard economic model of performance measurement and rewards should be adapted by incorporating expectancy.

This paper proceeds as follows. The next section presents the linear variant of the agency model that is widely used by economists studying performance measurement and rewards. It is formally derived that in this model the incentive constraint is independent of the noise in the performance measure. Section 3 discusses expectancy theory in more depth and relates this theory to the commonly used (agency) model for studying the economics of performance measurement. In particular, it argues that the latter model is insufficient if expectancy would be relevant. This section also discusses some measurement problems that have

---

[2] Likewise, Rollinson (2005, p. 219) states that "[Expectancy theory] is currently the most influential process theory in academic circles."

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 economic model to study performance measurement and rewards

Principal-agent theory focuses on a situation in which an agent takes an action a to 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 it is uncertain ex ante 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 $\overline{U}$;

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.[3] 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.[4] 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: $w(y) = s + b \cdot y$.

---

[3] See Laffont and Martimort (2002) for a book-length treatment of the principal-agent model in its various forms.

[4] 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).

An important implication of this set of assumptions is that *effort incentives* are independent of the distribution of uncertainty. Put differently, the incentive compatibility constraint, and thereby the agent's optimal action choice, is unaffected by 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:

$$a = \arg\max_{\widetilde{a}} \int U(w - c(\widetilde{a}))\, dF(\varepsilon), \tag{1}$$

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:

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

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 $U' > 0$ incentive constraint (1) thus reduces to:

$$b \cdot \theta'(a) = c'(a). \tag{2}$$

It immediately follows that the equilibrium effort level that solves (2) 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.

Assumptions A.1 through A.3 suffice for this prediction. Yet, additional intuition can be obtained by momentarily making the following three remaining assumptions of the standard linear agency model:

A.4 Noise term ε is normally distributed, i.e. $\varepsilon \sim N(0, \sigma_\varepsilon^2)$;

A.5 The agent has constant absolute risk aversion (CARA) preferences and his utility function equals $U(w, a) = -\exp\left(-r \cdot [w - c(a)]\right)$, where $r > 0$ is the risk aversion coefficient (for $r = 0$ we have $U(w, a) = w - c(a)$);

A.6 Linear production function and quadratic effort costs: $\theta(a) = \theta \cdot a$ and $c(a) = \frac{k}{2}a^2$ (with $\theta, k > 0$).

These additional assumptions allow an intuitive closed-form solution. Assumptions A.4 and A.5 together imply that the agent's *certainty equivalent* of his compensation contract $w(y)$ equals:

$$CE(s, b) = s + b \cdot \theta(a) - c(a) - \frac{r}{2} \cdot b^2 \cdot \sigma_\varepsilon^2. \tag{3}$$

In this particular case expected utility maximization thus effectively reduces to mean-variance preferences.[5] Together with Assumption A.6 we immediately obtain that the agent's equilibrium effort level equals $a^*(b) = \frac{b \cdot \theta}{k}$, which is independent of the variance in the noise term σ_ε^2 and the agent's risk attitude r. From expression (3) the intuition for this is straightforward. Although the agent's utility function does directly recognize the risk in the compensation contract, see in particular the last term in $CE(s, b)$ which reflects the risk premium, there is nothing the agent can do to lower the size of the disutility this risk generates (action a does not affect the risk premium). In making his action choice there is thus no risk-return tradeoff, because lower effort levels do not lead to less risk.[6]

Whereas noise by itself does not affect effort incentives, 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:

$$\int U(w - c(a)) \, dF(\varepsilon) \geq \overline{U}. \tag{4}$$

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

$$\max_{\{s,b\}} E[y - w(y)] = (1 - b) \cdot \theta(a) - s$$

$$\text{s.t. constraints (2) and (4)}.$$

In general, the solution b^* to this program will depend on σ_ε^2. For example, making Assumptions A.4 through A.6 as well, the optimal incentive intensity equals:

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

The optimal incentive intensity follows from a tradeoff between having to pay the agent a risk premium to participate (cf. term $\frac{r}{2} \cdot b^2 \cdot \sigma_\varepsilon^2$ in (3)) 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 (2) is unaffected by the

---

[5] Mean-variance analysis is typically used in models of portfolio selection. Such an analysis is consistent with expected utility maximization (where utility is defined over outcomes) only under certain conditions, of which Assumptions A.4 and A.5 are a special case. See Ingersoll (1987, pp. 95-99) and Baron (1977) for elaborate discussions of the general conditions under which this holds.

[6] There would be such a tradeoff in the presence of e.g. multiplicative noise (cf. Baker and Jorgensen (2003), Zabojnik (1996)). When $\theta \sim N(\overline{\theta}, \sigma_\theta^2)$, the certainty equivalent of the agent becomes $CE(s, b) = s + b \cdot \overline{\theta} \cdot a - c(a) - \frac{r}{2} \cdot b^2 \cdot \sigma_\varepsilon^2 - \frac{r}{2} \cdot b^2 \cdot \sigma_\theta^2 \cdot a^2$. The final term reflects that more effort now also leads to more risk. With $c(a) = \frac{k}{2} a^2$ it then follows that $a^*(b) = \frac{b \cdot \overline{\theta}}{k + r b^2 \sigma_\theta^2}$. The second term in the denominator now arises because the agent recognizes that a higher effort level will also lead to more risk. Because he is risk-averse, he reduces his effort level as compared to the situation without multiplicative noise (i.e. $\sigma_\theta^2 = 0$).

noise term. Therefore, for a given compensation contract the agent's effort level is independent of σ_ε^2.

The main prediction of the commonly used economic model of performance measurement and rewards that we would like to test experimentally can thus be summarized as follows:

(EcTh) 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 organizational behavior and psychology, which suggests that σ_ε^2 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 in 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, George and Jones 2005; Rollinson (2005)).

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 the commonly used theoretical model in economics 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 also has 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 should use in combination to increase an employee's motivation: (i) increasing the subjective expectations that greater effort will lead to higher levels of performance (E), (ii) strengthening the perceived link between performance and rewards (I), and (iii) ensuring that employees value the rewards given for high performance (V). These three factors are called the VIE-factors. To emphasize their interactive role in generating motivation, expectancy theory is typically summarized by means of the following "equation": $MF = E \cdot I \cdot V$, where MF refers to motivational force. Unlike the economic theory of performance measurement and rewards, however, expectancy theory is not cast in a rigorous formal analytical model.

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 incentive intensity b, while valence refers to the agent's utility function. Both in expectancy theory and in agency theory wage w is valued positively and effort a negatively (besides w and a other arguments can easily be incorporated in the utility function). 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, distinguishes expectancy theory from the standard economic model of performance measurement and rewards. 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. Hence we do not investigate how the three VIE-factors combine into worker motivation.

Recall the example of the sales representative given 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, the sales rep is likely to exert less effort when demand is more noisy, because the agent's (relative) impact on the final outcome becomes smaller. In terms of the model of Section 2 this prediction from organisational behavior (OB) and expectancy theory can be formulated as follows:

(OB-ExpTh) 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.

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

8

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 and five measures of work motivation, viz. effort, performance, preferences, choices and intentions. In general they find positive correlations for each factor in isolation. However, the results are mixed insofar that some effects are significantly different from zero whereas others are not. This ambiguity does also hold for the most interesting correlation for our purposes, the one between expectancy and effort.[7] Combining the VIE-factors in various ways does not lead to higher or more significant effect sizes. Because expectancy theory explicitly predicts that the VIE-factors play an interactive role in motivation, this may be taken as evidence that the 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 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 versus the measurement of causal effects. We 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 criticize this approach because, in case work motivation is 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 approach proposed by Vroom are hardly used and are best applicable in experimental settings:

> "One approach might be to assume that expectancies correspond perfectly with the *objective* probabilities. ... 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 *communicated probabilities*. ... The assumption that expectancies are

---

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

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)

The second issue raised by van Eerde and Thierry (1996) is that the correlations between measures of work motivation and the VIE-factors are typically obtained from 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 from the experience with the decision task gained in experimental practice rounds and 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.

Second, in a laboratory experiment it is straightforward to confront the same subjects with 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 measure the impact 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 constant. In an experiment this is easily accomplished.

4 Experimental design

In many laboratory experiments that study effort incentives the worker's effort decision corresponds to choosing a "decision number", with higher numbers being increasingly costly. Because effort is then measured in a highly abstract and artificial way, the external validity of the results may be questioned. van Dijk et al. (2001), for instance, argue that there is a clear difference between allocating budgets and allocating real effort. 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 closely follows the one of van Dijk et al. (2001). In particular, we adapted their setup such that Assumptions A.1 through A.3 made in Section 2 are satisfied as closely as possible. To improve external validity we also framed the experiment in a labor market context (cf. Loewenstein 1999).

   Below we first describe the real effort task subjects had to perform and how they were rewarded. After that we provide an overview of our treatments and sessions.

4.1   Description of real effort task and payment scheme

Our 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 is a fixed number $n \in \{25, 40\}$ of working weeks available, which have to be allocated over the two regions. Each region is represented by a two-dimensional grid, see Figure 2 below. Selling effort takes the form of searching for high values in this grid through trial and error.[8] To each position (H, V) in a grid corresponds a function value $h(H, V)$, 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. Each step represents a working week.

[Figure 2 about here]

---

[8] This optimization 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).

At the start of every year sales reps 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, what result their last step has amounted to in the region in which they were active. In the instructions subjects are explicitly informed that there are diminishing returns to steps in each region (see below). The year ends when exactly n steps have been made in the two regions together.

The subject's actual overall sales in region J in year t equal 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 reflect the impact of the state of the economy on overall sales and are drawn after a subject has finished his/her search in that year. The probability distributions may differ between the two regions (see the next subsection where we discuss the treatments). 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 on the basis of performance pay:

$$W_t = 25 + 0.5 \cdot (R_{At} + \varepsilon_{At}) + 0.5 \cdot (R_{Bt} + \varepsilon_{Bt}) \tag{5}$$

Here W_t represents the earnings of a sales representative in year t. Representatives thus obtain a fixed wage of 25 points and a share of 50% of the overall sales in each of the two regions. Note that the variability in compensation W_t caused by the random factors ε_{At} and ε_{Bt} is independent of how effort is allocated over the two regions, i.e. independent of R_{At} and R_{Bt}. Subjects' overall real earnings equal the sum of their yearly earnings. (There was no show up fee.) The conversion rate is such that 150 points in the experiment correspond with 1 euro in money.

The parameters of the value function $h(H, V)$ that determines R_{Jt} differ across regions and over the years. They were chosen such that the function $h(H, V)$ always equals a single peaked mountain with a maximum of 100. Moreover, in all cases $R_{Jt} = 0$ at the origin $(H, V) = (0, 0)$ where search starts and the shortest route to the maximum always consists of exactly 25 steps.[9] Towards the optimum the incremental value of an additional step decreases, representing that there are diminishing returns to effort in each region. Subjects were informed about these aspects of R_{Jt}.

Our effort allocation task is similar to the one used by van Dijk et al. (2001). In their experiment subjects also divide effort between two tasks A and B. The

---

[9] Following van Dijk et al. (2001) we used the following general functional form: $h(H, V) = 100 - [(a_1 H - b_1)^2 + (a_2 V - b_2)^2 + c(a_1 H - b_1)(a_2 V - b_2)]^{3/4}$. The maximum of 100 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. In all cases $\left|\frac{b_1}{a_1}\right| + \left|\frac{b_1}{a_1}\right| = 25$, to secure that the optimum is always exactly 25 steps from the origin.

only difference is that we fixed the overall number of steps subjects should take.[10] The idea behind this setup is that each region (task) provides an opportunity cost to allocating effort to the other region. That is, the costs of taking a step in one region is the lost opportunity of taking a step in the other region. 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 in Section 2 requires.[11]

The yearly earnings (5) of representatives are linear in overall performance. This is in line with Assumption A.3, with the incentive intensity b fixed at 0.5. Because the noise in overall performance in each task is additive, the essential part of Assumption A.2 is satisfied as well. Yet this assumption also supposes that the marginal productivity of effort is known to the agent. This would come down to informing subjects about the functional form $h(H, V)$ that underlies the value functions in the two grids. Clearly this would change the nature of the tasks completely and reduce them to formal optimization problems, rather than searching through trial and error. Most importantly, the amount of effort exerted to solve these problems would not be easily measurable. We therefore did not inform subjects about the exact functional form $h(H, V)$. We did secure though that the (diminishing) marginal return to effort is comparable across regions. In each year the two value functions were chosen such that the function in one region just equalled the one in the other region up to a rotation of either $90, 180, 270$ or 360 degrees (this was unknown to the subjects).[12] The degree of

---

[10] By doing so, we intended to keep the overall level of effort exerted constant. van Dijk et al. (2001) allowed subjects 50 seconds in which they could take as many steps as they liked (with a 1.5 seconds delay in between steps). For their purposes it was important (just like it is for us) to keep the extent to which subjects exert effort constant over the different treatments. Therefore, they explicitly test if the total number of steps taken does not vary between treatments (and find that this is indeed not the case). By fixing the overall number of steps in advance we avoided having to test for equality ex post (and run the risk of significant differences).

[11] Formally, for the effort allocation task effort incentives follow from the agent's incentive compatibility constraint:

$$(a_{At}, a_{Bt}) = \arg \max_{(\widetilde{a}_{At}, \widetilde{a}_{Bt})} \int U\left(W_t\left(\widetilde{a}_{At}, \widetilde{a}_{Bt}\right), c\left(\widetilde{a}_{At} + \widetilde{a}_{Bt}\right)\right) dF\left(\varepsilon_{At}, \varepsilon_{Bt}\right)$$

$$\text{s.t. } \widetilde{a}_{At} + \widetilde{a}_{Bt} = n$$

Here it is assumed that the (utility) costs of taking a step is independent of the region, i.e. $c((\widetilde{a}_{At}, \widetilde{a}_{Bt}) = c((\widetilde{a}_{At} + \widetilde{a}_{Bt})$. Plugging the fixed overall effort constraint in the agent's objective function, the incentive compatibility constraint reduces to:

$$a_{At} = \arg \max_{\widetilde{a}_{At}} \int U\left(W_t\left(\widetilde{a}_{At}, n - \widetilde{a}_{Bt}\right), c(n)\right) dF\left(\varepsilon_{At}, \varepsilon_{Bt}\right)$$

From this we obtain the first order condition:

$$\left(\frac{\partial W_t}{\partial R_{At}} \frac{\partial R_{At}}{\partial \widetilde{a}_{At}} - \frac{\partial W_t}{\partial R_{Bt}} \frac{\partial R_{Bt}}{\partial (n - \widetilde{a}_{At})}\right) \cdot \int U'\left(W_t, c(n)\right) dF\left(\varepsilon_{At}, \varepsilon_{Bt}\right) = 0.$$

Again, the agent's optimal choice a_{At} is independent of the distribution of the noise terms $F\left(\varepsilon_{At}, \varepsilon_{Bt}\right)$. In particular, the agent choose a_{At} and $a_{Bt} = n - a_{At}$ such that $\frac{\partial R_{At}}{\partial \widetilde{a}_{At}} = \frac{\partial R_{Bt}}{\partial \widetilde{a}_{Bt}}$.

[12] 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 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

rotation varied randomly over the years. From an economic perspective subjects thus had no reason at all to favor one of the regions above the other.[13] And given that the marginal return of a step is decreasing in the number of steps along the optimal path, the economic model predicts that subjects take about the same number of steps in the two tasks (cf. van Dijk et al. 2001, p. 196).

4.2 Treatments and sessions

Our main treatment variable is the distribution of the noise terms ε_{At} and ε_{Bt}. The 30 years that a sales rep works for a firm are divided into 6 contract periods of five years. Within each contract period, the distributions belonging to ε_{At} and ε_{Bt} are kept fixed. Between contract 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 in the variance σ_ε^2. Table 1 provides an overview of the distributions used, ordered from low to high variance. Every distribution is a (possibly degenerate) symmetric three-point distribution, which can only take the values $-x$, 0, and x. For example, under distribution $-120/0/120$ in Table 1 each of the three values -120, 0 and 120 are equally likely.

In each contract period we compare two distributions. The right half of Table 1 provides an overview of the comparisons made. For ease of reference we use α (β) to represent the distribution with the lower (higher) variance. In the third contract period the noise terms ε_{At} and ε_{Bt} actually have the same distribution. This contract period serves as a benchmark. 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 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 we employ a differences-in-differences approach; all comparisons are made relative to the observed differences in the benchmark contract period 3 (with $-20/20$ vs. $-20/20$).

The other 5 contract periods compare distributions that differ in their variance to various degrees. In the first contract period, for example, 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 $R^{\max}$ equals 100, under distribution β bad luck (i.e. $\varepsilon_\beta = -120$) then can never be compensated with search effort to

---

both regions, except that they may be mirrored in different quadrants of the two-dimensional grid.

[13] We also tested the equality of the perceived difficulty of the two regions by comparing the ranked difference of the function value of coordinate $h(1,1)$ in every year across the regions. This is the value that results after two steps if a player moves into the (most natural) northeastern quadrant and proceeds into both the horizontal and the vertical direction. The rank was not systematically higher or lower for either region ($p = 0.978$). We therefore conclude that the two regions were indeed of comparable difficulty.

14

Table 1: Distributions of the noise terms and the comparisons made

Label	Distribution	σ_ε^2	Contract	Years	α versus β
$-20/20$	$\frac{1}{2} \circ -20 \oplus \frac{1}{2} \circ 20$	400	1	1 to 5	$-20/20$ vs. $-120/120$
$-160/0/160$	$\frac{1}{8} \circ -160 \oplus \frac{3}{4} \circ 0 \oplus \frac{1}{8} \circ 160$	6400	2	6 to 10	$-80/80$ vs. $-120/0/120$
$-80/80$	$\frac{1}{2} \circ -80 \oplus \frac{1}{2} \circ 80$	6400	3	11 to 15	$-20/20$ vs. $-20/20$
$-120/0/120$	$\frac{1}{3} \circ -120 \oplus \frac{1}{3} \circ 0 \oplus \frac{1}{3} \circ 120$	9600	4	16 to 20	$-160/0/160$ vs. $-80/80$
$-120/120$	$\frac{1}{2} \circ -120 \oplus \frac{1}{2} \circ 120$	14400	5	21 to 25	$-120/120$ vs. $-180/0/180$
$-180/0/180$	$\frac{1}{3} \circ -180 \oplus \frac{1}{3} \circ 0 \oplus \frac{1}{3} \circ 180$	21600	6	26 to 30	$-20/20$ vs. $-80/80$

Remark: The notation $p \circ -x \oplus (1-2p) \circ 0 \oplus p \circ x$ in the second column means that the noise term equals $-x$ with probability p, 0 with probability $(1-2p)$ and x with probability p. In the final column α (β) represents the low (high) variance distribution, i.e. $\sigma_\varepsilon^2(\alpha) \leq \sigma_\varepsilon^2(\beta)$.

obtain non-negative overall sales. This may especially weaken effort incentives. In contrast, in the last contract period the β-distribution takes the values -80 and 80 with equal probabilities, such that it is possible to compensate bad luck with high effort. By including both comparisons we can establish whether effort incentives decrease proportionally with σ_ε^2.

For the comparisons made in periods 1 ($-20/20$ vs. $-120/120$) and 6 ($-20/20$ vs. $-80/80$) the increase in variance is clear cut; noise term ε_β takes more extreme values than ε_α does and distribution β is just a stretching of distribution α. In the terminology of Rothschild and Stiglitz (1970, p. 229), β differs from α by a *single* mean-preserving spread. In contract periods 2, 4 and 5, in contrast, we compare a two-point distribution with a three-point distribution. In all these cases the three-point distribution has the (weakly) higher variance. (Note that in contract period 4, i.e. $-160/0/160$ vs. $-80/80$, distribution α and β actually have the same variance.) At the face of it this may not be obvious though, because under the three-point distributions not only the extreme values are more extreme, but also $\varepsilon = 0$ becomes more likely. Therefore, distribution β is not just a stretching of distribution α and it cannot be obtained by a single mean preserving spread.[14] We include these comparisons in order to test hypotheses EcTh and OB-ExpTh also for cases where σ_ε^2 increases in less obvious ways.

Apart from the distribution of the noise term we used the number of steps n as a second treatment variable. 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 considered both the case in which the number of steps is relatively low ($n = 25$) and the one in which n is relatively high ($n = 40$). To avoid confusion every subject was confronted with one of these treatments only.

Overall we conducted four sessions. In the first two sessions $n = 40$, in the two remaining sessions $n = 25$. Moreover, between sessions we also varied the identity of the region with the low variance distribution α. On the computer screen region A always appeared on the left and region B always appeared on the right. Clearly, subjects may have a bias for either side. For instance, they might tend to start working at the left hand side, and this could affect the total number of weeks (steps) dedicated to that region.[15] 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 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

---

[14]In contract periods 2 and 5 distributions α and β satisfy the so-called integral conditions of Rothschild and Stiglitz (1970). This implies that β can be obtained from α by a *sequence* of single mean preserving spreads. Acccording to the four measures of riskiness distinguished by Rothschild and Stiglitz, distribution β therefore displays greater uncertainty than α does. The distributions used in contract period 4 cannot be ordered according to these criteria.

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

in the four different sessions.[16] Each subject participated in one session only. Most of the subjects were undergraduate students 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).[17] After three practice rounds, the actual experiment started. Because this is an individual experiment, subjects could work at their own pace. Subjects got paid 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.

5   Results

Standard economic theory predicts that the way in which subjects allocate their effort over the two regions is independent of the distribution of the noise terms (cf. Hypothesis EcTh). Expectancy theory, on the other hand, predicts that subjects will bias their effort towards the region with the lower variance, i.e. with distribution α. Effort put into a region is measured in a number of different ways. The most important one is the number of steps taken in that region. Other, more indirect measures of effort that we explore are: the percentage of overall sales coming from a region (excluding the noise terms), the probability that a subject starts working in a region, and the average marginal return for the final step taken in a region.[18,19] Except for the last one, all these measures are supposed to increase as effort increases. On the contrary, due to diminishing marginal returns towards the optimum, a higher marginal return to the final step indicates that *less* effort is devoted to that region, i.e. it is an *inverse* measure of effort.

We first look at the benchmark contract period 3 in which the two noise terms have exactly the same distribution $-20/20$. Table 2 provides an overview of the average amount of effort put into region A according to the various measures of effort considered. Both the number of steps and the relative sales appear evenly divided over the two regions. Also the average marginal return for the final step taken in a region is almost the same for the two regions. However, subjects do

---

[16] 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.

[17] The instructions in English can be downloaded from the first author's website: www1.fee.uva.nl/scholar/mdw/sloof.

[18] We are thankful to an anonymous referee who suggested the last (inverse) measure of effort.

[19] The idea behind the starting region as a measure of effort is that, if subjects perceive one region to be "better" and thus worth more their effort, they may want to start in that region. However, we also find a bias to start on the left (see below), suggesting that subjects use simple clues to decide when they are indifferent. The noise structure may just be another simple clue to solve indifference. The starting region thus only provides a weak measure of effort at best.

seem to have a tendency to start with region A which appears on the left hand side of the screen.

The above 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 taken. Both in this session and in the second session (and overall) subjects do start significantly more often with region A. There thus appears a tendency to start on the left. No significant differences are found for relative sales and the marginal return to the final step taken in a region.

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, while the third panel reflects the propensity to start searching in region α. To allow an insightful comparison between the low variance and the high variance region, the remaining effort measure is also depicted in a relative way. In particular, panel D depicts as relative measure of effort (final marginal return in β − final marginal return in α) / final marginal return in α. Positive values then reveal that effort put in the region with the lower variance α is higher (recall that final marginal return is an inverse measure of effort).

Overall, the different panels in Figure 3 do not show a clear pattern. Panels A, B and C suggest, if anything, that subjects put slightly *less* effort in the low variance region than they put in the high variance region. Panel D yields rather mixed results. Note, however, that these impressions from Figure 3 are not derived from a rigorous within-subjects analysis.

[Figure 3 about here]

Our formal statistical analysis is based on a differences-in-differences approach. For each effort measure of Table 2 we calculate for each individual and each year, the difference between effort in the region with the low variance distribution α and the high variance distribution β.[20] We subsequently take the average of these differences for each contract period, i.e. we calculate for each individual subject the average difference over the five years in which the two distributions α and β stay the same. By means of sign rank tests we then compare the observed differences in a particular contract period with those in the benchmark contract period 3 where $\alpha = \beta = -20/20$. Because these statistical tests are based on a within-subjects comparison our conclusions are not biased due to differences in ability or risk attitude between subjects. Moreover, since we create an individual benchmark for every subject, the statistical tests can be based on the overall pool of 74 subjects. Table 3 reports the $p-$values obtained from the sign rank tests.

---

[20]Table 2 shows that region A corresponds with distribution α in sessions S1 and S3. In sessions S2 and S4 region A corresponds with distribution β.

Table 2: Average outcomes in the benchmark ($A = B$ as $\alpha = \beta = -20/20$)

Session	n	(A,B)	# of subjects	# of steps in A	Relative sales in A	Prob. start in A	Return last step in A (B)
S1	40	(α,β)	17	20.05	50.24	0.59	2.87 (2.64)
S2	40	(β,α)	23	20.72	48.80	0.78***	2.46 (2.77)
S3	25	(α,β)	14	12.51	50.49	0.6	3.49 (3.46)
S4	25	(β,α)	20	13.41**	52.57	0.59**	3.57 (3.63)
all			74	17.04	50.47	0.65***	3.05 (3.10)

Remark: ***(**) significantly different from region B at 1% (5%) according to a sign rank test. Relative sales in A equal $100 \cdot \frac{R_{At}}{R_{At}+R_{Bt}}$ (in year t).

Out of 20 comparisons, only two appear significant at the 5% level. In contract period 2 the realized relative sales in the region with low variance distribution α are significantly lower than those in the corresponding region in the benchmark period. This suggests that subjects allocate relatively more effort to distributions with a higher variance, in contrast to both hypothesis EcTh and OB-ExpTh. The second significant difference is that in contract 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 variance distribution. Overall, however, with 18 out of 20 insignificant differences, the data support hypothesis EcTh. This holds true especially for our most important measure of effort, i.e. the number of steps taken in a region.

To further explore the validity of our conclusions we perform two types of robustness checks. First we investigate whether learning could be an issue. Perhaps, the tendency to put in effort in the low variance region increases as subjects have done more rounds (years) in the experiment. This appears not to be the case. When we rebuild Tables 2 and 3 using only the scores of either the first or the last year within a contract period (instead of the average over 5 years), almost identical results are obtained. The single important exception is that we now find a third significant difference, viz. in contract period 6 relative sales in region α are higher than in region β when we focus on the final year only. Overall, however, our results are robust.

Second, the same conclusions are also obtained when we perform our tests at a less aggregate level. In particular, when we consider sessions S1 and S2 (with $n = 40$), and sessions S3 and S4 (where $n = 25$) 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 S1 and S3 in isolation we get that, apart from contract period 6, now also in contract periods 1 and 2 subjects tend to start with the β-region more often than with the corresponding region B in contract period 3. We thus obtain some evidence that subjects prefer to start with the high variance region. However, as argued above the starting region is a weak measure of effort at best. The other, arguably more convincing measures of effort reveal that the ultimate allocation of effort between the two regions is independent of the distributions of the two noise terms.

Our within-subjects design adds to the confidence we have in our insignificant results. As explained in Subsection 3.2 our design follows the recommendations of van Eerde and Thierry (1996), who argue that a between-subjects methodology is inappropriate. They also observe in their meta-analysis that within-subjects correlations (between the VIE-factors and work effort) are significantly higher than between-subjects correlations. From that perspective our within-subjects analysis thus gives the Ob-ExpTh hypothesis a fair chance.[21] If anything, in a between-subjects design we would expect an even weaker correlation between the noise in the performance measure and effort.

---

[21]The typical concern raised by experimental economists against within-subjects designs is that by exposing subjects to multiple conditions, one heightens their sensitivity to the differences in conditions (cf. Camerer (2003, p. 42)). Also from that perspective a within-subjects analysis provides a conservative test of hypothesis Ob-ExpTh.

Table 3: $p-$Values of comparisons across distributions (α versus β)

Contract period	α versus β	# of steps in α	Relative sales in α	Prob. start in α	Return last step in α
1	$-20/20$ vs. $-120/120$	0.201	0.147	0.258	0.634
2	$-80/80$ vs. $-120/0/120$	0.187	0.004	0.105	0.632
3	$-20/20$ vs. $-20/20$		Benchmark		
4	$-160/0/160$ vs. $-80/80$	0.690	0.698	0.650	0.257
5	$-120/120$ vs. $-180/0/180$	0.728	0.730	0.191	0.794
6	$-20/20$ vs. $-80/80$	0.470	0.261	0.001	0.704

Remark: $p-$values are obtained from sign rank test for matched pairs.

21

It is important to keep in mind that our finding that subjects divide their effort about equally over the less noisy and the more noisy region, provides no information at all about subjects' underlying risk attitudes. In our experiment the incentive scheme is such that subjects can do nothing to affect the amount of risk they face. The variability in their compensation (due to the additive noise in the performance measure) is completely independent of their effort allocation choices. Therefore, economic theory predicts the *same* behavior for any level of a subject's risk aversion $r \geq 0$ and our design cannot be used to infer risk attitudes from observed choices (nor does it intend to do so).[22]

In sum, our results confirm hypothesis EcTh and reject hypothesis OB-ExpTh; subjects allocate their effort irrespective of the amount of noise in the performance measure.

6 Conclusion

The leading economic 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 (additive) 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 more uncertainty in the relationship between effort and performance will de-motivate the agent to exert 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 some 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 potential explanations for the discrepancy between our experimental results and previous findings. First, the shortcomings of most previous empirical tests identified by van Eerde and Thierry (1996) may indeed

---

[22] In contrast, Holt and Laury (2002) confront subjects with a menu of choices between a "safe" lottery and a "risky" lottery, where the "risky" lottery has the higher variance. The more risk-averse a subject is, the more often s/he will choose the "safe" lottery. Because subjects with different risk attitudes are predicted to behave differently, the Holt and Laury design permits measurement of the degree of risk aversion (and the specific functional form it takes).

have biased the expectancy-effort relationship upwards as the authors of the meta-analysis suggest. 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 and Montmarquette et al. (2004)). 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.

References

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.

Baron, D. (1977). On the utility theoretic foundations of mean-variance analysis. *Journal of Finance 32*, 1683–1697.

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

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.

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.

Camerer, C. (2003). *Behavioral game theory. Experiments in strategic interaction.* Princeton University Press.

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

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.

George, J. and G. Jones (2005). *Understanding and Managing Organizational Behavior*. New Jersey: Pearson Prentice Hall.

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 & Organization 7*, 24–52.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*, 1644–1655.

Ingersoll, J. (1987). *Theory of Financial Decision Making*. New Jersey: Rowman and Littlefield Publishers.

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.

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.

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.

Rollinson, D. (2005). *Organisational behavior: An integrated approach* (3rd ed.). FT Prentice Hall.

Rothschild, M. and J. Stiglitz (1970). Increasing risk: I. a definition. *Journal of Economic Theory 2*, 225–243.

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.

Zabojnik, J. (1996). Pay-performance sensitivity and production uncertainty. *Economics Letters 53*, 291–296.
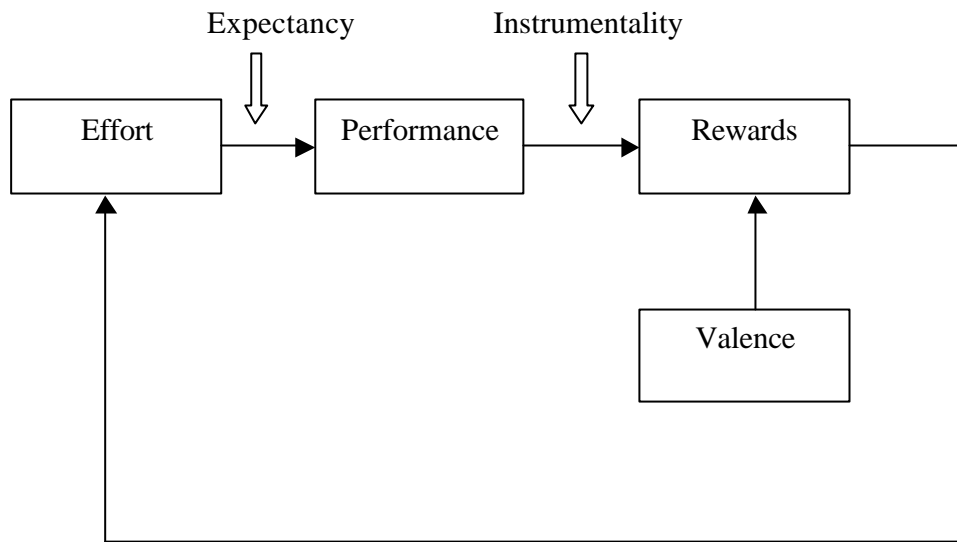
Figure 1: An expectancy theory model

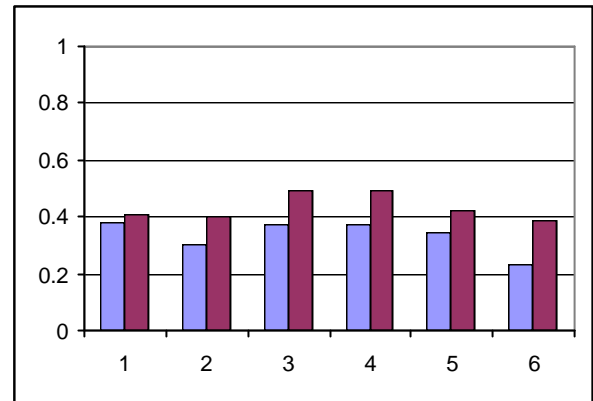Figure 2: Computer screen reflecting the subjects' allocation problem
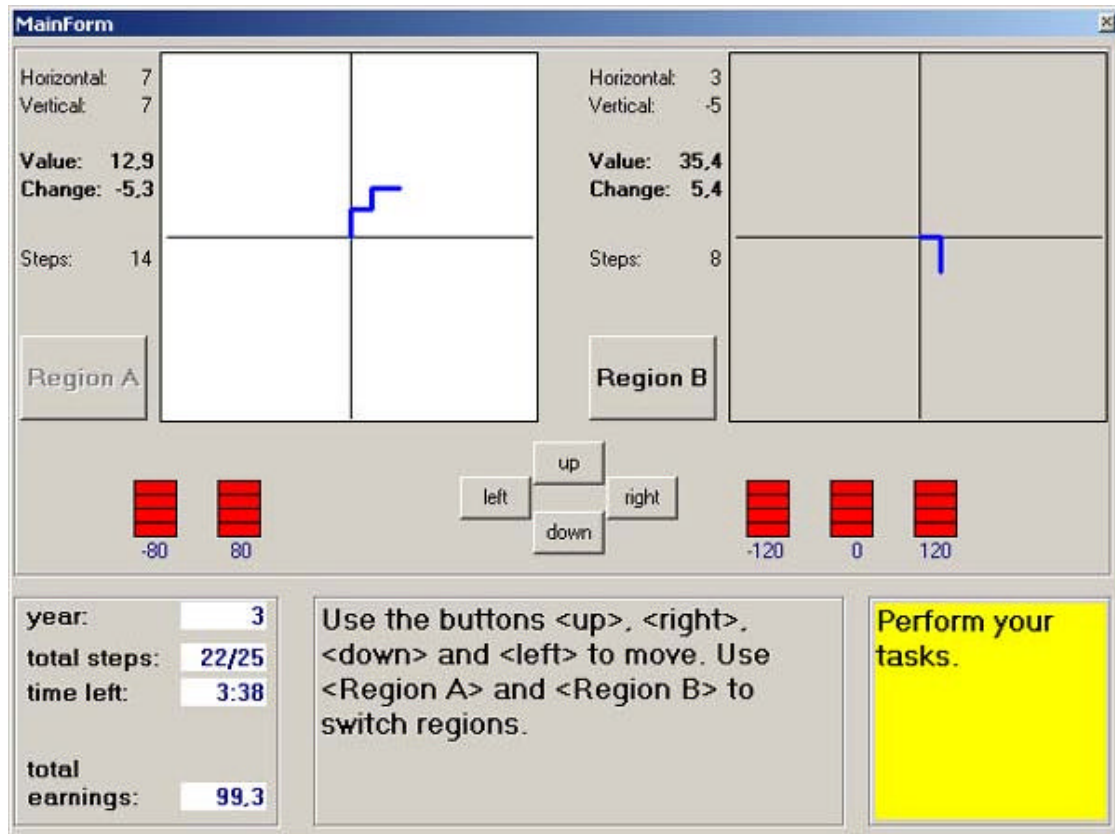
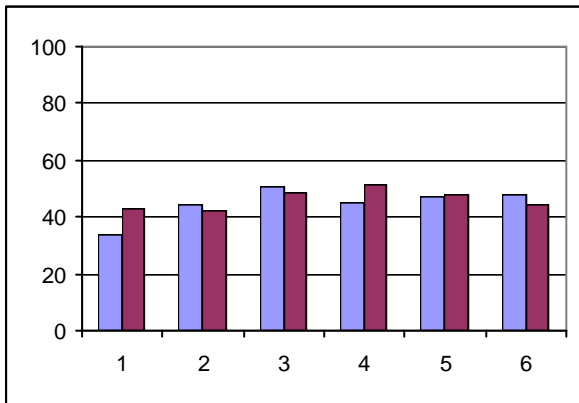# Figure 3 Descriptive statistics: effort put in region **a**[*]

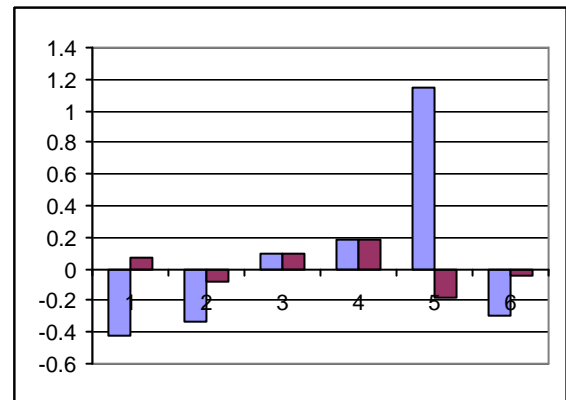Panel A: average number of steps in region α
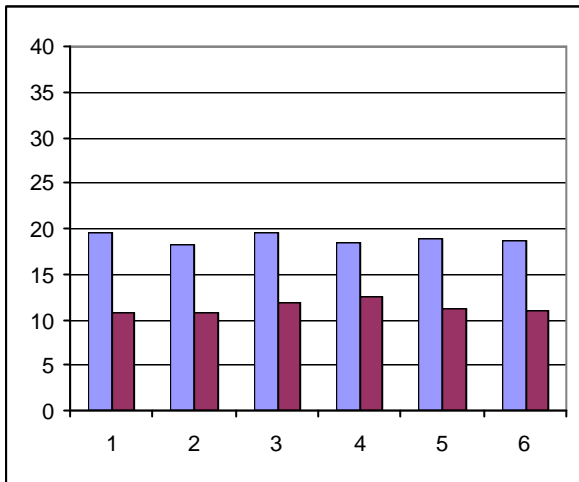
Panel C: average starting propensity in region α

Panel B: average relative earnings in region α

Panel D: average relative final marginal gain in region β

In all panels: ▢ n = 40   ▮ n = 25

[*] *Remarks*: Region α represents the region with the lower variance. In panel A effort is measured as the number of steps taken in this region, in panel B effort is represented by the relative earnings derived from region α, whereas in panel C effort devoted to region α is measured as the proportion of rounds started in that region. In panel D effort is represented as the marginal gain of the final step taken in region β relative to region α, i.e. the final marginal return in region β minus the final marginal return in region α divided by the final marginal return in region α. The measure of effort would be positive in case individuals put more effort into region α (resulting in lower absolute levels of the final marginal return in this region as compared to the other region).
All measures of effort in each of the panels 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 (light) column shows the average value of the effort measure for the first two sessions in which the total number of steps was 40, whereas the second (dark) column shows the average effort level for sessions 3 and 4 with n = 25.