LUNDS UNIVERSITET

Department of Psychology

PSY 161 HT-2004

MASTER'S THESIS

HUMAN INSPECTION WORK

−A case study of why faults are missed?

Magnus H. Lindblad

Tutor: Professor Curt R. Johansson
Work and Organizational Psychology Division
The Department of Psychology
Lund University

APPROVALS

[Insert university's approval sheet here]

ACKNOWLEDGEMENTS

With out the support and commitment by all the different departments that have given me all the information and help I asked for, this study never could have been made.

The same is true for the help and support from my tutor Curt R Johansson and the company contact who possessed a fantastic ability to "open doors".

Sincerely Magnus Lindblad

ABSTRACT

Visual inspection in order to assure quality control is a challenging field that is attracting increasing interest after a long period of silence due to the trust and hope in computers and automation. Human visual inspection is however a challenging area as it requires the designer of the system to take in to account demands and limitations of technical, economic and psychological nature.

This study was conducted in a real world setting in cooperation with a company with manufacturing in East Asia. Semi-participant observations were used to study the work in the factory. The results from the study highlight the tremendous importance of not only taking the usual economical and technological considerations when designing a human inspection system. Humans have demands and limitations that the inspection/test system has to be adapted to.

Human inspection/test performance was found to have large individual differences. The overall performance tended however to decline over the work shift although a main effect could not be statistically secured ($p = .060$). Paired samples t-test showed on the other hand significant results between work shift periods 1-3 and 2-3 and was very close between 1-4 ($p = .052$). Performance was also impaired by reduced test times as the correlation results show ($r_s = -.48$, $p = .001$, 1-tailed).

This research has also pointed out the danger with monotonous work tasks due to habituation and expectancy effects and the subsequent need for stimulation in the work situation. The results also highlight the need for autonomy, goals and feedback to improve motivation and so performance.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

HUMAN INSPECTION WORK

This project was conducted as a case study in cooperation with a company in the electronics field that was interested in the performing more product-tests manually, in order to develop a faster more flexible and adaptive test-system. It was suggested that before such a step, there was a need for a critical investigation of the manual tests already in use. This project analyzes the work in two manual stations in the test-system: Manual Test One (MT1) and Manual Test Two (MT2).

Quality control by the use of visual inspection is however a challenging field because it requires the designer of the system to take in to account demands and limitations of technical, economic and psychological nature. Because of this wide research field the project was split into two papers. Human Inspection Work: ─A case study of quality control performance in a modern global electronics company by Lindblad (2005), which has an industrial management perspective. It handles the questions: What is the performance of the system? How can it be improved? This paper has a psychological perspective handling "why" faults were missed. The goal of this paper is to describe and explain the possible reasons for the system performance on a human level, not primarily to give recommendations on improvements.

The aim of Lindblad (2005) was to precise the demands the current system puts on the operators/inspectors. What were the: work tasks, output demands, stress level and what was the performance of the system in terms of hit rate and fault rate? Those results showed that the work tasks and performance demands will lead to a highly stressful work situation according Karasek's demand -control model. The demands on the operators' attention were also high. Some of the faults were very small 0.025 mm2, which equals a visual angle of 0.0036 degrees at a viewing distance of 400 mm. Other fault symptoms occurred rarely, on average once every 20 workdays. The performance in terms of fault detection hit rate was lower what the company had hoped for, the estimated hit rate, depending on fault symptom, ranged from 58-87%. (Lindblad, 2005) This led to the question; why so many faults were missed on a human level?

The first step was to review the literature and research on inspection work and visual inspection. Working with this project in a practical way with the goal to actually improve inspection performance showed the value of theories. The psychological theories are of this reason given a large space in the paper.

Working in the "real world" also created a need to combine quantitative and qualitative methods. Numbers and values were very useful to communicate results, but also "scientifically ungrounded" results were very valuable because there are few true answers in the real world and an educated guess could quite possibly be the best there is. This structure or lack there of is reflected in the paper.

Quality control

The aspect of *quality* has had a dramatic development in the last part of the 1900 century and the trend continues in to the 2000 century. Quality improvements have on many markets gone from a way to win orders, to a qualifier that companies have to meet not to loose orders (Hill, 2000). The high demands on quality have dramatically lowered the fault rate. 20 years ago it was said that a 20% reject rate was realistic (Gallwey & Drury, 1986). In this study on the other hand one of the fault symptoms occurred on average once every 20 workdays and no symptom had a higher fault rate than 2% (Lindblad, 2005). This improved quality makes the demands even higher on the operators.

The rapid technological development and the stricter quality demands have changed the *quality control* work in radical ways. Instead of randomly testing a sample of the products, more and more products are tested by a 100% inspection. This inspection has changed from *manual inspection* and testing to an increasingly automated process. This because *human inspection* has on several studies proven to be less than 100% accurate (Drury & Sinclair, 1983; Gallwey, 1998; Jiang, Gramopadhye, & Melloy, 2003).

Automated inspection systems are improving as computer technology and image-processing are developing and falling in price. These systems can perform simple boring tasks for extended periods of time for which human operators are poorly suited, yet humans still outperform machines in many *inspection tasks*. Automated systems are also task-specific and inflexible (Gallwey, 1998; Jiang et al., 2003).

Visual inspection in order to control quality is a challenging field that is attracting increasing interest, this after a long period of silence due to the trust and hope in computers and automation. There are two main reasons the globalization of the economy and the relocation of manufacturing to countries with lower operator salaries. Automated test systems have had problems to successfully move from the laboratory to volume production. This study has highlighted the tremendous importance of not only taking the usual economical and technological considerations when designing a human inspection system. Humans have demands and limitations that the inspection/test system has to be adapted to.

Human error

Human error are responsible for 80% of industrial accidents, 20-50% of all equipment failures, 50% of pilot accidents and 20-70% of all system failures at nuclear power plants of which the Three-Mile Island and Chernobyl are the most serious (Buffardi, Fleishman, Morath, & McCarthy, 2000; Czaja, 1997).

Several factors affect human error rates. On a personal level: cognitive abilities, perceptual-motor abilities, sensory abilities, motivation and fatigue; and on an organizational level: factory type, organizational climate, equipment and task design (Buffardi et al., 2000).

Human error can be divided into, *slips* and *lapses* (unintentional execution errors on skill-based tasks) and *mistakes* (cognitive errors on rule and knowledge-based tasks). *Skills* are learnt and task specific stored patterns of behavior that requires little conscious effort. Existing rule-based behavior are often used in familiar situation. Knowledge-based behavior is used in specific situations requires cognitive processing based on system knowledge. Ability is a more general capacity useful in many tasks (Buffardi et al., 2000; Park, 1997). Inspection errors can be further divided in *misses* (undetected faults or type 2 error) and *false-alarms* (good items rejected or type 1 error) see Figure 1.

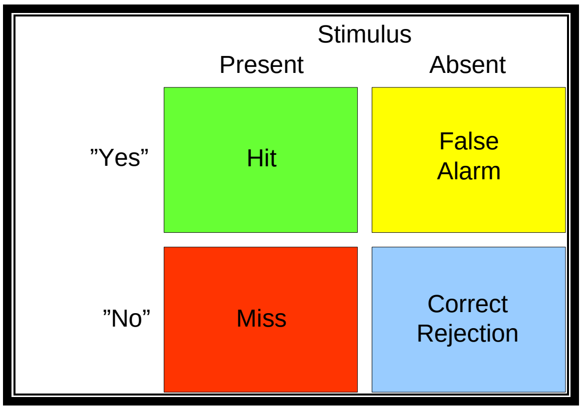


Figure 1. *Possible inspection outcomes.*

Gallwey and Drury (1986) reviewed experiments on multiple fault types and concluded that performance was worse for inspecting for multiple fault types compared to a single fault type. Their own experiments confirmed this. When six fault types were used compared to two, did the number of search errors increase with 60%. The decision errors were depending on the complexity of the standards by which flaws were judged. This implies that the standards should be as simple as possible and memory aids be used.

Vigilance

The *vigilance decrement* is a theory concerning the question why faults are missed. The vigilance decrement is the deterioration in signal detection performance often observed on tasks were randomly occurring, normally easily detectable signals with a low probability are to be detected against a background of non-critical signals over a prolonged period of time and under monotonous conditions. The one contributing most to the theory is N. H. Mackworth. He investigated why radar operators during World War II reported more enemy aircrafts in the beginning of their shifts than later, even though there ought to be the same intensity (Craig & Colquhoun, 1975; Jiang et al., 2003; Nachreiner & Hänecke, 1992; Poulton, 1970).

This deterioration can be dramatic as Fox (1975, p. 89) writes:

> With prolonged inspection periods it is commonly observed in such analyses that fault detection deteriorates as a function of time. The rate of deterioration can be rapid- drops of 40% in 30 minutes have been noted. As many paced inspection tasks are carried on for periods well in excess of half-an-hour, the implications of the phenomenon for quality assurance are serious and some understanding of its basis is correspondingly important.

The vigilance problem starts getting serious after the first 30-40 minutes, see Figure 2. No human operator can be constantly alert for prolonged periods. The tasks should therefore be self paced, breaks provided and any kind of changes are important to help maintaining a high vigilance (Gallwey, 1998; Poulton, 1970).

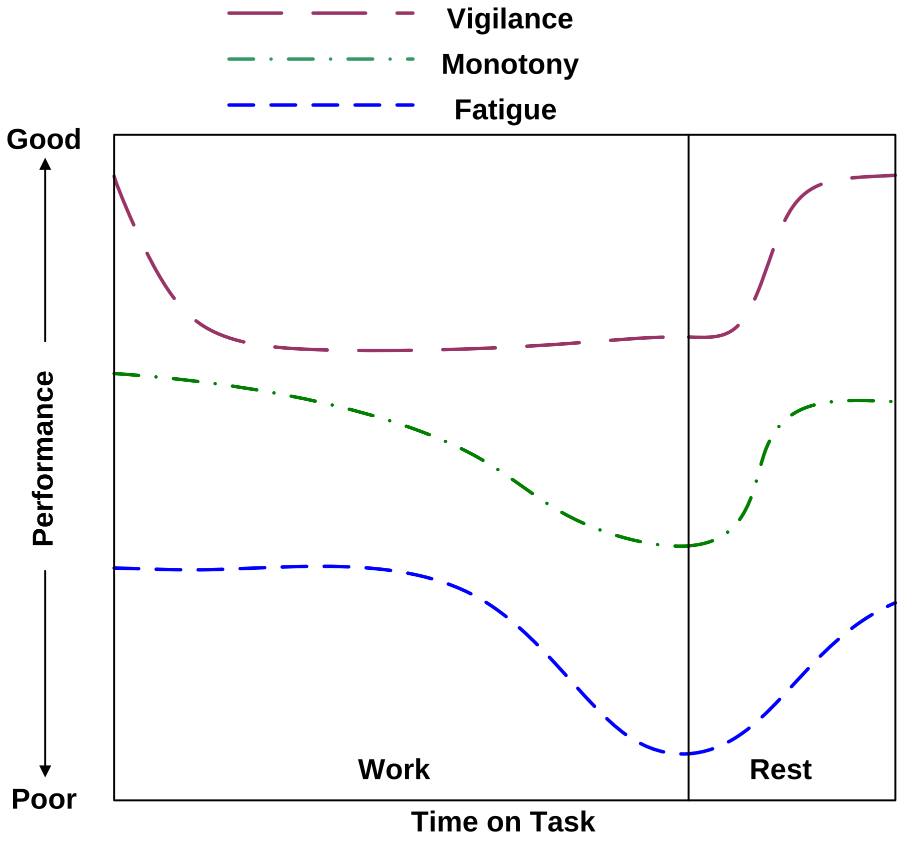


Figure 2. Typical performance curves for different forms of impairment.
After a figure by Nachreiner and Hänecke in Smith and Jones III (1992).

There are two main schools of explaining the vigilance decrement; one that focuses on physiological explanations, as *arousal* and *inhibition*; the other on cognitive mechanisms, as *expectancy* and *motivation* due to benefits/cost of different actions. These theories are not mutually exclusive so by combining them a richer theory is formed (Craig & Colquhoun, 1975; Nachreiner & Hänecke, 1992).

Arousal theory

Arousal theory states that the central nervous system requires a minimum level of activity to function well. This is supported by studies on *sensory deprivation*. Craig and Colquhoun (1975) cites research by Bexton et al. (1954) and Vernon et al. (1959) that showed that the test participants that were restricted from sensory stimulation preformed worse on a majority of tests of mental performance. Poulton (1970) suggests that the impaired performance can be due to the sudden dramatic change in stimulation from the tests themselves and so increasing arousal over an optimal level. I am skeptical towards this suggestion, because from my experience arousal does not increase so rapidly. For example when a person is woken from deep sleep (the lowest point of arousal) and are asked questions, his or her poor performance can hardly be blamed on over-arousal.

A person's arousal varies over time, as mentioned, from deep sleep to highly agitated, but there are also individual differences. People with an *extrovert personality* tend to have a lower level of arousal and a higher need for stimulation than *introvert individuals*, especially in the morning (Matthews, 1992). *Extraversion* has been linked to poor performance in vigilance tasks (Craig & Colquhoun, 1975).

It is important to have the right level of arousal as the *Yerkes-Dodson law* implies, see Figure 3. The figure shows that performance increase with arousal to a certain point and then drops if arousal keeps increasing. The turning point on performance depends on task difficulty so a difficult task may suffer while a simple task may benefit from increased arousal. Increased arousal leads to a focus of attention. This can in turn lead to tunnel-vision and failure of taking into account all relevant information (Hygge, 1992; Molloy & Parasuraman, 1996; Nachreiner & Hänecke, 1992; Smith & Jones, 1992).

Anyone that have faced a really demanding task have experienced a need for minimizing disturbance, if this is because of over-arousal due to stress or that we only have limited resources for attention could and are debated.

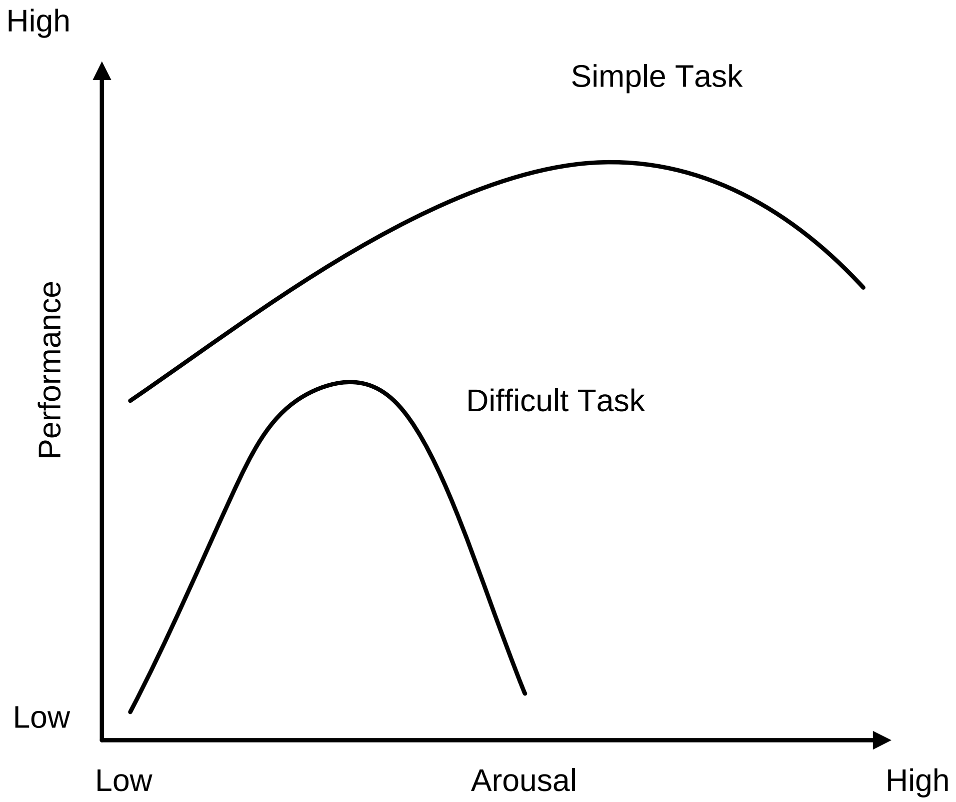


Figure 3. The Yerkes-Dodson inverted U relationship between arousal and performance. After a figure by Hygge in Smith and Jones I (1992).

This relationship results in a need to be able to affect ones own arousal level. Mackworth showed that a stimulant drug could abolish the vigilance decrement (Craig & Colquhoun, 1975). Other less extreme ways of increasing the arousal level is music, a worm environment, usage of different skills, feedback. and short breaks (Adams, 1975; Craig & Colquhoun, 1975; Gallwey, 1998; Hackman, Oldham, Janson, & Purdy, 1975; Henning, Jacques, Kissel, Sullivan, & Alteras-Webb, 1997; Hygge, 1992; Smith & Jones, 1992).

Stress

The demand aspect of *stress* is closely related to arousal and can in many cases be described with a U-function, so that at the optimal level of arousal there is a minimum feeling of stress but on either side of the optimum is performance suboptimal and so the stress increase. Stress was introduced to the life sciences by Hans Selye 1936 (Sanders, 1983).

The definition of stress has changed over the years. Stress has been defined as external stimuli, or as events termed *stressors*, that subject us to (too) high demands. Selye on the other hand defined stress as an internal specific response of the body to any demand upon it. These definitions lack the emotional and cognitive aspect that many intuitively associate with stress. Richard S. Lazarus and Lennart Levi have had major influence on the modern way to look at stress as an *intervening variable* dependent on the situation and the interaction and transaction between the individual and the environment (Sanders, 1983; Schnall, 2004 a; Smith, 1993).

S Michie (2002, p. 67) provides a modern definition based on the ideas of Lazarus and Levi that captures this cognitive aspect:

> "The generally accepted definition today is one of interaction between the situation and the individual. It is the psychological and physical state that results when the resources of the individual are not sufficient to cope with the demands and pressures of the situation. Thus, stress is more likely in some situations than others and in some individuals than others."

The stress response is believed to be a primeval reaction to sudden danger where the organism mobilizes energy to prepare for *fight or flight*, see Figure 4. The body tries to balance the threat with a series of reactions in the organs in order to maintain *homeostasis*, the delicate balance between the different organ systems inside the body.

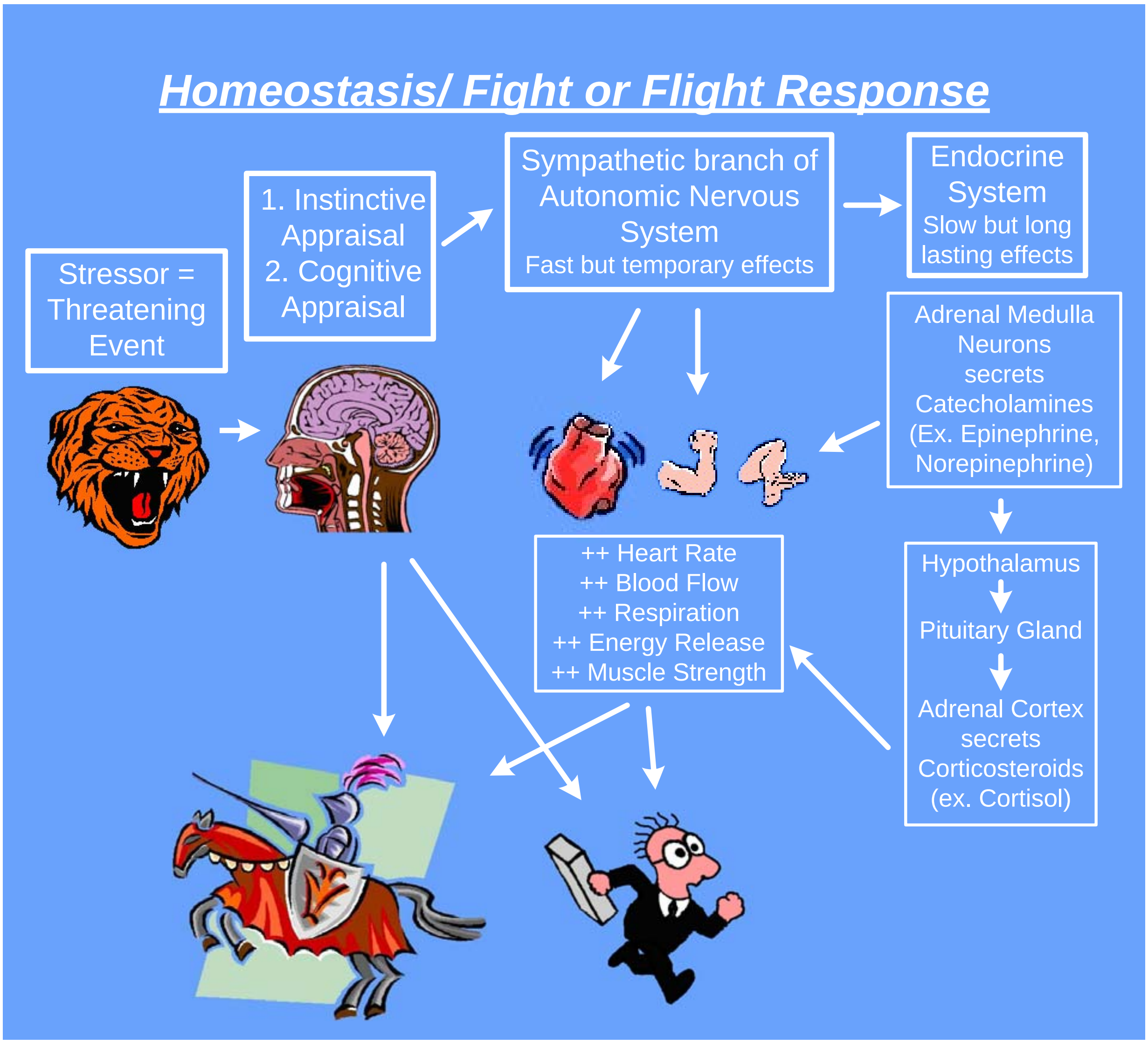


Figure 4. *The different stages of the fight- flight response, based on a figure in Schnall 2004 a.*

The physical response to stress and the activation of the *sympathetic adrenal-medullary (SAM) system* has its roots in the work by Walter B. Cannon. This response is developed for temporary activation. The problem is that many work situations today leads to a chronic activation of this response; this affects health, employee absenteeism, employee turnover and motivation (Michie, 2002; Sanders, 1983; Schnall, 2004 a; Schnall & Patel-Coleman, 2004; Schnall & Perlo, 2004; Smith, 1993). Carruthers describes the stress response cited in Smith (1993, p. 471) as a:

> "Stone Age physiological and biochemical responses to emotion have become inappropriate in a Space Age setting, and can pave the way to psychosomatic diseases".

The effort-distress model is stress model that explores the relationships of stress, feelings and hormones created by Marianne Frankenhauser and her colleagues in Sweden. They have identified two main components of stress: *effort* expended in combating the stress and *distress* associated with negative feelings. Effort is connected to active coping and attempts to take control, leading to positive feelings of interest, engagement and determination. Distress is associated with boredom, depression, passivity, uncertainty and anxiety and an often helpless approach to behavior (Craig & Cooper, 1992; Schnall, 2004 a).

The interesting finding is that the two types of responses are connected to two biological systems, the *sympathoadrenal medullary system* (*catecholamines, adrenalin* and *noradrenalin*) and the *pituitary-adrenal cortical system* (*corticosteroid*s such as *cortisol*), seen in Figure 4. Experiments have showed that in the face of demanding but controllable and predictable stressors does the adrenalin levels increase, but cortisol levels decrease. The feeling is that effort without distress is experienced. Low demand-low control situations leads to elevated cortisol levels, with only mild elevations in catecholamines leading to feelings of depression and helplessness. (Schnall, 2004 a)

In the typical vigilance task on the other hand, which is characterized by low control and rather high demands, do both the adrenalin levels and cortisol levels increase, producing a heightened feelings of both distress and effort (Craig & Cooper, 1992).

General Adaptation Syndrome (GAS) was discovered by Hans Selye in the 1930s. He noticed that patients with quite different illnesses shared many of the same symptoms, such as muscle weakness, weight loss, and apathy. Selye conducted experiments on a variety of physical stressors such as: heat, cold, poisons, strenuous exercise, and electric shock. He found that the stressors all produced a similar response: enlargement of the adrenal glands, shrinkage of the thymus gland (a gland involved in the immune response), and bleeding stomach ulcers (Schnall, 2004 a; Smith, 1993).

Selye proposed a model with three-stages describing the stress response, the GAS model. The stages of GAS are: *alarm*, *resistance* and *exhaustion*. The alarm stage is the response to the stressor which activates the sympathetic nervous system and increases physical arousal. During resistance the body tries to maintain homeostasis by mobilizing the body's resources and secreting *stress hormones*. Physical arousal is still high under this phase which can last for a relative long time. Exhaustion is the result of the depletion of the body's resources due to a too long period of resistance or a shorter period with a very high level of stress. If the resources are completely depleted then the body is extremely vulnerable to disease and even death (Schnall, 2004 a; Smith, 1993).

Karasek's demand-control model captures Frankenhauser's and Selye's results, see Figure 5, but it is also tied to theories on motivation as will be showed later. Robert Karasek started the development of the model during the seventies for work environments where the stressors are: chronic, not initially life threatening and which are the product of sophisticated human organizational decision making. Table 1 shows the two dimensions of the model, *psychological demands* and *decision latitude*. Decision latitude is a combined measure made-up of *task control* (*decision authority*) and *skill use* (*skill discretion*) (de Rijk, 1998; Schnall, 2004 b).

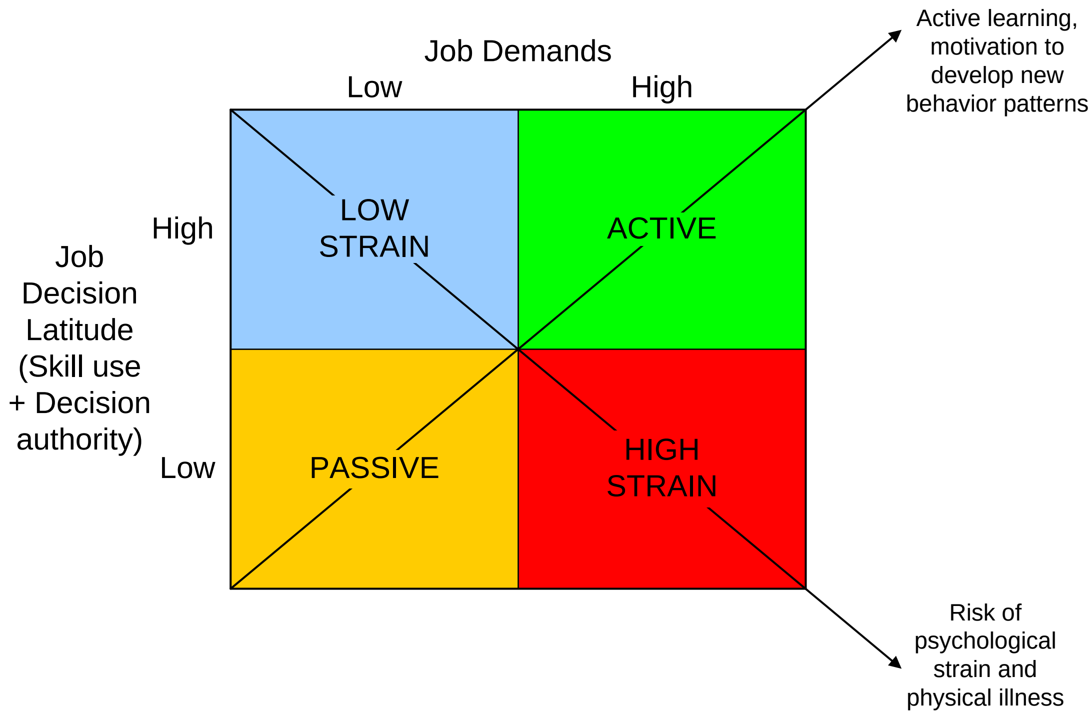


Figure 5. The demand- control model, after a figure in Karasek (1979).

Table 1.
The two dimensions psychological demands and decision latitude in Karasek's demand- control model.

Psychological demands	Decision latitude*
1. Excessive work	1. Have freedom to make decisions
2. Conflicting demands	2. Choose how to perform work
3. Insufficient time to work	3. Have a lot of say on the job
4. Work fast	4. Keep learning new things
5. Work hard	5. Can develop skills
6. Intense concentration	6. Job requires skills
7. Often interrupted	7. Task variety
8. Very hectic	8. Repetitious
9. Waiting on others	9. Job requires creativity
*Decision authority is measured by items 1-3; skill discretion is measured by items 4-9.	

It is dangerous to experience high strain or *job strain* for a long time. Because the body's ability to restore and repair tissues will be gradually affected. An example of a high strain occupation is the assembly-line worker which has almost every behavior rigidly constrained and is often helpless to increases in demands, the effects are also often chronic. It is worth pointing out that cardiovascular disease (CVD) is the major cause of mortality in the industrialized world (Schnall, 2004 b).

Recent cross-cultural studies have proven this model in China. It was found that job time demands and job decision latitude had direct and interactive effects on psychosomatic health. The effect of job time demands was greater than that of job decision latitude. Results from Mexico have on the other hand showed that job insecurity was best predictor of depression, anger and exhaustion; not psychological demands nor decision latitude (Schnall & Patel-Coleman, 2004).

Later research has found a need to add a third dimension, *social support*. Social support is the most important psychosocial stress buffer and has been found to reduce cardiovascular diseases and enhance immune function (de Rijk, 1998; Hammar, Alfredsson & Johnson, 1998; Patel-Coleman, 2004).

The negative effects of stress are many, already mentioned are the negative effects on health and motivation. The most relevant to this paper is that stress impairs performance (Michie, 2002; Smither, 1994). Stress can lead to fatigue and impaired attention, cognition and timing of movement (Craig & Cooper, 1992; Patel-Coleman, 2004). Employees that are physically present at their jobs may show decreased productivity and quality -a concept known as decreased *presenteeism*. Presenteeism is a health-related productivity loss while paid at work. Presenteeism is often referred to as LPT of LWPT (lost (work) productive time). Some data suggest that presenteeism is a larger productivity drain than either absenteeism or short-term disability (Schnall & Perlo, 2004). According to Frankenhaeuser can higher demands be countered by increasing effort or by letting performance drop. The increase of effort is primarily a temporary adjustment (Craig & Cooper, 1992).

Personality and feelings of stress are connected as previously stated, much like arousal but more complex due to the cognitive component. How a situation is perceived or appraised determines to a high degree which emotion and the level of emotion that is experienced (Lazarus, 1993; Smith, Haynes Lazarus & Pope, 1993; Michie, 2002). People that don't think about stress have been found to be less stressed (Smither, 1994).

Several studies have found a significant relationship between the personality disposition *negative affectivity* and feelings of stress. Negative affectivity means that the person has a generally negative self-concept and attitude towards life. Negative affectivity can be predicted with measures on neuroticism and anxiety. People with high negative affectivity tend to feel stress regardless of the situation (Smither, 1994).

Also the personality characteristics, *need for achievement* and *impatience-irritability* are connected with higher feelings of stress. The need for achievement affects performance positively while impatience-irritability affects performance negatively. People with a high combined value on these dimensions are labeled "*Type A*" individuals (Brody & Ehrlichman, 1998; Smither, 1994).

Type A individuals are according to Smither (1994, p. 474) characterized by "a chronic sense of time urgency, a distaste for idleness, impatience with any thing or person they see as a barrier to accomplishing their goals, hostility, and competitiveness". *Type B* individuals are described as "easygoing, relaxed, satisfied, and unhurried. It is often very stressful for a Type As to work in a Type B environment and the opposite for Type Bs. Type As tend to live in densely populated industrial urban areas. Type A individuals are in higher risk of cardiovascular diseases (Brody & Ehrlichman, 1998; Smither, 1994).

Perception

Perception is the fundamental processes for bringing stimulus to our consciousness. Perception is the process where the neural signals or the raw data from the senses are processed and if necessary made conscious. It is an active, creative process generating and selecting among responses based on unique personal experience. Perception is our biased interpretation of reality, not factual reality (Kellogg, 1995; Marmaras & Kontogiannis, 2001; Smith, 1993).

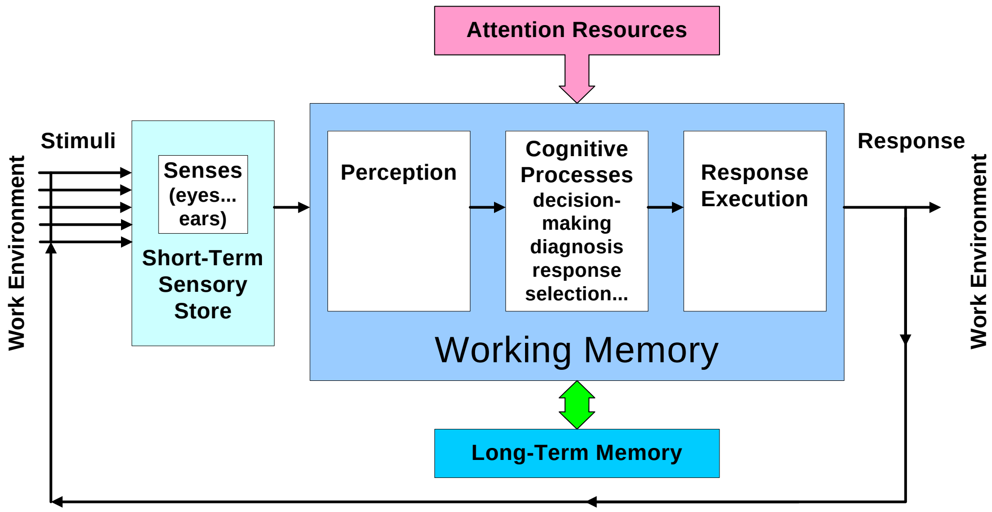


Figure 6. *Model of human information processing adapted from Wickens, 1992.*

We are constantly bombarded with enormous amounts of signals collected by our senses from our surroundings, see Figure 6. Each sensory system has a temporary storage, *short-term sensory store (STSS)*. It seems like the senses store information here independently of where attention is focused.

The STSS is large compared to the *working memory* but data decay fast. The next system processing this is called working memory and has *limited capacity* of between 5 and 9 *chunks* (groups of similar or connecting information) at full attention. The mental workload is also dependent of which mental functions are needed, if two tasks need the same functions performance decrease.

Finally a response is chosen for example to move the eye in a certain way. When a task becomes familiar through practice, it will now longer need the same amount of attention and cognitive resources, leading to increased speed and accuracy (Marmaras & Kontogiannis, 2001).

Attention is the main regulating mechanism, selecting and constraining which information that will be perceived and managing which tasks and mental operations that will be performed, see Figure 7. Features of stimuli that attract attention are: intensity, novelty, movement, contrast, and repetition (Marmaras & Kontogiannis, 2001; Wickens & Carswell, 1997). *Habitation* becomes a problem when attention needs to be selective and focused continuously, as when an operator performs an inspection task. The progressive decline in novelty of the stimuli from the task would lead the operator to attend less and less to the task itself and at the same time becoming easily distracted (Craig & Colquhoun, 1975).

Expectancy is a problem because expectations affect attention and perception (Marmaras & Kontogiannis, 2001; Wickens & Carswell, 1997). In Figure 8 will the black spot disappear at a certain distance, while the overall chequered pattern is visible; perceptual expectations have filled in the missing information (Smith, 1993).

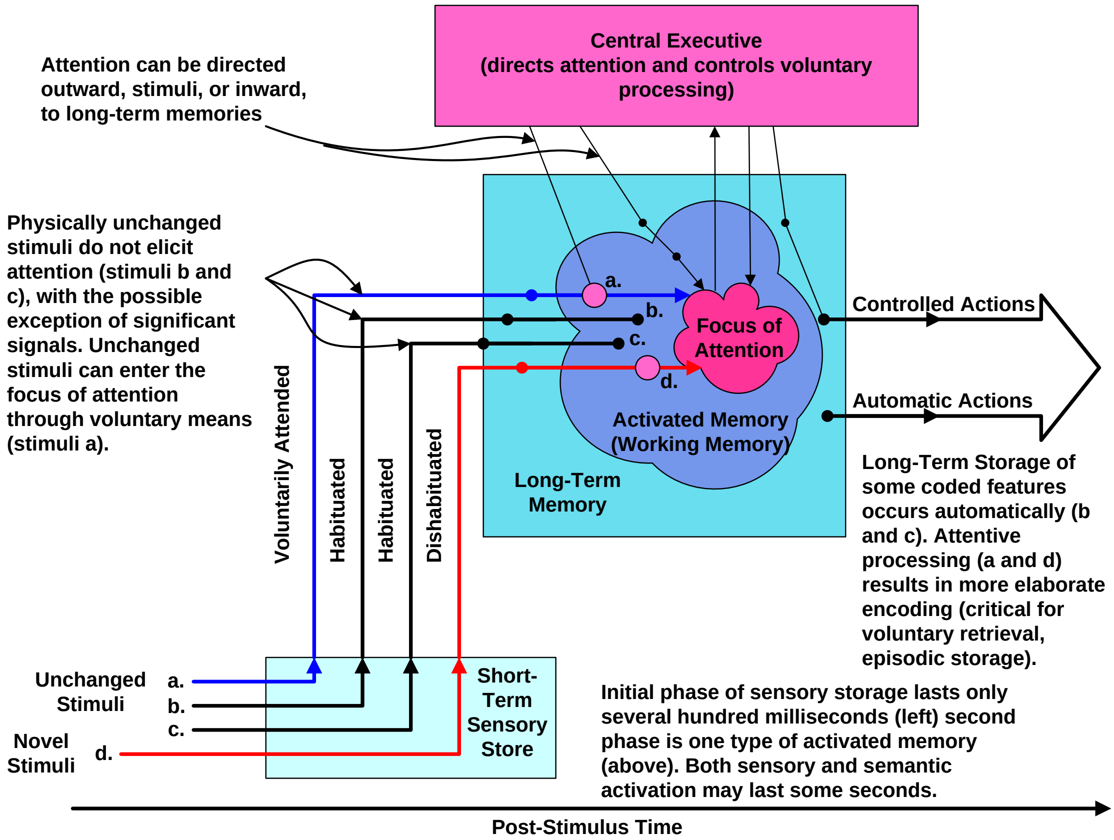


Figure 7. *Model of memory (sensory, short-term and long-term), attention and behavior (automatic and controlled). Based on Corwan (1988).´*

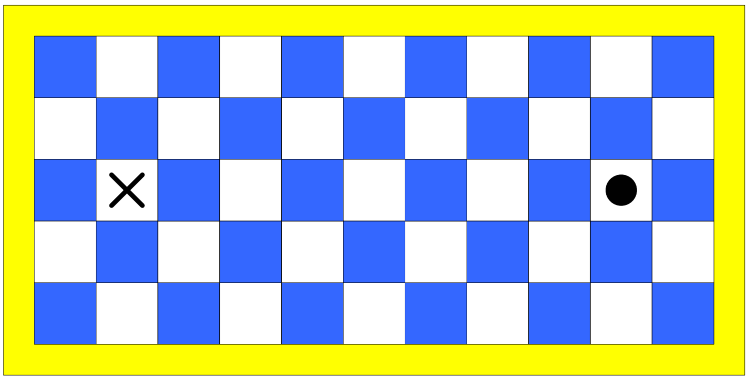


Figure 8. *Blind spot. Close your left eye. Focus on the X from a distance of about 40 cm. Keep focus while you slowly move closer to the figure. At a certain point will the black spot disappear because it crosses the optic disk or the blind spot. Note that you will continue to se the chequered pattern. Your perceptual system fills in the missing information. After a figure in Smith (1993).*

Other examples of *perceptual illusions* concerns lightness, brightness and cues in the environment, see Figure 9 and Figure 10. *Lightness contrast* is the phenomenon where the perceived lightness (achromatic color on a black to white dimension) of an object is influenced by its surroundings (Proctor & Proctor, 1997). Flour seems quite white when you spill it on your black pants, but almost brown when you add baking soda to a bowl of flour.

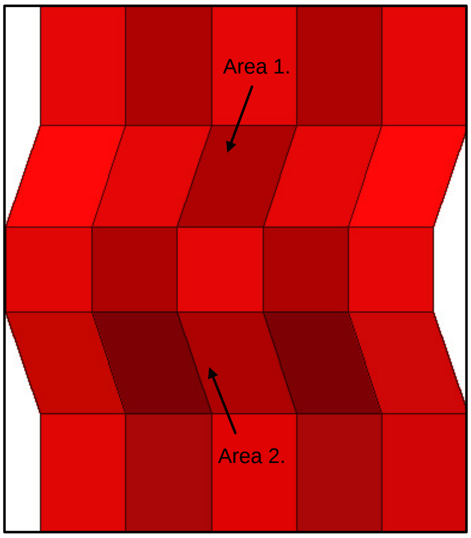


Figure 9. *Brightness illusion. Compare Area 1 with Area 2 which is brighter? After a figure in Proctor and Proctor (1997).*

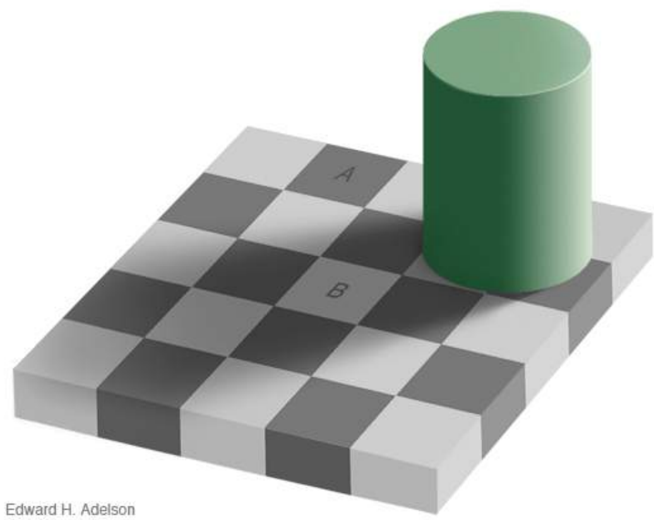


Figure 10. *Shading illusion. Compare square A and B, by courtesy of Adelson (1995).*

Figure 11 reveals the dramatic effects of depth and shading cues on brightness judgment. Square A and B are actually of the same color and brightness but shading and depth cues makes B appear very much brighter (Adelson, 1995). This is also why Area 1 in Figure 9 is perceived much darker than Area 2 even though Area 1 is of the same brightness and color as Area 2.

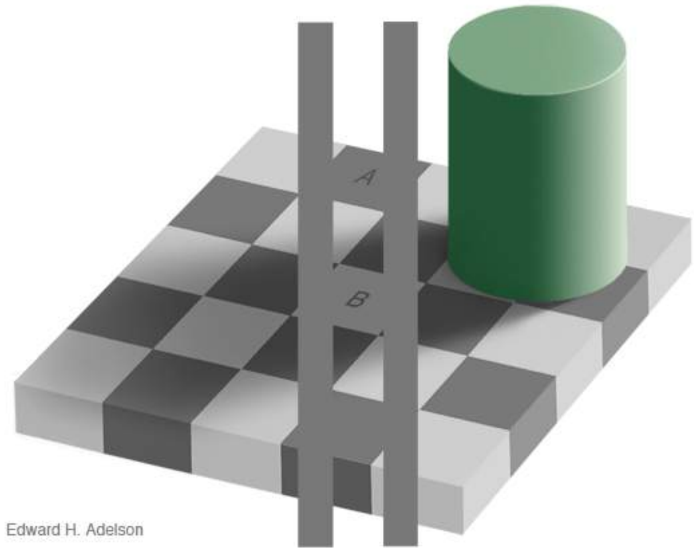


Figure 11. *Shading illusion proof, figure by courtesy of Adelson (1995).*

Lightness/brightness constancy is a phenomenon that makes us perceive that the lightness/brightness of objects remain the same under different conditions of illumination, this because the ratio of light intensity between an object and its surrounding is usually constant. The actual illumination does not matter as long as the illumination has the same intensity for both object and its surroundings (Proctor & Proctor, 1997; Smith, 1993).

The cycle of perception is a model for how perception works, see Figure 12. A *schema* is a general, hierarchical and mental representation of knowledge about the world. The "chair schema" contains legs, seat and back, but chair is also a component of the schema for room. There are numerous schemas activated and in standby at all times. These provide *expectations* of what will be encountered in the environment. These expectations direct our sampling of the world, as our eye movements. The information sampled from the environment either confirms, adjust or rejects our schemas (Kellogg, 1995; Smith, 1993).

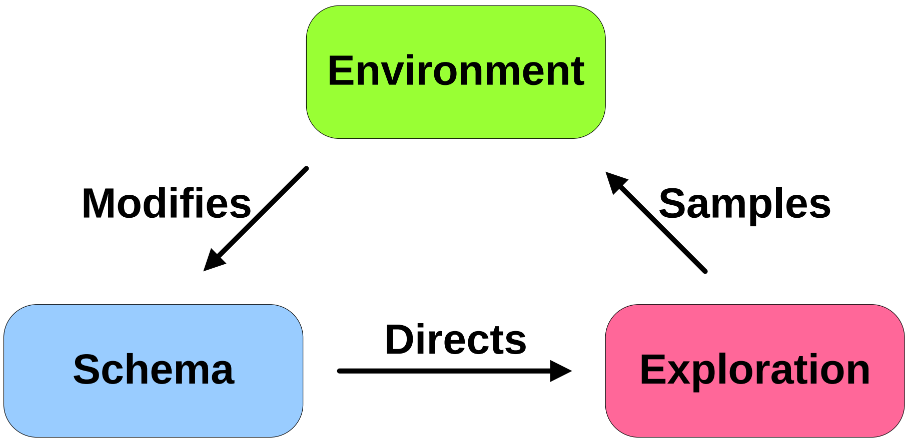


Figure 12. *The cycle of perception. After Kellogg (1995).*

When the sampling confirms what is expected the process works in a so called *top-down* or conceptually driven way. This reduces the needed amount of information sampling. An experienced operator tends to look most at the sources that change most frequently and so provides most information. However the operator has to balance the cost of not sampling. If I fail to look at a source what will it cost me? Simultaneously are edges, lines, areas of light and dark, colors, sounds and other physical features analyzed in a *bottom-up* or data-driven fashion in order to verify or reject the hypothesis and provide alternatives. If a source from the bottom-up process shall attract conscious attention it normally need to be salient and significant enough to compensate for the information access cost, the effort needed for further examination (Kellogg, 1995; Wickens & Carswell, 1997).

If observers are unaware of the real *signal rate* they will only have the knowledge of their own past experience. If faults are missed in the beginning operators may lower their expectations even more and so be less inclined to report faults. Research has shown that if inspectors are trained with a low *signal probability* they expect a low rate of signals later. This leads to a low detection rate even if signal rate is higher. If on the other hand a high rate is used during training the detection performance is much better, but only in the beginning and soon falls dramatically. The best performance is achieved by a correct signal rate. Even if this is used a vigilance decrement occurs that can not be explained with expectancy theory adding evidence for using the different theories in conjunction (Craig & Colquhoun, 1975).

Motivation

Motivation affects the willingness to find and report faults and it also affects attention (Marmaras & Kontogiannis, 2001; Wickens & Carswell, 1997). Humans are for example much faster at detecting an angry face in a crowd than a happy (Smith, 1993). One influential cognitive theory on motivation is the *expectancy-value theory*. Its basis is a very simple equation; behavioral potential equals the *expected probability* of an event multiplied with the *value or cost* of that event. By *behavioral potential* means the strength of the tendency to engage in the behavior relevant to the event (Brody & Ehrlichman, 1998). So if missing a fault leads to a high personal cost (danger, negative value), attention is more likely to be attracted and the correct behavior performed.

The level of motivation was above explained to depend on the perceived "value" of an event and the probability of that event occurring. Motivation differs not only in level but also type and orientation. One such type is *extrinsic motivation.* The illustration of the man and the carrot is a typical example of extrinsic motivation. Rewards can lead to resistance if the employees feel bribed and manipulated by the reward or absence of the reward. This can lead to impoverished workers that face work with resentment, resistance and disinterest. If monetary rewards are used frequently it can lead to the expectation of reward very time they are asked to do some thing new or unusual (Ryan & Deci, 2000; Yukl, 2002).

It is likely that rewards result in compliance rather than commitment. These types of rewards do not stimulate employees to go above and beyond what is expected or that the task demands. Instead is it quite possible that employees "cut corners" in order to achieve the goal (Yukl, 2002).

The view of extrinsic motivation so far doesn't capture the complexity and diverse ways in which people can be extrinsically motivated. Self-Determination Theory suggests that extrinsic motivation can be described as a range from amotivation to intrinsic motivation, see Figure 13. The following section is based on an article by Ryan and Deci (2000). The main variable is autonomy, or how self regulated the task is, this is similar with the demand control model.

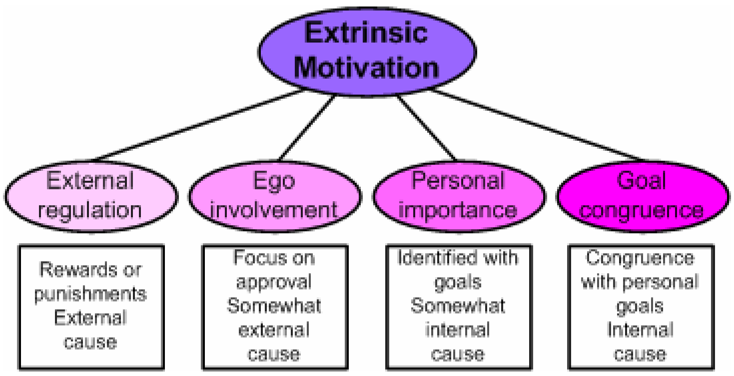


Figure 13. *A taxonomy of human extrinsic motivation, based on a figurein Ryan and Deci (2000).*

The first stage is where a behavior is executed in order to satisfy an external demand or to obtain an externally imposed reward. It is this type of extrinsic motivation that best fits the previously described. The second stage is more personal. The actions are still highly controlled but are performed to enhance self-esteem and pride and to avoid guilt and anxiety. When a person realizes the personal importance of the goals and accepted them he has taken the first step to stage three. It also requires that the person feels competence and has the required competence.

The final factor is a sense of belongingness to a group that valued by the person and to which the task is recognized as important. The final stage is when the belongingness and feelings of competence and most importantly the autonomy have increased to a level where the goals and tasks can't be separated from the personal needs, goals and values. The more autonomous extrinsic motivation has been found to be associated with greater engagement, better performance, higher persistence and higher quality of learning. Cross cultural research has proven the model in Japan.

Signal detection theory (SDT)

Signal Detection Theory (SDT) has had a tremendous impact on research on visual inspection. It was developed by Swets and originates from communication engineering and statistics. The strength of the theory is that it can explain misses and false-alarms, as well as providing a possibility to quantify different situational factors. The basic idea of the theory is that people always have a level of neural activity generated by *noise*. This noise can be both external, for example a person talking to you, or internal like thoughts. This noise varies over time. It is assumed that this noise is normally distributed from low to high. When a "*signal*" here a fault occurs its intensity is added to the intensity of the background noise. Figure 14 shows two hypothetical distributions of neural activity, one generated by noise the other by noise plus signal. The red line represents the *response criterion* which determines the probability of the different outcomes (Drury & Fox, 1975; Kellogg, 1995; Lehto, 1997; Sanders & McCormick, 1987; Wickens & Carswell, 1997).

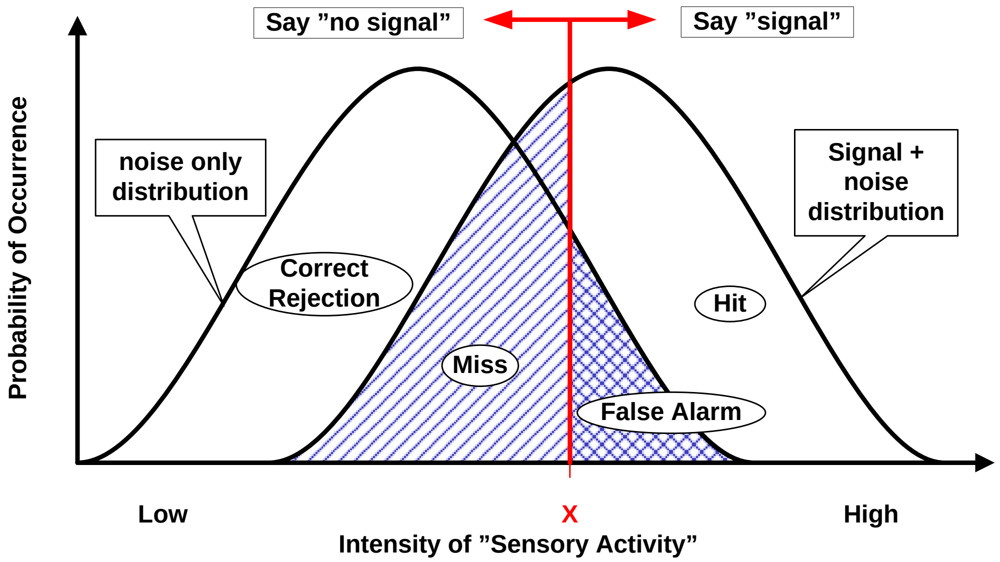


Figure 14. *Key concepts of SDT. Above is two hypothetical distributions of neural activity, one generated by noise the other by noise plus signal. The red line represents the response criterion which determines the probability of the different outcomes.*

The essence in the theory is that people chooses where to set their response criterion and because the two curves normally overlap this means that you can only get more *hits* by increasing the *false-alarms*. The ratio of signal to noise is labeled *beta*. If a person has a beta greater than 1.0 he is *conservative*, thus making few false-alarms but missing many faults. If a person instead is *risky*, meaning that the beta is low, he misses few faults but at the expense of making many false-alarms, see Figure 15 (Drury & Fox, 1975; Sanders & McCormick, 1987; Wickens & Carswell, 1997).

It has been shown that if the fault rate is increased the detection rate is also increased. The drawback is a higher rate of false-alarms. This is one of the most robust findings in the area of signal detection (Craig & Colquhoun, 1975).

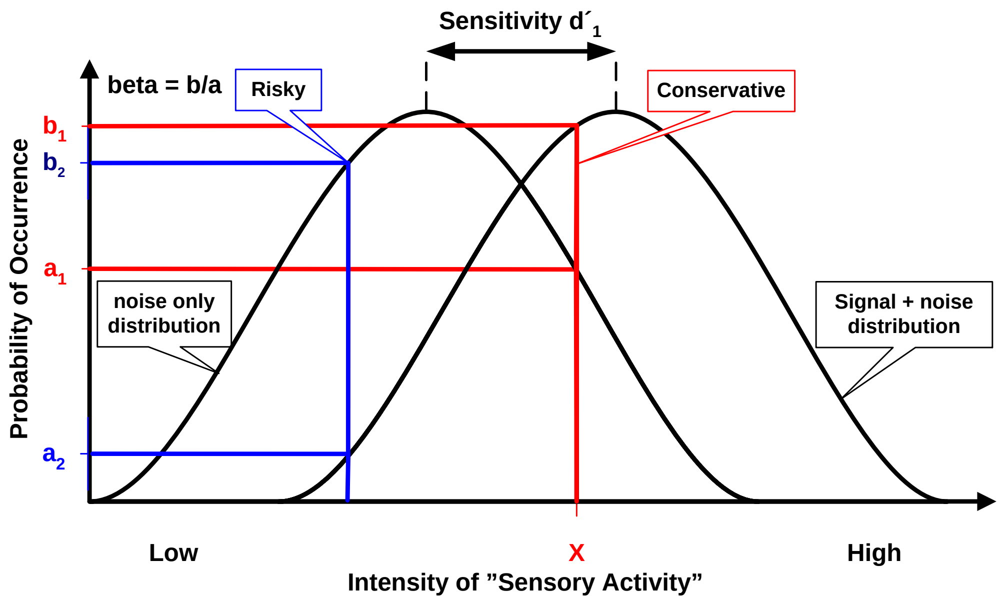


Figure 15. *Risky and conservative response criterion. Beta ß is a measure of the response bias and is calculated by dividing signal+noise with noise at the points where the two curves are intersected.*

Sensitivity or *d´* is the separation of the peaks of the two distributions. If the level of noise is high d´ decreases, so the signal can be conceal leading to a miss. If the signal instead is salient, compared to the level of noise, d´ will increases and so performance. The first task for the designer of an inspection system is to increase the separation of the two curves and so increasing the sensitivity of the system. When this is done by making the faults/signals appear more distinct, by for example combining both visual and auditory signals and using work aids as magnifiers; and by training the operators to better discriminate between faulty and correct units/signals (Adams, 1975; Drury & Fox, 1975; Poulton, 1970; Sanders & McCormick, 1987; Wickens & Carswell, 1997).

The response criterion or beta (β) is the next step. It is important to make sure that the operators set the optimal response criterion. There are two variables β: 1) the likelihood of observing a signal, 2) the cost and benefits of the different outcomes. It is important that the operators have knowledge about fault rates and costs/values of the possible outcomes so they can set a criterion that optimizes performance. Here are training and incentives important tools (Drury & Fox, 1975; Sanders & McCormick, 1987; Wickens & Carswell, 1997). Chi and Drury has proven that human operators are sensitive to and change their β dependent on rewards/costs and probabilities according to expectancy-value with the Bayesian probability model, see Appendix 1. The operators often fail adjust *β* far enough, a phenomenon called the "*sluggish β*".*(Chi & Drury, 1998)*

> "As a final reflection, one may ponder the extent to which it comforts the manager's mind to know that the declining detection efficiency of his inspectors is most probably due to a shift in their decision criteria, and that their real ability to detect faults has not changed at all!" (Craig & Colquhoun, 1975) p. 85)

Background of the study

The company

The company where this study was performed is in the electronics industry. They manufacture products with short life cycles; because of this it is very important to have a fast production ramp up, so they are able to deliver the right volume just as the demand for the product exists. When the study was performed they were using a 100% test and inspection strategy and relied heavily on advanced automatic test machines, called fixtures, but they also used human operators for testing and inspecting. There are big investments associated with these fixtures, both financial and in development; the short lifecycles of the products makes this a problem. The fixtures are also inflexible and only find the faults they are programmed to find. If the fixtures are not properly serviced performance drops. Because of these reasons the company wants to shift towards more manual tests/inspection to be able to take advantage of human flexibility and critical thinking, but to do that without any decline in quality.

The factory

The factory is located in East Asia and employs several thousand people. Most of them operators, but it also has an expanding staff of engineers and other white-collar employees. The factory has experienced a dramatic expansion in the last few years, both in production volume and number of employees. Also the changes in management and the organization of the factory have been dramatic.

The production is divided in to assembly and packing. This study was performed in the assembly section where two of the manual inspection/test stations were observed: MT1 (manual test one), MT2 (manual test two). This was done for one product.

The operators had three different forms of employment; regular employees with contract with the factory, but the majority were either provided by outsourcing companies or students from vocational school. These students were working a limited time period, primarily to get work experience.

There was no upper age limit for the operators but most operators were around twenty and all above eighteen. No study of the distribution between males and females were made, but there seemed to be slightly more female operators. But the distribution varied between lines and shifts.

Manual test one

Manual test one (MT1) consists of a fixture or a jig that holds the unit in place and handles the communication between the unit and the test computer. The test computer was an ordinary PC in a network with a monitor, a keyboard and a mouse. The work tasks[1] were:

1. Remove plastic cover from systems contact.
2. Place the unit in the fixture and fasten.
3. Push power-on on the unit.
4. Test key group 1 by pushing in an indicated sequence.
5. Test key group 2 by pushing in an indicated sequence.
6. Look in a mirror if LED lights up, push key for “Pass” or “Fail”.
7. Disconnect unit and place in corresponding tray, green “Pass” red “Fail”.

[1] Some tasks have been eliminated in order to conceal the identity of the product and company.

The keys are tested in groups by pushing a specified sequence that is analyzed by the test computer. The whole sequence is judged either “pass” or “fail”. A “fail” could be caused by one or more keys failing or that the operator pushed the wrong sequence. The intention is that the operator shall be “blind” to which key failed.

The operators have three trials for each key group, so they can make two slips per group before it truly fails the test. These fail messages due to operator slips were labeled *false-alarms*. It is important to emphasize that it was not the rate of false rejection that was measured, but these false-alarms or slips. The number of false-alarms was received by counting the fail messages and by keeping track if the unit finally passed or not. If it passed the number of fail messages was the same as the number of false-alarms.

When a unit fails a key test, for any reason faulty unit or slip pressing the wrong key, does a big red message appear on the monitor with the text “Fail” and there is a delay of 1-2 seconds before the operator can try again.

Manual test two

Manual test two (MT2) has the same basic equipment as MT1 with the exception of a vision aid that are to be folded down over the unit; on order to see if the display has an uniform luminance. The work tasks[2] were:

1. Remove plastic cover from systems contact.
2. Place the unit in the fixture and fasten.
3. Push power-on on the unit.
4. Fold down the vision aid.
5. Remove protective foil from display.
6. Inspect with aid if luminance is even, push key for “Pass” or “Fail”.

[2] Some tasks have been eliminated in order to conceal the identity of the product and company.

7. Inspect LCD for black spots on an all light background, Press the corresponding key for “Pass” or “Fail”.
8. Inspect LCD for white spots on an all dark background, Press the corresponding key for “Pass” or “Fail”.
9. Disconnect unit and place in corresponding tray, green “Pass” red “Fail”.

In MT1 it was the test computer that made most decisions whether a unit passed or not, in MT2 it was the operator that made the decisions and the computer that activated the different features.

Air bubbles were often trapped under the LCD protective foil. These were often much bigger than the defective pixels 0.025 $mm^2$and trapped dust the operators were looking for, so it was important that it was removed properly. A defective pixel equals a visual angle of 0.0036 degrees at a viewing distance of 400 mm.

The purpose

This research paper concerns as previously stated “why” faults were/are missed in the two test/inspection stations in the factory’s production: MT1 (manual test one), MT2 (manual test two). The performance and demands on the operators/inspectors have been audited by Lindblad (2005). That study showed that many faults were/are missed. These results led to this study of why this many faults were/are missed?

Results on the demands and conditions in the factory

The factory imposes different demands on the operator: technical, economical, physical and psychological; but also the human have demands that have to be met. Do these different demands work together or collide?

Lindblad (2005) studied what the demands were on the operators. The results are presented here in Table 2, Table 3 and Figure 16. The results in Table 2 shows that the work in test stations MT1 and especially MT2 are high on the dimension psychological demands, in the demand-control model.

Table 2.
Psychological demands on the operators in MT1 and MT2, Lindblad (2005).

Excessive work	No, but observations suggests that volume goals was frequently not met
Conflicting demands	Very much so, because the inspectors lower output
Insufficient time to work	Self paced but indirectly controlled by the station after
Work fast	Short work cycles, 9- 22 s
Work hard	Dependent on line status
Intense concentration	MT1 not so much, MT2 much due to small and infrequent faults, 0.025 mm^2
Often interrupted	System and workflow can be unstable
Very hectic	Yes, especially to compensate for a system breakdowns
Waiting on others	Frequent

Figure 16 shows the significantly decreasing test times over the work shift. This increased work tempo was often the result of attempts to compensate for the accumulated systems/ equipment failures.

Table 3 shows the very limited decision latitude in the two stations. Here MT2 is slightly better. The combined effect of these dimensions mean that the work in test stations MT1 and MT2 are in the direction of high strain, in the demand-control model.

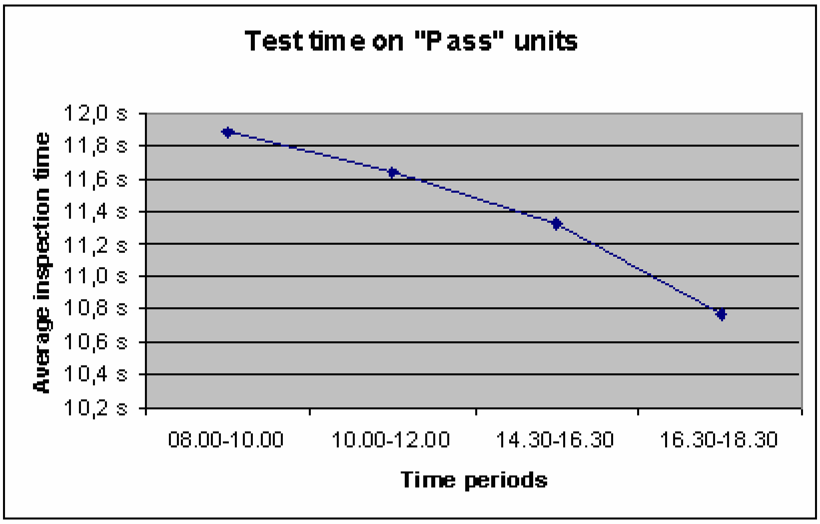


Figure 16. *Changes in test time over the work shifts.*

In spite of these results the workers seem to have good morale. The line leaders appeared to be supportive. There were big differences in how much the operators talked, how relaxed they worked and how well they followed the rules in the different shifts and lines. It was said that this often depended on the line leader.

Table 3.
Decision latitude experienced by the operators in MT1 and MT2, Lindblad (2005).

Have freedom to make decisions	Only "pass" or "fail" and to retest. In MT2 the operators make most decisions if the unit passes or not. In MT1 it is mostly the computer.
Choose how to perform work	No
Have a lot of say on the job	Limited
Keep learning new things	Very limited
Can develop skills	Very limited
Job requires skills	Basically most dexterity (MT1) or visual search (MT2) and some computer skills
Task variety	Nonexistent
Repetitious	Very much so
Job requires creativity	Nonexistent

Research hypothesis

Certain recurring themes surfaced when the literature on: inspection, stress, motivation, signal detection and human error were reviewed. Those were: the vigilance decrement and decline in attention due to habitation and the need for stimulation, stress and expectancy and how motivation is affected by stress, lack of autonomy, expectancy, feed-back, rewards and costs. The purpose of the study was to find evidence of the vigilance decrement and possible influencing factors, as well as to find factors affecting motivation, search criterion and expectancy.

The statistically testable quantitative hypotheses used were:

- The performances of the operators will decline over the work shift, in terms of more execution slips (false-alarms) in MT1 and fewer found faults in MT2.
- Performance is affected negatively by increases in demands/stress. Performance in MT1 will be worse when the test times are low and/or the tested volume is high.

Method

This research was preformed as a case study with an explorative direction in a real world setting. This method was also chosen because it is extremely expensive and complicated to do experiments in the production lines or to design and perform laboratory tests with sufficient degree of ecological validity.

The goal was to describe and increase the understanding of the situation and problems in the inspection system. The research field was huge and each inspection system is different, it therefore needed a basic investigation with the aim to give insight in to the problem areas and the related questions, so various hypotheses could be formulated. This led to the use of both qualitative and quantitative methods. Qualitative data on the performance and motivation of the operators was collected with semi-structured interviews and semi-participative observations.

The quantitative data that are most relevant to the study of inspection work are missed/found faults and false-alarms. Because this was not a study in a laboratory but in a factory running at full pace these measures were not possible. There was no way to know how many faults were in the batches and how many that passes uncaught. Because the fault rate in especially MT1 is dependent on how well the unit was manually assembled fault rate could change during a shift. This is why SDT is not used mathematically but instead only as a concept.

Of these reasons were it decided to use false-alarms, note that false-alarms do not stand for false rejection as it normally does, but for execution slips see page 30-31. False-alarms is an indirect measure of performance as it is here used as an indicator of general performance. It is assumed that if an operator makes more false-alarms (slips) they will also make more other mistakes, this may however not be the case. Data was collected in a semi-participant observation with structured observational measurements that used a repeated measures design. These results were later combined with the results on test times and production volume from (Lindblad, 2005) in order to test correlations.

Demarcations

As just mentioned is this area of research huge, so some demarcations must be made. First, the thesis focuses on why performance decline and what factors influence expectancies and motivation. Suggestions on how to improve the design of tasks, goals and rewards and other areas as training, the needed instructions and cultural differences are handled in Lindblad (2005). The second was to only look at two test stations in final test: MT1 (manual test one), MT2 (manual test two). Third, was to study one product. The product chosen had been on the market since the summer 2004. It was chosen because it was new enough to be representative and interesting to the current work of the factory and old enough for production to be stabile and for a sufficient amount of data to be accumulated.

Because this is a case study it is not possible to make predictions about the future. The study can only describe the history and generate qualified guesses about why certain results were received.

Information and data collection

The information and data collection was divided in to three phases. First a literature investigation was performed to give an insight to the problems. Then secondary data was collected from the company. Finally a job analysis was performed in the factory. The advantage of using literature to find out answers to problems compared to doing experiments is to avoid having to "reinvent the wheel"; this saved resources and made it possible to give the project a wider scope.

The literature investigation was made by extensively searching the databases: Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI) and PsycINFO. For descriptions see Appendix 3. First keywords and authors were used, and then the thesaurus functions to find more keywords. By combining these keywords and authors in different ways a promising list was generated. The next step was to find what other sources these articles and books themselves referred to. The literature investigation continued throughout the project and thus generated an ample body of literature.

Secondary data was collected onsite in the company's headquarters, by reading the different process and work descriptions and field rapports from production and by interviewing employees at different positions within the company. This gave an initial picture of the demands and problem areas.

The collection of primary data was performed in the company's factory in East Asia. The job analysis lasted for eight days. During that time was both observations and interviews performed. Persons that were interviewed were the director of engineering and the director of quality & OD, around twenty engineers from different departments as well as about ten people in human resources and operators/inspectors.

Observation design

Operator work performance was observed and measured by sitting behind the operators while they performed their normal tasks. The operators working on the stations MT1 and MT2 were sitting next to each other about half a meter apart in the line. Observations were made from the position shown in Figure 17.

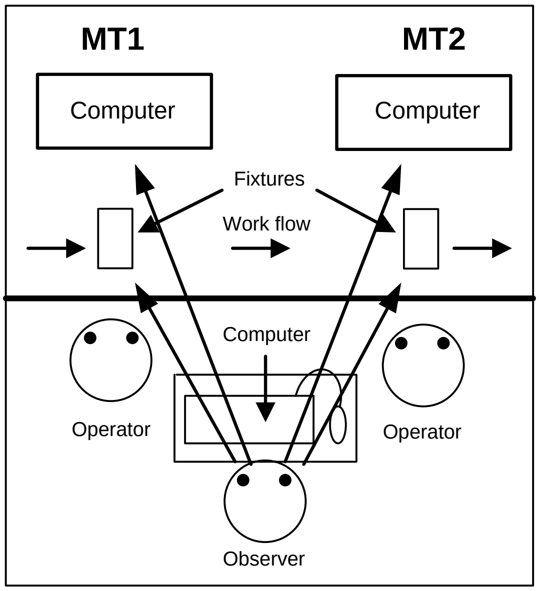


Figure 17. *Schematic layout of the observations.*

The shift lasted from 8.00 to 20.00, with two one hour breaks. For convenience was only day shifts observed. In order to see if the operators' performance changed over time were they observed 4 times during a shift. 10 operators were observed for 20 minutes during each of the four time periods. The time periods were: 08.00-10.00, 10.00-12.00, 14.30-16.30 and 16.30-18.30. That means that some operators were observed between 8.00 and 8.20 while some between 9.40 and 10.00 in the first time period. All observations within one period were assumed to be equal in terms of time on task. In order to have equally long time between observations were the observer first studied also studied first in the other time periods, and the last were correspondingly last in all.

Two of the observations collided though with early lunch. These were instead observed before time period 3. This gave an interesting opportunity to see if lunch affected results.

Before the real observation was a pretest performed. This was done the day before and on a shift not included in the real observations in order to spot observer effects and to try out and adapt the equipment and procedures.

Equipment

Quantitative data was collected on number of tested units, found faults and fase-alarms; see next section for a more thorough explanation. The measurements were collected in production with a laptop computer, an optical mouse and a specially designed test program in Microsoft Office Excel 2003, see Appendix 2. Each subtest started when the unit was placed in the fixture and stopped when removed. The program also included an area to the left as can be seen in the appendix, where qualitative notes could be written down. The statistical calculations were performed in SPSS 11.5. During the interviews was a dictaphone used as a memory aid.

Subjects

The subjects were the normal operators in the factory, but because of the limited number of operators and changes in research design could not equal groups be formed, see Table 4. The operators in shift 1 was randomly selected but in a way so the number of male and female were the same. After a decision to stop observing MT2 were all operators in shift 2 observed. That shift had only one female at MT1. The operators were supposed to man one station per shift, but two females changed position in shift 1. This was noticed too late causing internal loss. All operators were of a similar age ranging from 18 to 23 years old.

Table 4.
The distribution of the observed operators.

Station	Shift 1	Shift 2	Total
MT1	4	6	10
Female	2	1	3
Male	2	5	7
MT2	4	0	4
Female	2	0	2
Male	2	0	2

The observer effect was dramatic when entering the factory for the first time. All conversations and laughter stopped in the entire factory. In order to tackle this extreme reactivity effect, were various adaptations by the observer made in order to see how the operators work normally, when relaxed. They were: stop shaving, dress down, be overly relaxed, look lost, make jokes, try to speak their language and use a lot of body language. It was also expressed by use of vivid body language that the interest was in the system/computer, not them so they should relax. It is uncertain if this was believed. These actions worked very well though and it didn't take long for the operators to relax and work "normally". It worked so well that one operator actually fell asleep within a meter from the observer during one observation!

Never the less was performance affected by demand characteristics. Operators joked and laughed but they clearly wanted to be "good". It was decided that this should be disregarded because the research is about decline in performance and not the level of performance. There is risk though that there was an adjustment effect so that the operators performed at max in the beginning because of the observers presence but that they got more and more used to being observed and so relaxed and let performance drop, or that they could not keep performance at these eventual max level for an entire work shift.

Ethics

The actions mentioned in the previous sections raises some ethical questions. The operators were neither given the true/complete reason for the observation nor were they given the choice to participate or not. This was unfortunately necessary to capture realistic work performance. The limited number of operators and the great language barriers were further reasons. The operators were on the other hand not required to do anything different from their normal work tasks. In order to protect individual operators were all data recorded anonymous and handled in a way that they could not be identified. The concern of the company was process/system issues. The responsible person for this project was clear in emphasizing that poor operator performance is a symptom of system/procedure problems and not the cause.

Results

The results are divided in to quantitative and qualitative results. In this section will the quantitative results on how the rate of false-alarms (mistakes) changed over the work shift be presented. After this are the correlations explored between false-alarms and the two measures of stress demands; number of units tested and test time.

Rate of false-alarms in MT1

The repeated measures ANOVA was very close to show a main effect between periods (F (3.24) = 2.83, p = .060, $\eta 2$ = .26). Table 6 shows that the results were also very near to a linear trend (F (1.8) = 4.88, p = .058, $\eta 2$ = .38). The data had to be transformed by using the square root to avoid sphericity and to improve normality.

Table 5.
ANOVA - Tests of within-subjects effects for false-alarms in MT1, very close to a main effect between time periods.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
SQRT(F.A.)	Sphericity Assumed	3	,027	2,828	,060	,261	,604
Error(SQRT(F.A.))	Sphericity Assumed	24	,010				

a. Computed using alpha = ,05

Table 6.
ANOVA - Tests of within-subjects contrast for false-alarms in MT1 showing an almost linear trend.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	F	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
SQRT(F.A.)	Linear	1	,081	4,878	,058	,379	,493
Error(SQRT(F.A.))	Linear	8	,017				

a. Computed using alpha = ,05

Figure 18 shows a very linear looking curve that conceals the fact that there was a high standard deviation. The mean was 20.7 % false-alarms with a standard deviation of 12.6 %. Three of the operators actually improved performance over time. Individual differences were big, see Appendix 4.

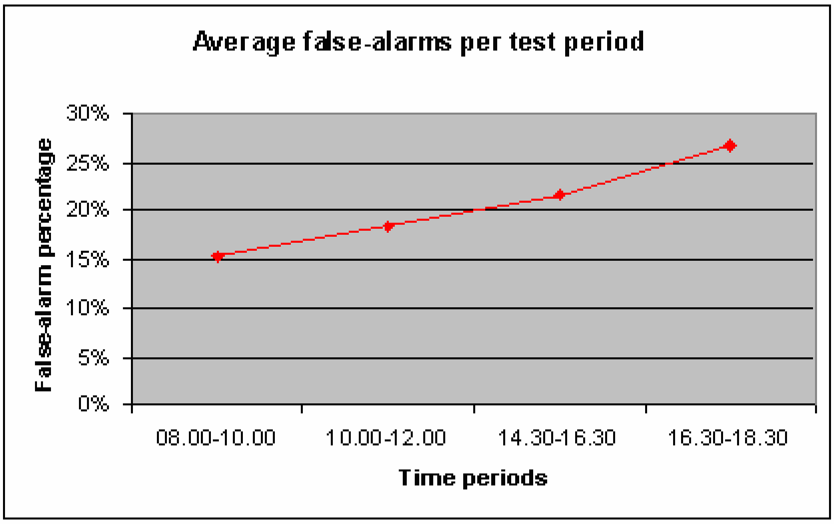


Figure 18. *Changes in false-alarms over the work shifts.*

Table 7.
Paired samples t-test for false-alarms in MT1 showing significant results between periods 1-3 and 2-3.

Paired Samples Test

		Paired Differences					t	df	Sig. (1-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	F.A. 1 - F.A. 2	-,02670	,057061	,018044	-,06752	,01412	-1,480	9	,087
Pair 2	F.A. 1 - F.A. 3	-,06456	,078306	,026102	-,12475	-,00436	-2,473	8	,020*
Pair 3	F.A. 1 - F.A. 4	-,09870	,171919	,054365	-,22168	,02428	-1,815	9	,052
Pair 4	F.A. 2 - F.A. 3	-,03422	,053221	,017740	-,07513	,00669	-1,929	8	,045*
Pair 5	F.A. 2 - F.A. 4	-,07200	,179303	,056701	-,20027	,05627	-1,270	9	,118
Pair 6	F.A. 3 - F.A. 4	-,04744	,188859	,062953	-,19261	,09773	-,754	8	,237

When the repeated measures ANOVA failed to show significant results were paired samples t-test used, see Table 7. These gave significant differences (1-tailed) between periods 1-3 ($p < .05$) and between periods 2-3 ($p < .05$) as well as a very near difference between periods 1-4 ($p = .052$). Performance decreased (false-alarms increased) over the work shift as predicted confirming the research hypothesis.

Demands and stress in MT1

Stress in terms of work demands in MT1 can be measured by volume, number of tested units; and work tempo, the average time it takes to test a fault free and false-alarm free unit. Results on these variables in Lindblad (2005) were combined with the measurements on false-alarms to test if these variables were connected as predicted. Data was transformed when needed in order to use parametric tests. Figure 19 shows the mean value changes per time period for false-alarms, number of tested units and test time. The test time was reduced with 9.4%, 18% more units were tested in spite of production problems and 62% more false-alarms were made between the first period and the last.

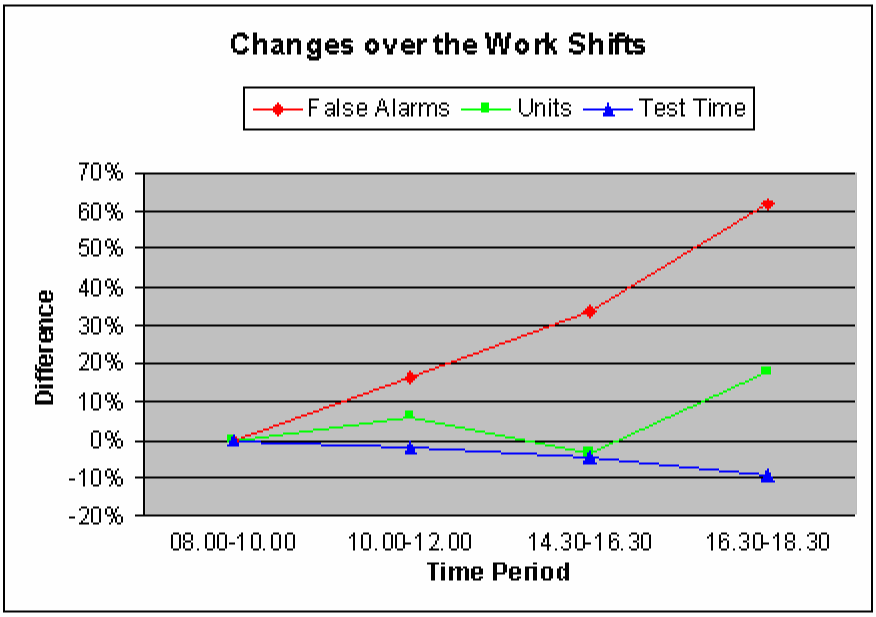


Figure 19. *Performance changes over the work shifts.*

Test time had a much skewed distribution not even the inverse transformation was sufficient to make it distributed normally, so the non parametric Spearman test had to be used. A negative correlation between false-alarms and test time was found

($r_s = -.48$, $p < .01$, 1-tailed, the table shows p = .001 not p < .001 explaining where there are two and not three stars), see Table 8. This confirms the hypothesis that shortening the test time will produce more false-alarms (mistakes).

No correlation was however found between number of tested units and false-alarms or between test time and tested units. The last seems odd and is probably due to the production problems that occurred during the observations, because the two variables time and volume are obviously connected. If the tests are performed faster more units can be tested. The general impression to the observer was that work tempo increased as Figure 16 previously showed. The Because of the limited number on female operators studied were no differences due to sex studied.

Table 8.
Correlations between test time, false-alarms and tested units in MT1.

Correlations

			TEST TIME	FALSE ALARMS	UNITS
Spearman's rho	TEST TIME	Correlation Coefficient	1,000	-,484**	-,149
		Sig. (1-tailed)	.	,001	,189
		N	39	38	37
	FALSE ALARMS	Correlation Coefficient	-,484**	1,000	,171
		Sig. (1-tailed)	,001	.	,155
		N	38	39	37
	UNITS	Correlation Coefficient	-,149	,171	1,000
		Sig. (1-tailed)	,189	,155	.
		N	37	37	38

**. Correlation is significant at the 0.01 level (1-tailed).

Qualitative result

So far has only results from the test station MT1 been presented. That is because an interesting effect was encountered by the observer.

Habituation in MT2

After observing MT2 for a day and a half something interesting happened. The observer caught himself drifting away several times. He was either not looking at the operator he was supposed to observe or he was looking straight through him/her thinking of other things. He forgot to start and end tests, and he had only worked with observing MT2 one and a half days and not for several months.

There was tremendous pressure on the observer to perform well and he was very motivated to do so. The work of the observer was as simple as the operators but he had a lot more control over the work and he had breaks or variation every 20 minutes. In spite of this was it impossible for the observer to keep focus, because so little happened in MT2. The observer described it as the body had learnt that nothing interesting ever takes place in MT2 and so used its resources on other things against the observer’s conscious will. This effect is though expected from the habituation and arousal models described earlier.

This is very important because it explains why the operators often didn’t look at the work piece they were paid to look at, even though they so obviously were observed. This finding makes it hard to criticize the low paid highly stressed, highly controlled operators for not “doing” their job and carefully study each work piece, when the observer that were in a much better position couldn’t keep attention focused for two days. These lapses by the observer in combination with the low fault rate led to the decision to stop observing MT2. Because of the scarce faults were the results highly sensitive to random variations, threatening validity of the results. The frequent lapses made the results highly unreliable. In MT1 were there fail messages on 20% of the units that provided stimulation that abolished the habituation effect.

Motivation

The goals and feedback are suboptimal. There are different and conflicting goals for production: volume goals and yield goals (how many of the tested that passes) and quality goals. These are updated regularly. Volume and yield goals are provided and evaluated every two hours and are highly visible for the operators. Quality feedback is provided on samplings test done the previous day.

The operators have a bonus system divided in two parts, one part if volume goals are met the other if less than a target fault rate is found in the sampling tests. This is very good, but there seems to be problems in production to understand what the priority is, volume or quality because of the frequent changes. There is also a high tacit pressure on the operators to approve units because the goal of their colleagues is primarily volume or yield.

The operators in MT1 and MT2 can open a window on their test station computers to get continuous feedback. This is a very useful tool but it directs behavior in the wrong direction in its current form, this because it focuses on at volume and yield. The task of the quality inspector is to lower the yield by finding faults, put roughly. These conflicting goals are a major stressor to the operators. The information provided is on the last 100 units tested: number tested, passed, failed, min-max-ave. test time as well as total units tested. The problem is that if you haven't found any faults the value 100% is shown in big numbers on a bright green background, but when the operator finds a fault the value is lowered.

Operator behavior was obviously affected by these goals and choices of feedback and this was probably the reasons why some operators tampered with failed units so they sometimes passed the second time. It was observed that some of the operators in MT1 had found a way to get around the “blind” test of the keys, by opening a program and readjusting and arranging the different windows on the screen. By doing that is it possible to see each command row and thus which key that were problematic and so sometimes could pass the test if it was pressed harder. It was also observed for other fault types that the operators squeezed, hit the units and/or sent the units to another station in order to get the units to work, which they sometimes did.

Another possible way to tamper with the units and collected information is to restart the station when a faulty unit is found. All current test data is lost if the system is restarted, so it is possible that some operators restart the system to conceal how many faulty units that actually exists. Lindblad (2005) showed that restarts often occurred in connection with faulty units. If a unit was faulty it was 22 times more probable to get a restart compared to a unit free of faults. A parametric test gave a correlation value of $r_s = .423^{**}$ (Spearman).

The number of restarts could and were at least to some extent affected by the implementation of a new computer system and could also be affected by the start up of the factory after a big holiday, though many of the operators seemed quite familiar with restarts. It is important to point out that this wasn’t the only reason for restarts, many happened for no apparent reason. There are some ideas that operators can cause system crashes by being too fast, but Lindblad (2005) found no correlation between test time and restarts. It is important to state that the amount of collected and analyzed data is limited, so the reliability is sensitive to random variations. Also the validity of restarts as an indicator of operator motives is in question due to the fact that it has not been proven whether it actually is the faulty units that cause restarts or the operators.

Discussion

In order to understand inspection performance in the real world one has to use several models and theories in combination, this because a model per definition is a simplified version of reality and inspection in reality is not simple! This case study has showed that a significant vigilance decrement occurred over the work shift and that stress in terms of shortened test times and work backlog affected performance negatively. There were also observations and experiences by the observer that suggest that attention was impaired by habituation, expectancy and motivation. The high standard deviations and the very different individual performance curves confirmed that this is an area which is very much affected by personal differences.

Goals and feedback

These are main problems. The fact that the operators put so much effort in to making units pass, by tampering with the units and test station, maybe even restarting the station to conceal faulty units, is a strong indicator that the wrong goals are set from a quality perspective. This is further strengthened by the observation that operators often did not even inspect the units, this finding is however largely due to the habituation effect. The feedback was focused on volume and yield and so further increased the problem. These are strong evidence that rewards/costs, feedback and goals are set in a way that affects the search criterion in a conservative direction and so leads to missed faults.

This means that it is essential to carefully contemplate which information is provided to the operators and the way it is provided so it guides behavior in the right direction. Training and an increased vertical communication is fundamental to control and communicate that the aim is for the right goals. It is also very important for feedback for both operators and management.

The need for stimulation

The low fault rate and intensity of the fault signals did appear to be big problems, especially for MT2. The low fault rate is likely the reason why the operators often failed to keep attention focused or focused it in the wrong places due to expectancy. The low fault rate is also a likely cause to lowered arousal and performance.

Operators were not taught about the expected fault rates, which they ought to be in order to set an optimal search criterion. The fault rate was not a so big problem for the keys test in MT1, because that station has a high rate of events basically the false alarms (23% of the units) that provide stimulation and the test computer did most of the decisions. There was however also a LED inspection test performed in MT1 was it as bad as for MT2. Because as just stated there was a sufficient level of stimulation in MT1 the poor performance on the LED inspection seems to be more linked to expectancy. That highlights the need for faults or apparent faults to occur at some minimum level.

These results signify the need for further research on ways to provide stimulation. Stimulation can be both task specific, as how often there are event out of the normal for example faults and false-alarms and general as music that affects the overall stimulation level. An interesting study would be of the effects of music on inspection performance? Another interesting study would be how high level of faults that is required for a high detection performance?

Stress

Stress in terms of increased demands increased as a result of trying to compensate system failures and this had a significant correlation to decreased performance. Possible ways to improve performance is to have a system that regulates so inspection for the different areas take a minimum time, (some tests already have this). Another way is to insert virtual faults and measuring performance on how well the operators find these, in an attempt to ensure that the operators actually inspect the workpiece. This however further decrease operator control and increase the demands on the operator, on the other hand can virtual faults be a source for fast, relevant and accurate feedback and so probably increase their motivation as well as increasing stimulation. This is a major area for further research.

It was found in the job analysis in Lindblad (2005) that the operators often worked faster and some much harder in order to be faster than they needed to be, this had surprised many visitors from Europe. The behavior could be explained with the demand-control model and the effort-distress model. By increasing the effort can possibly a sense of control be achieved, which according to the research lower the level of corticosteroids. This in turn leads to decreased feelings of boredom, depression, dissatisfaction, helplessness, passivity, uncertainty and anxiety.

Also the personality trait need for achievement and the competitiveness of Type A individuals can explain why the operators often worked faster than they needed to. One can only speculate in which “Type” the workers are but as previously stated are Type As more frequent in industrial urban areas. The factory is placed in one of these areas, and it is often the need for achievement that makes people move to these areas.

Attention

When studying the focus of attention for the operators, it was found that they spend most of the time watching the computer screen instead of the workpieces. Theories on expectancy and top-down processing suggest that this is because it is the computer screen that gives the most information per time unit and the information that the operator is most dependent on. In other words, and observed in the factory, it is more common that the computer system misses a key being pressed, losses connection with the tested unit, hang or some other problem, than there is faults present in the units. One conclusion of this is that it is the computer that threatens the operator from reaching the goals, volume, and yield not faulty units.

Individual differences in arousal and hunger

Individual differences could be possible explanations for the very different patterns of performance seen in MT1, as between operator MT1op3 and MT1op8 see Figure 1 and Figure 2 in Appendix 4. MT1op3 was very open and eager to communicate, signs of an extrovert personality. He showed clear signs of under-arousal by yawning and he actually fell asleep several time while being observed from a distance less than one meter. After period three there was a break leading to a possible vigilance recovery, the work tempo increased a lot leading to more stress (stimulation) and he had only about an hour left to work possibly giving him a clear goal. This could be the explanations of the dramatic improvement.

MT1op8 showed a completely different pattern. It is possible that the increased stress at the end of the work shift made him over aroused. It could also be that he was hungry and tired or a combination of the different factors. He was much quieter and didn't appear to be as extrovert as MT1op3.

Both MT1op8 and MT1op2 were males and 18 years old and so probably in a phase were they are frequently hungry. They were also the only ones that had lunch before measure 2 and both show improved performance, see Figure 3 and Figure 4 in Appendix 4. It would be interesting with further research to see if the improvement was because of food or the break? It would also be interesting further research to find out if the drastic decline later on the shift was due to the increased stress because of production problems leading to over-arousal or hunger or some other factor.

Operators MT1op5 23 years old and MT1op7 19 years old are the two with the most dramatic increase in false-alarms, see Figure 4 and Figure 5 in appendix 4. It is interesting to note that the increase is between periods 1-2 and 3-4 and so in between meals. If performance declined as a function of time then the biggest decline would be between period 2 and 3 because the time difference was 4 hours between periods 2 and 3 compared with 2 hours between the others.

The individual differences found are other interesting fields for further research. Could performance be predicted by testing personality traits? The problem with this, apart from accurately measuring individuality and the associated costs, are that the conditions in the line change, so it is probable that a "low stress" individual perform worse than a high stress individual in normal work conditions and only better at high stress conditions.

It would also be interesting to study if the improved performance between period 1 and 2 for MT1op2 and MT1op8 was because of food or the break itself. The operators MT1op2, MT1op5, MT1op7 and MT1op8 all show a drastic decline in performance between period 3 and 4. How much of this decline was due to increased stress, fatigue and possibly hunger? These operators were all male could there be a difference between male female in how meals affect performance? The factory provides two meals per work shift. The portions are ample with a healthy mix of different selections of meats and vegetables and a fruit. It would however be interesting to see if a mini break for a fruit or sandwich could have a positive effect.

Lindblad (2005) found that males had a significantly lower test time mean value in MT1 11.4 s compared with 13.0 s for females. If this result is combined with the result here that males make significantly more mistakes (slips), it suggests that males should be placed in MT1, were many of the slips are caught by the system and females in MT2, were it is important to be careful. But this ought to be confirmed with further research. It is also dangerous because it signals that there is a difference between males and females in a factory that otherwise has a very equal distribution of males and females in the different departments and positions.

CONCLUSION

Performance declined over the work shifts as predicted. The reason however is however multi-faceted and the individual differences are big. There are two main areas: motivation and physical/psychological mechanisms. Motivation is highly dependent on goals, feedback, social support, the information provided and how that is interpreted. Because of the crucial influence of motivation on behavior is it imperative that the department designing the tests also have control over which goals are set and which feedback that is provided to the operators.

The physical/psychological mechanisms such as habituation, expectancy and perception, arousal and the vigilance decline are more difficult influence directly. The task is instead to avoid/minimize their negative effects. It is important to increase variation as rotating between stations, provide adjustable a level of stimulation as music and sufficient periods for recovery. To inflated fault rate by inserting virtual faults can help to counter habituation and expectancy effects.

It is important to keep in mind that stress is expensive to the company. It increases turnover and so the expenses for training and fluctuating production performance. Health is also affected and the performance as well as the section about presenteeism showed. One vital aspect of stress is the control component, this must be taken in to account by the test designers.

The operators are the “customers” of the test system and as with any product it is imperative to design with the customer in focus. These “customers” will make mistakes and their eyes will occasionally lie! So if quality is paramount then the goal is to have a test system with risky decision criteria and a “safe net” catching the increased number of false-alarms. It is also important to provide accurate feedforward information on expected fault rates. Humans will always make errors the goal is to minimize the amount!

REFERENCES

Adams, S. K. (1975). Decision making in quality control: Some perceptual and behavioural considerations. In C. G. Drury & J. G. Fox (Eds.), *Human reliability in quality control* (pp. 55-69). London: Taylor & Francis Ltd.

Adelson, E. H. (1995). Checker shadow illusion. Retrieved 31 May, 2005, from http://web.mit.edu/persci/people/adelson/illusions_demos.html

Aron, A., & Aron, E. N. (1999). *Statistics for psychology* (2 ed.). Upper Saddle River: Prentice Hall Inc.

Brody, N., & Ehrlichman, H. (1998). *Personality psychology- the science of individuality*.Upper Saddle River: Prentice Hall Inc.

Buffardi, L. C., Fleishman, E. A., Morath, R. A., & McCarthy, P. M. (2000). Relationships between ability requirements and human errors in job tasks. *Journal of Applied Psychology,* 85(4), 551-564.

Chi, C.-F., & Drury, C. G. (1998). Do people choose an optimal response criterion in an inspection task? *IIE transactions* (30), 257-266.

Craig, A., & Colquhoun, W. P. (1975). Vigilance: A review. In C. G. Drury & J. G. Fox (Eds.), *Human reliability in quality control* (pp. 71-87). London: Taylor & Francis Ltd.

Craig, A., & Cooper, R. E. (1992). Symptoms of acute and chronic fatigue. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 3-State and trait, pp. 289-339). London: Academic Press Limited.

Czaja, S. J. (1997). Systems design and evaluation. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2 ed., pp. 17-40). New York: John Wiley & Sons Inc.

de Rijk, A., (1998). So I was compelled to pull it together, and this was the model: an interview with Prof. Robert Karasek. Retrieved 20 August, 2005, from http://www.fss.uu.nl/penh/karasek.htm

Drury, C. G., & Fox, J. G. (1975). The imperfect inspector. In C. G. Drury & J. G. Fox (Eds.), *Human reliability in quality control* (pp. 11-16). London: Taylor & Francis Ltd.

Drury, C. G., & Sinclair, M. A. (1983). Human and machine performance in an inspection task. *Human Factors,* 25(4), 391-399.

Fox, J. G. (1975). Vigilance and arousal: A key to maintaining inspectors' performance. In C. G. Drury & J. G. Fox (Eds.), *Human reliability in quality control* (pp. 89-96): Taylor & Francis Ltd.

Gallwey, T. J. (1998 a). Evaluation and control of industrial inspection: Part I - guidelines for the practitioner. *International Journal of Industrial Ergonomics,* 22, 37-49.

Gallwey, T. J. (1998 b). Evaluation and control of industrial inspection: Part II - the scientific basis for the guide. *International Journal of Industrial Ergonomics* 22, 51-65.

Gallwey, T. J., & Drury, C. G. (1986). Task complexity in visual inspection. *Human Factors,* 28(5), 595-606.

Hackman, R. J., Oldham, G., Janson, R., & Purdy, K. (1975). A new strategy for job enrichment. *California Management Review,* 17(4), 57-71.

Hammar, N., Alfredsson, L. & Johnson, J.V. (1998). Job strain, social support at work, and incidence of myocardial infarction. *Occupational and Environmental Medicine,* 55(8), 548-553.

Henning, R. A., Jacques, P., Kissel, G. V., Sullivan, A. B., & Alteras-Webb, S. M. (1997). Frequent short rest breaks from computer work: Effects on productivity and well-being at two field sites. *Ergonomics,* 40(1), 78-91.

Hill, T. (2000). *Manufacturing strategy - text and cases* (2 ed.). New York: Palgrave.

Hygge, S. (1992). Heat and performance. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 1-The physical environment, pp. 79-104). London: Academic Press Limited.

Jiang, X., Gramopadhye, A. K., & Melloy, B. J. (2003). Evaluation of best system performance: Human, automated, and hybrid inspection systems. *Human Factors and Ergonomics in Manufacturing,* 13(2), 137-152.

Kellogg, R. T. (1995). *Cognitive psychology*.London: SAGE Publications Ltd.

Lazarus, R.S. (1993). From psychological stress to the emotions- a history of changing outlooks. *Annual Review of Psychology,* 44, 1-21.

Lehto, M. R. (1997). Decision making. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2 ed., pp. 1201-1248). New York: John Wiley & Sons Inc.

Lindblad, M. (2005). *Human inspection work- a case study of quality control performance in a modern global electronics company.* Unpublished Master of Science, Lund Institute of Technology at Lund University, Lund.

Marmaras, N., & Kontogiannis, T. (2001). Cognitive tasks. In G. Salvendy (Ed.), *Handbook of industrial engineering - technology and operations management* (3 ed., pp. 1013-1040). New York: John Wiley & Sons Inc.

Matthews, G. (1992). Extraversion. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 3-State and trait, pp. 95-126). London: Academic Press Limited.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research a philosophic and practical guide*.London: RoutledgeFalmer.

Michie, S. (2002). Causes and management of stress at work. *Occupational and Environmental Medicine* 59, 67-72.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors,* 38(2), 311-322.

Nachreiner, F., & Hänecke, K. (1992). Vigilance. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 3-State and trait, pp. 261-288). London: Academic Press Limited.

Park, K. S. (1997). Human error. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2 ed., pp. 150-173). New York: John Wiley & Sons Inc.

Patel-Coleman, K. (2004). Session 9: Secondary intervention: Individual programs. On *CHS M278/EHS M270: work and health winter 2004*: UCLA.

Poulton, E. C. (1970). *Environment and human efficiency*.Springfield: Charles C Thomas.

Proctor, R. W., & Proctor, J. D. (1997). Sensation and perception. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2 ed., pp. 43-88). New York: John Wiley & Sons Inc.

PsycINFO. (2004, September). Psycinfo database information. Retrieved 4 January, 2005, from http://www.apa.org/psycinfo/products/psycinfo.html
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology* 25, 54-67.
Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica,* 53, 61-97.
Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design* (6 ed.). New York: McGraw-Hill Inc.
Schnall, P. (2004 a). Session 1: Introduction to work & health; a social epidemiologic approach to the workplace and cardiovascular disease. On *CHS M278/EHS M270: work and health winter 2004*: UCLA.
Schnall, P. (200 b). Session 2: Conceptual theoretical models for OHP: Operationalization, measurement and assessment. On *CHS M278/EHS M270: work and health winter 2004*: UCLA.
Schnall, P., & Patel-Coleman, K. (2004). Session 7: International and cross-cultural comparisons; economic costs of workplace psychosocial factors. On *CHS M278/EHS M270: work and health winter 2004*: UCLA.
Schnall, P., & Perlo, S. (2004). Session 8: Primary intervention: Work organization redesign. On *CHS M278/EHS M270: work and health winter 2004*: UCLA.
Smith, A. P., & Jones, D. M. (1992). Noise and performance. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 1-The physical environment, pp. 1-28). London: Academic Press Limited.
Smith, C.A., Haynes, K.N., Lazarus, R.S. & Pope, L.K. (1993). In search of the hot cognitions - attributions, appraisals, and their relation to emotion. *Journal of Personality and Social Psychology,* 65(5), 916- 929.
Smith, R. E. (1993). *Psychology* (Vol. Company). St. Paul: West Publishing.
Smither, R. D. (1994). *The psychology of work and human performance* (2 ed.). New York: HarperCollins College Publishers Inc.
Web_of_Science®. (2004). Web of science databases. Retrieved 4 January, 2005, from http://wos02.isiknowledge.com/help/h_database.htm#sci
Wickens, C. D., & Carswell, C. M. (1997). Information processing. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2 ed., pp. 89-129). New York: John Wiley & Sons Inc.
Yukl, G. (2002). *Leadership in organizations* (5 ed.). Upper Saddle River: Prentice-Hall Inc.

APPENDICES

Appendix 1. The bayesian probability model

State of Item	*Decision*	*Name*	*Probability*	*Payoff*
Good	Accept	Correct accept	$p_1(1-p')$	V_g - kt_d
	Reject	False alarm	$(1- p_1)(1-p')$	$-C_g$ - kt_d
Faulty	Accept	Miss	$(1- p_2)p'$	$-C_f$ - kt_d
	Reject	Hit	p_2p'	V_f - kt_d

p' = defective rate;
V_g = value of accepting a good item;
C_g = cost of rejecting a good item;
V_f = value of rejecting a faulty item;
C_f = cost of accepting a faulty item;
t_d = time for decision making;
p_1 = probability of accepting a good item;
p_2 = probability of rejecting a faulty item.
k = a constant.

Expected profit of acceptance :

$$A_{profit} = \left[V_g - kt_d\right]p_1(1-p') + \left[-C_f - kt_d\right](1-p_2)p'$$

Expected profit of rejection :

$$R_{profit} = \left[-C_g - kt_d\right](1-p_1)(1-p') + \left[V_f - kt_d\right]p_2p'$$

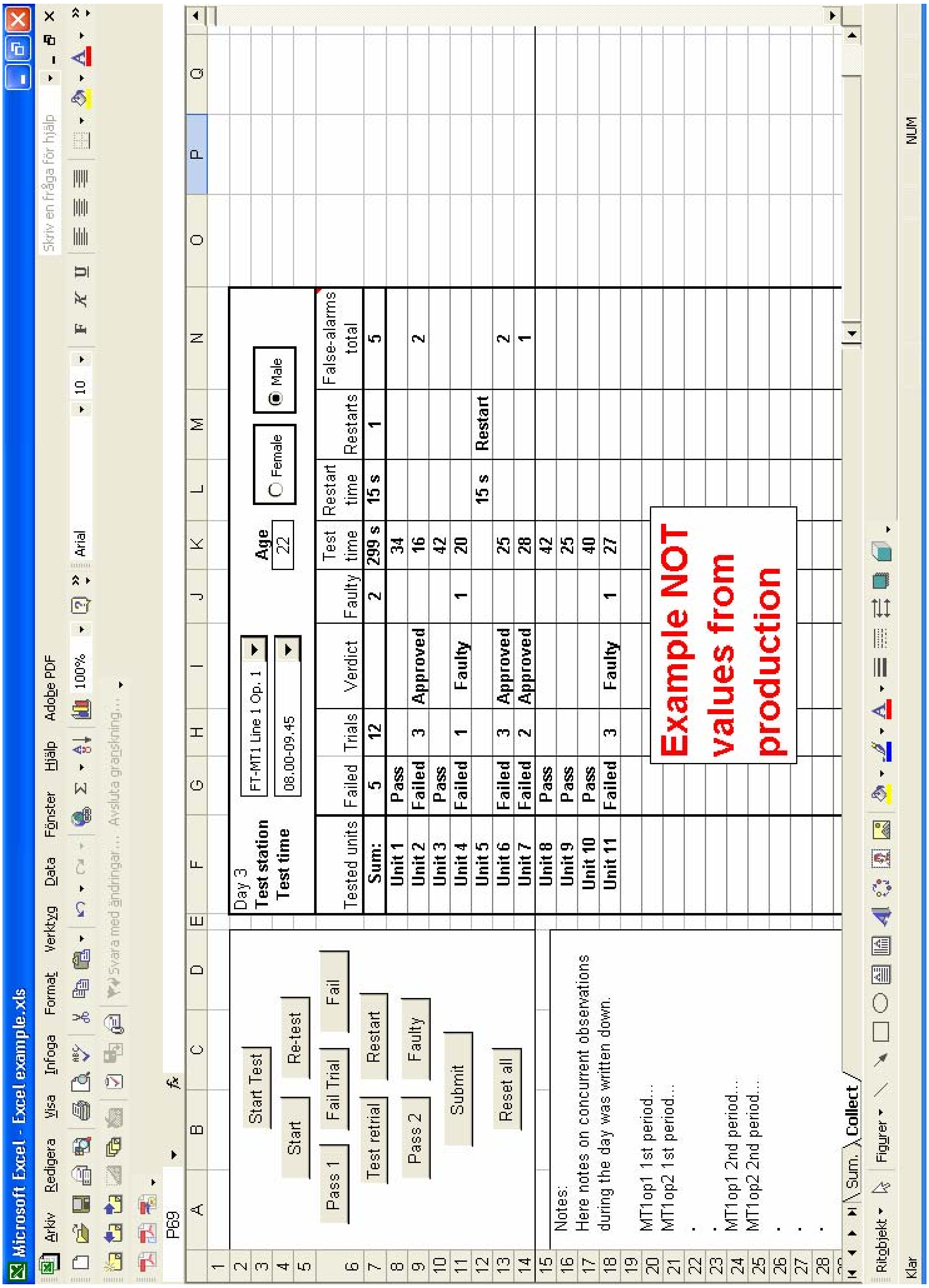
Optimal Bayesian decision criterion, dependent on the likelihood ratio of acceptance vs rejection and associated values and costs :

$$\beta_0 = \frac{Max(A_{profit})}{Max(R_{profit})} = \left[\frac{(1-p')}{p'}\right]\left[\frac{V_g + C_g}{V_f + C_f}\right]$$

Drury's "sluggish β" phenomenon, m is the slope of the regression line :

$$\beta_C = m\beta_0 + (1-m)$$

Appendix 2. Program in Microsoft Excel for collecting data.



Appendix 3. Literature databases used

Science Citation Index Expanded (SCI-EXPANDED) is a multidisciplinary index to the journal literature of the sciences. It fully indexes 5,900 major journals across 150 scientific disciplines. The Science Citation Index Expanded includes all cited references captured from indexed articles. Coverage 1945-present (Web_of_Science®, 2004).

Social Sciences Citation Index (SSCI) fully indexes more than 1,725 journals across 50 social sciences disciplines, and it indexes individually selected, relevant items from over 3,300 of the world's leading scientific and technical journals. Coverage 1956-present (Web_of_Science®, 2004).

Arts & Humanities Citation Index (A&HCI) is a multidisciplinary index covering the journal literature of the arts and humanities. It fully covers 1,144 of the world's leading arts and humanities journals, and it indexes individually selected, relevant items from over 6,800 major science and social science journals. Coverage 1975-present (Web_of_Science®, 2004).

PsycINFO a database which provides access to international literature in psychology and related disciplines. The sources include nearly 2,000 professional journals, chapters, books, reports, theses and dissertations, published internationally. Additionally, there are more than 12 million cited references in 305,856 journal articles, books, and book chapters. PsycINFO is a department of the American Psychological Association (APA) dedicated to creating products that make it easier for researchers to locate psychological literature relevant to their research topics. Coverage 1840 - present (PsycINFO, 2004).

Appendix 4. MT1 false-alarms

Table 1.
Test of normality for false-alarms in MT1.

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
FALSEAL1	,940	9	,584
FALSEAL2	,898	9	,240
FALSEAL3	,843	9	,062
FALSEAL4	,869	9	,121

Table 2.
Test of normality for square root of false-alarms in MT1.

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
SQRT(F.A.)1	,924	9	,424
SQRT(F.A.)2	,957	9	,771
SQRT(F.A.)3	,931	9	,490
SQRT(F.A.)4	,940	9	,582

Table 3.
Descriptive statistics for false-alarms in MT1.

Descriptive Statistics

FALSEALA

SUBJECT	PERIOD	SEX	Count	Mean	Standard Deviation	Coefficient of Variation
Total	1	Female	3	,15333	,090738	59,2%
		Male	7	,16629	,115150	69,2%
		Total	10	,16240	,103482	63,7%
	2	Female	3	,13433	,095929	71,4%
		Male	7	,21257	,102875	48,4%
		Total	10	,18910	,102610	54,3%
	3	Female	2	,12450	,021920	17,6%
		Male	7	,24371	,144832	59,4%
		Total	9	,21722	,136219	62,7%
	4	Female	3	,17033	,097316	57,1%
		Male	7	,30000	,161785	53,9%
		Total	10	,26110	,153223	58,7%
	Total	Female	11	,14755	,075853	51,4%
		Male	28	,23064	,135005	58,5%
		Total	39	,20721	,126093	60,9%

Totals that are aggregated over either a single category of a variable or a split file variable are omitted.

Table 4.
Test data had to be transformed with the square root to avoid problems due to sphericity for false-alarms in MT1.

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	df	Sig.
SQRT(F.A.)	,246	5	,096

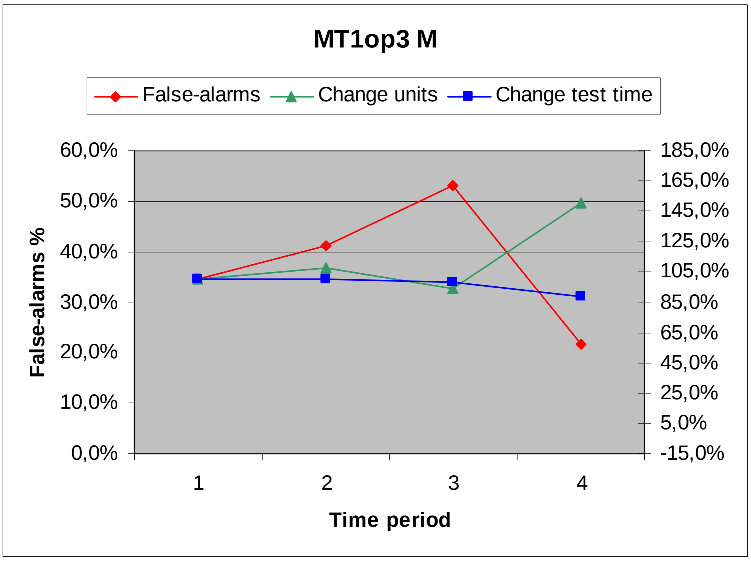


Figure 1. *Performance curves for MT1op3 male 20 years old with signs of extrovert personality.*

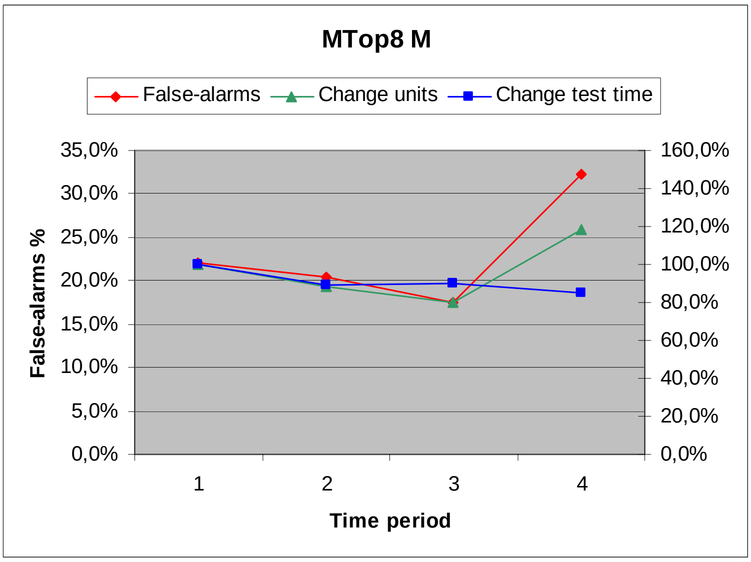


Figure 2. *Performance curves for MT1op8 male 18 years old, that appeared to be sensitive to stress. He had lunch before period 2.*

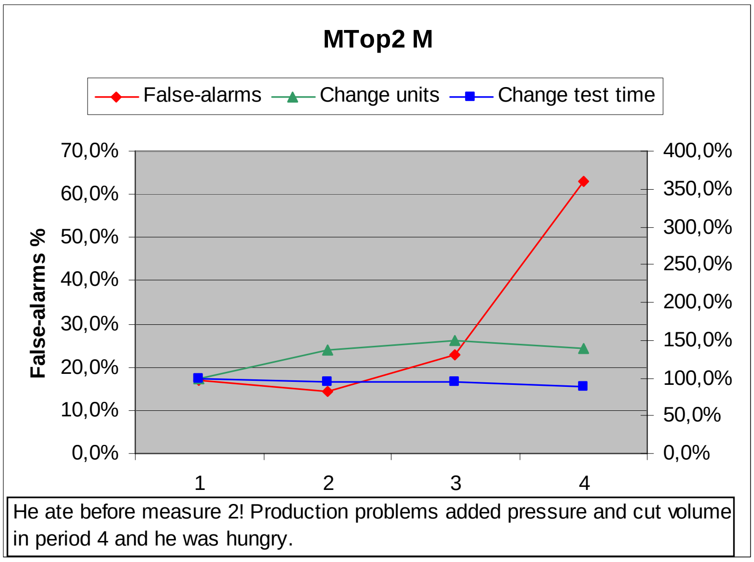


Figure 3. *Performance curves for MT12 male 18 years old, that had lunch before period 2.*

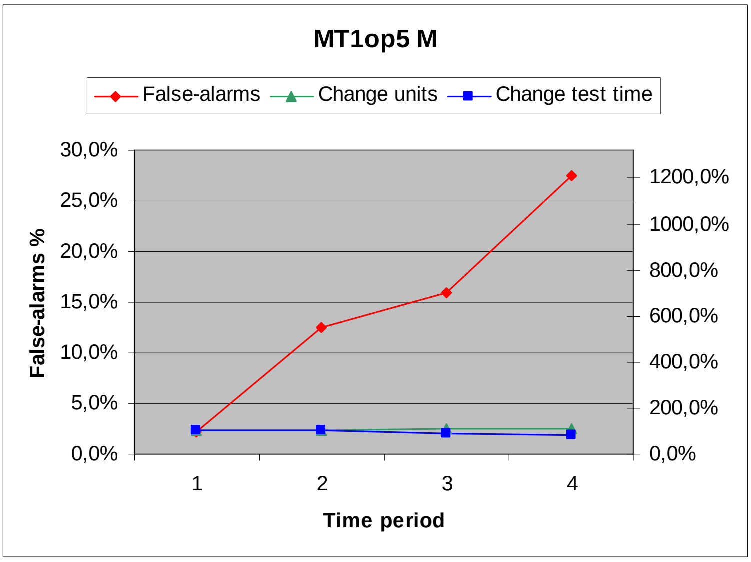


Figure 4. *Performance curves for MT15 male 23 years old. He started at a very low level of false-alarms, but showed the most dramatic increase. This increase was mostly between meals. Though it can't be seen in the figure, he reduced the test time with about 20 %, explaining some of the deterioration in performance.*

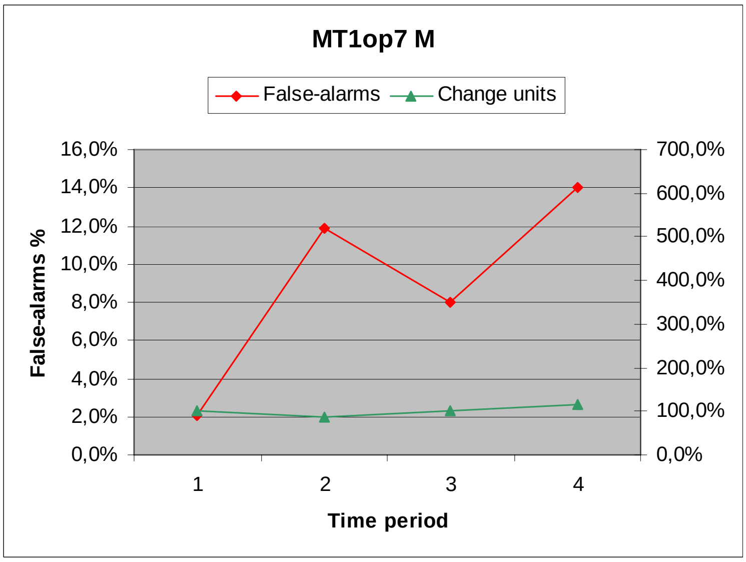


Figure 5. *Performance curves for MT17 male 19 years old. Test time data unfortunately missing. Note the dramatic increase in false-alarms between meals, even though the time between period 2 and 3 is 4 hours compared to 2 hours between the others.*