THE INFLUENCE OF PERFORMANCE MEASUREMENT ON ACTOR'S PERCEPTION OF TASK IN GOAL ORIENTED SYSTEMS

By

Michal P. Bobinski

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Management Sciences

Waterloo, Ontario, Canada, 2005

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Michal P. Bobinski

Abstract

This thesis addresses the problem of the influence of a control system on the behaviour of an actor in a social or socio-technical system. In particular, the influence of a performance measurement mechanism on the behaviour of an actor and on the development of workarounds is being studied.

Current literature on those topics generally addresses only selected and rather obvious reasons for the existence of dysfunctional behaviour or the workarounds. However, no precise models of the cognitive processes or the explanation of the mechanisms, which govern this problem, are proposed in a satisfactory manner. In addition, most researchers have focused only on the system's point of view of the task, paying less attention to the actors' perception of that task. Furthermore, the existing body of work mainly uses a case study format to explain the phenomenon or to validate the proposed solutions and theories.

In this thesis, the problem of the influence control system on the behaviour of the actor is framed in terms of four major concepts: (1) the concept of complexity of the task not being fully captured by the performance measurement mechanism; (2) the concept of an actor perceiving that extra complexity is not being captured by the system and thus choosing alternate paths other than the system-prescribed path; (3) the concept of a network of valence forces associated with alternate paths; and finally, (4) the concept of similarity judgment between the alternative paths and the system-prescribed path based on the actor's model of the control system's point of view.

This thesis develops the theoretical framework for analyzing and understanding the issues of dysfunctional behaviour and workarounds. It also presents an empirical experimental study in support of the theoretical discussion and the hypothesis. The experiment examines subjects' rating of quality, defined as a degree of similarity to a target object, of several objects on a page under various performance measurement conditions. The stimulus used for experiment was made up of two dimensional quadrangle figures, including rectangle, parallelograms and trapezes, in various shades of red colour.

The results show that a person's similarity judgments are highly correlated with the valance induced by the performance measurement system on a given dimension such as either shape or colour. This suggests that the subject's perception of similarity of two objects was influenced by the performance measurement system. It is concluded that the behaviour and the actor in the system and his/hers decision making process are highly influenced by the system of valance forces induced by the performance measurement system as well as the judgment of similarity of available alternatives.

Acknowledgment

First and foremost I would like to thank my advisor Professor Rob Duimering for showing me what good research is all about. He has guided me through difficult times when it seamed like there will never be a light at the end of the tunnel; he kept me focused when my mind was wandering off and he was always there ready to share his knowledge and experience with me. It has been a great journey and rewarding journey. I can only hope that one day I will be able to guide someone as well as he has guided me.

I would also like to thank all of those who took the time to read, comment and discuss my work with me: Professor Frank Safayeni , Professor Scott Jeffrey as well as Kevin O'Connor and Brandon Sweet who were my proofreaders , thank you so much for your time, commitment and effort, I know it was not an easy task.

Thanks also to my follow students: Kim, Smitha and Ray for many lunches, tee breaks, and for listing to me rant and rave from time to time. Thank you all.

Dedication

To my parents, who always believed in me and supported me with great love and good advice at each and every step of the way.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The need for a good control mechanism in any goal-oriented system is undisputed. One of the major components of a control system is performance measurement mechanism. This mechanism is often closely linked with the second major component of control, a reward and punishment system. The problem of the impact of the performance measure on the behaviour of the actors in the system is a known problem and it has been present for quite some time in the literature. This problem can be found in accounting, supply chain, and behavioural science literature.

It has been recognized that the design of a performance measurement mechanism and the choice of the performance indicators have an impact on how actors perform their tasks and how their priorities and attitudes can change depending on what part of the task is being currently monitored. Often simply measuring performance may result in undesired, often called dysfunctional, behaviour that may work against the achievement of organizational goals. For example, the original goal of a given task, for which the control system was created, may become replaced by the goal of achieving the highest rating on the performance indicator. Furthermore, this can create a situation where the control mechanism may be unable to detect the anomaly simply because the system cannot distinguish between the achievement of the original goal and the achievement of new goal created by the introduction of the performance indicator. Blau (1955), in his study of New York employment officers illustrates that very situation. The social worker's task was to arrange the job interviews, the goal of which was to match a person (and the skills this person had) with an appropriate job offering. The performance measure used to evaluate this task was the number of interviews each officer had arranged. Blau (1955) noticed that the social workers, realizing that only the volume of interviews counted, started to send just about anyone to any interview. This situation resulted in the view that the employment office was doing its job, but in fact the real goal of that social organization, finding people jobs, was being ignored.

The problem is not limited only to domain of business enterprises, but is present in many areas, such as government and social programs, that in one form or another use performance measure as a part of their control system. Existing literature approaches this problem by showing instances of the unwanted behaviour and then attempting to explain it by drawing links between the performance measure system and the reward system. Many examples of the dysfunctional behaviours are presented usually in case by case bases and the root causes of the problems are often attributed to the improper selection of the individual performance measures in each particular case study. As a response to the problem, the authors propose new, improved, more comprehensive and encompassing measures are often proposed and argued for.

Some authors, such as Lawler (1976), go a step further and offer some models of the behaviour, however, those are generally limited to theories of motivation and do not offer an insight into the mechanics of decision making done by the actors.

Economics literature has also addressed the problem of performance measure and design of appropriate incentive systems. Author such a Kerr (1995), Gibbons (1998) or Courty et al (2003) approached the problem of performance measurement form the perspective of designing the evaluation and reward system based on the principles of game theory and agency theory. In their work they focus on the selection of wage versus piece-work reward systems and the implication of such choices for the behaviour of the actor in the system. Though this types of deliberations are important and do contribute to a better design of performance measurement systems, they do not provide the basic answers to the question of how the actor perceives his/hers task, and how he or she decides on a manner in which to perform the task at hand.

Related to the issue of performance measurement are “workarounds”. The problem of “workarounds”, refers to the situations where actors are forced pursue a course of action that is different from the prescribed actions in order to achieve the goal. This working around the problem is often created by the fact that the system demands certain results, but due to some design flaw or error does not have the functionality built in to produce that result. For example, in a certain warehouse operation workers were supposed to stick special barcode labels on each box being loaded on the pallet. Then the barcode had to be entered in to a computerized shipping system, thus allowing for electronic representation of each pallet shipped (a requirement made by the trading partner and general industry standard). Some times the workers (actors) encountered a situation where the actual boxes were smaller then the barcode labels. Faced with the necessity of having both the physical label and matching electronic representation the workers decided to shrink wrap entire pallet and stick only one label on the entire pallet and enter the data to the computer in such way that it looked like there was one big box on the pallet. Though theoretically against the rules, this workaround enabled the operations of the warehouse to continue. In fact this workaround was virtually undetectable because the control of the process was being performed using the electronic representation of the shipments and not the physical appearance of the pallets being shipped.

Though different on the surface, performance measurement often creating negative results and "workarounds" often trying to overcome difficulties to create positive outcomes, appear to have a very similar structure at their core. In both cases, the actors perceive the situation as being more complex in terms of variety of responses or problems than the system within which they operate actually does. In other words, it can be said that the similarity between the workarounds and performance measure problems is the gap between an actor's perceptions of the task and how it is to be performed and the actor's perception of the formal system's view of the same task.

So far most of the literature has focused on a single view of the problem, usually from the formal system's point of view. The conclusions and solutions have been generally limited to a particular instance of a problem. The explanations or models of the mechanics of the decision-making by the actors have also been generally limited. Gibbons (1998) admits that the economical models, for example do not take to the account many factors such as psychological and cognitive factors affecting the actor.

This thesis will attempt to create a model of behaviour capable of explaining the actual mechanism of decision-making done by the actor in the situation, as he/she perceives it in relation to the design of the system he/she operates within. In contrast to existing case-by-case approaches to the problem, it is a goal of this research to create a model that could be used more generally in all situations involving the design of a control system.

In the first part of the study, the theoretical framework will be developed and discussed. This discussion will be followed by a set of general hypotheses. In the second part, results of an experimental study performed to test some of the hypothesis will be presented. A general discussion and a brief outline of future research possibilities will conclude this thesis.

Chapter 2: Review of Background Literature

In this chapter the existing literature on the effects of performance measurement and on the "workarounds" will be examined. In the first section, the problems of performance measures and the dysfunctions thereof will be addressed. This section will be concluded with a brief critical evaluation of the relevant existing models and approaches to the studied phenomenon. In the second section, a brief overview of literature on "workarounds" will be presented.

2.1 Performance Measure Problems in the Literature

Organizational control systems and mechanisms are designed and implemented to ensure that "planned activities are producing the desired results "(Lawler, 1976). Performance measures along with the reward system are the two major components of most control systems in use today. It is accepted that any goal-seeking social and socio-technical system needs some form of control if it is to reach its goal. Merton (1952) writes, "…an effective bureaucracy demands reliability of responses and strict devotion to regulation." The role of the performance measure is to measure the degree to which actors adhere to the rules and procedures designed to reach the organizational goals.

In general there are two types of control systems. The first type, used in organizations that have a transparent and easily observable process but not an easily observable goal, is referred to as process- or behavioural-based control. The second type is usually used in situations where the results can be observed and quantified by the actual process of attaining the goal is not easy to observe or measure. This type of control is referred to as output based control (Gibson et al, 1976; Agarwal, 1999; Jensen, 2004)

It has been noticed, however, that the very design of the control mechanism and thus the selection of the performance measures can have negative effects by producing behaviour which can be classified as dysfunctional and unwanted (Neely et al 1997). The literature on this issue can be divided in two distinct categories: (1) dealing with the problems surrounding the selection of appropriate performance indicators; (2) dealing with the models of behaviour. Each problem category is explored in greater detail below

2.1.1 Inadequate Performance Indicators

One of the most difficult tasks when designing organizational control systems is selection of appropriate performance measure indicators. The main challenge comes from identifying indicators that would accurately measure the achievement of a given task. This challenge often stands from the fact that many tasks cannot be defined very well or cannot be monitored easily (Gibbons, 1998). Many authors have noted that most commonly used performance indicators, such as sales per employee or units assembled per unit of time, which are often derived from a statistical - or accounting approach to control, suffer from the fact that they are only able to capture a small portion of the entire process or task being performed (Argyris, 1971; Ouchi, 1977; Ghalayani and Noble, 1996). In other words, those indicators can be inflexible as they are unable to capture the full complexities of the tasks and the situations for which they are being used (Ghalayani and Noble, 1996). One reason for this problem is that not everything can be measured or quantified. Often, as Perrin (1998) and Feller (2002) point out, performance measures and indicators are chosen on the basis of data being available, not necessarily because it is reflecting exactly what should be used as a performance indicator.

It has also been noticed that since most of the reward systems in organizations are tightly connected to the performance indicators, a poor selection of the performance indicators results in various unwanted and mostly unanticipated behaviours displayed by the actors. Ouchi (1977) and Feller (2002) point out one such problem. They have observed that the tasks or parts of the process, which are not being monitored but are an integral part of the process, become devalued in the eyes of an actor and often become neglected. In the case of sales force described by Ouchi (1977) or Agarwal (1999) the tasks of arranging the stock on the shelves had been ignored as it was perceived by the actors as a task that did not directly improve their performance. Further, Ouchi suggests that the task of training new sales employees was being perceived as unwanted and even detrimental to the other actors' interest as every new sales person equalled a new source of competition. Ouchi (1977) suggest only that this kind of behaviour is motivated by the needs of the actors to benefit themselves by looking good. Unfortunately, he does not offer any other model or in-depth explanation of the phenomenon.

The problem of the selection of inadequate performance measures and the subsequent effect of that selection on other tasks can be found in Blau (1955). In his study of New York City employment officers, Blau shows how the selection of the metric that did not reflect the true complexity of the job negatively affected the attainment of the organizational goal. Since only the number of interviews arranged was selected as a measure of performance, other tasks such as ensuring an appropriate match between the candidate skills and the job required skills were devalued and subsequently neglected by the actors. Thus it created a situation in which the true goal of the organization was lost, yet the problem became undetectable to the system by any means other than Blau's study or a system wide audit.

The difficulty in selecting appropriate indicators for the performance measure was further highlighted by Ridgeway (1956). In his study of dysfunctional consequences of control systems, he identifies the problem and consequences of using: (1) single criteria systems such as those used in Blau (1955), (2) multiple but independent criteria, and (3) composite or interconnected criteria.

Ridgeway argues that each of the choices for the control system can lead to the negative behaviour on the part of the actor. It is because either the independent criteria do not accurately capture the nature of the task being performed, or the composition of several criteria doesn't allow the actor to reach the set goals due to the fact that he/she has a limited "effort" to give on each of the indicators making up the composite. In this case the actor can get frustrated and seek an alternative way of increasing the composite indicator

However, even if theoretically appropriate indicators, which truly capture the task and its complexity, were chosen, the problems do not disappear. Perrin (1998) suggests that seemingly clear and appropriate performance indicators could be misunderstood or misinterpreted by the actors. As an example, he uses the definition of the word "Client." He shows that different government and social organizations define the meaning of that word depending on their functions and their goals. Thus the meaning of 'client' for a mental health clinic is different from the meaning of 'client' for an employment agency despite the metric being called "clients served." Feller (2002) goes on to suggest that sometimes even the relevant performance measures can be misunderstood and used inadequately. In addition he also suggests that politics also plays an important role as different political interests may have different agendas, and those can influence the selection of performance indicators or the definitions of indicators that are being used, adding even more complexity and confusion to the situation at hand.

Some authors not only highlight the problems, but also propose solutions. For example, Ghalayani and Nobel (1996) or Neely (1997) offer, as a remedy to what they call "traditional indicators" such as cost per project or productivity, a set of new, more comprehensive and more encompassing indicators such as cycle time (Ghalyani and Noble, 1996). However, the new indicators, though offering an improvement over old methods, in the mind of the authors do not offer greater insight or solve the real problem of performance measures affecting behaviour. They are based on the same model and understanding of the behaviour of the actors as the old ones; namely the assumption that motivation to act comes mainly from the attractiveness of the reward.

As a result of either the inappropriate selection of the indicator or a misinterpretation of one, two other major problems have been identified in the literature.

First, Merton (1952) suggests that the selection of performance indicator, and especially the fact that those are closely related to the incentives and rewards, can create a situation when "adherence to the rules, originally conceived as a means, becomes transformed into an end-in-itself; there occurs familiar process of *displacement of goals*…"(p365). The goal displacement is well illustrated by the aforementioned study of Blau (1955). The goal of the employment agency was to find people jobs, but this original goal became replaced in the mind of the actors by the goal of looking at the statistical record (the number of interviews arranged), which was used to evaluate and reward the employment officer.

Kerr (1995) shows in various examples how the misalignment of rewards, meaning selection of the performance indicators which “ do not further the true goal of the organization” (Courty et al., 2003) and also referred to as incentive distortion, can lead the displacement of the original goal, and can lead to a situation when the actor might be working against the very goals the control mechanism was supposed to encourage. Kerr illustrates this situation by showing an example of a football player whose ranking was often based on the possession of the ball during the game. Such a performance indicator resulted in players passing the ball less often and could easily be leading to the team losing the game in the end.

The problem of goal displacement is also noticed by Irving et al (1986), indicating that even the computerized forms of control are not impervious to the impacts of control on an actor’s behaviour. As a matter of fact the argument goes further, pointing to the fact that computerized systems applied at individual levels lead into greater negative effects due to perceived loss of privacy and dissatisfaction. Such a loss of motivation could lead to greater effort on behalf of the actor to cater to the performance indicator while ignoring the goal for which that indicator was put into place.

Perrin (1998) and Feller (2002) also agree, in their studies of government and social organizations, that the selection of irrelevant or not fully encompassing indicators often leads to the case of “making numbers” a primary goal. Since in most cases the rewards and punishments are connected to the results obtained on the performance measure, actors will arrange and present their work in such way as to maximize their reward rather then achieve the organizational goal. Perrin (1998) uses an example of budgets and cost shifting as opposed to cost savings to illustrate that problem.

The second important problem discussed in the literature is the fit between the strategic plans of an organization and the control system thereof (Ouchi, 1977; Ansari, 1977; O’Mara et al, 1998). Performance measures often do not reflect the strategic direction of organizations and often their design induces behaviours that are detrimental to the attainment of those strategic goals. O’Mara et al (1998) point out how the short-term perspectives of many accounting and statistical based indicators negatively impact the attainment of long term strategic goals. This view is also shared by Ouchi (1977) who gives an example of a sales force, whose members are being evaluated and rewarded on their sales volumes, resorting to high pressure sales tactics and stealing sales in order to look good on their performance review. Ouchi (1977) points out that those actions are short-sighted as those kinds of tactics can only increase sales in the short term.

Other researchers including Ridgeway (1956), Kaplan and Norton (1992), Kerr (1995) and Perrin (1998) show how inadequate control system design can lead to sub-group optimization as opposed to organization-wide benefits. Ansari (1977) points out that the control systems must be designed in a way that takes into account the interaction between the structural domain of control, " …best exemplified by cybernetics " and behavioural domain, which " emphasizes the human and social process " of achieving control. Otherwise the control system and the strategic direction of the company will diverge. He proposes a new framework for the design on the control system, one that integrates the structural, social and support (rewards) domains of organization. Ansari's approach despite being more comprehensive and acknowledging the role of actors perception of the task or the process still does not offer an insight into the behaviour of the actor and does not explain the nature of dysfunctional behaviours observed by him and other authors.

Kaplan and Norton (1992) also propose a new approach to the design of control systems. They propose to combine the financial indicators (cost per unit) and non-financial ones (customer satisfaction). Their proposed "Balanced Scorecard" approach links indicators from four different perspectives: (1) customer, (2) internal, (3) innovation and learning and (4) financial. In their opinion, such a mix of indicators would allow managers to view the performance in several areas at once and thus allow them to react to any case of local/sub-group optimizing. However, certain assumptions made by Kaplan and Norton, in particular cause and effect assumptions, have been heavily criticized by other authors. Norrekilt (2000) argues that the "cause-and-effect-chain", linking some of the non-financial indicator to financial indicators as described by the Kaplan and Norton does not in fact hold. For example, "efficient business process – high customer satisfaction- good financial results."(Page 72). He points out that there are no solid theories of behaviour linking the new measures with the proposed benefits of the "Balance Scorecard" approach. Norrekilt (2000) also concludes that “Balanced Scorecard” does not make an optimal strategic control tool because of its hierarchical and top-down design.

2.1.2 Models of Behaviour

Merton, (1940) as summarized by Simon and March (1958), offers a classic model of dysfunctional behaviour caused by the demand for control and emphasis on reliability. In Merton’s model the unintended result of the performance measure is called rigidity behaviour. It comes as a result of felt need for the defensibility of individual’s actions. Usually this stems from the fact that actor’s performance is often judged on the basis of his/her adherence to the rules. It is easy to observe how this model also would apply to the problem of goal displacement, as the adherence to the rules becomes goal unto itself, overshadowing the original reason for which the rules were set. Furthermore, Merton’s model implies that the behaviour and choices the actors make are predominantly governed by the attractiveness of the reward or, on the other hand, the fear of punishment. Though this model does offer certain insight into the studied phenomenon it should be noted that in this model an actor’s reaction of obsession with the rule doesn’t lead into a workaround. Nevertheless, the impact of the control system on the behaviour is clear.

One other model is also summarized by Simon and March (1958) who offer yet another glimpse into the studied phenomenon. The Gouldner model of behaviour illustrates an actor in responds to the design of the control system, in this case rules and procedures. Just like in the case of performance indicators derived from statistical measures, rules and procedures often do not cover all possibilities but are usually designed and focused on general cases. Gouldner shows that over time actors identify the “minimum level of permissible behaviour” and utilize that knowledge by performing only the tasks in such capacity as to do only what is required and not necessarily what should be done. One can draw a parallel between this model and the effects of the inadequate selection of performance indicators discussed in the previous section.

Lawler (1976) in his study of the effects of the control systems on behaviour, identifies three types of dysfunctional behaviours: first, which he calls rigid bureaucratic behaviour illustrated by the studies conducted by Blau (1955) and Berliner (1961); second, resistance to control which he illustrates by the studies of MIS specialists conducted by Argyris (1971); and third, production of false data.

Lawler proposes two approaches to analyzing this problem. First, he devises a method of predicting whether or not a dysfunctional behaviour might occur. To do that Lawler draws on the analogy between a thermostat furnace system and the control system. He proposes a set of seven questions, which correspond to seven parts of the thermostat system: (1) what is actually measured, (2) who is setting the standard, (3) who or what acts as a discriminator comparing the performance to the standard, (4) what actions are taken and what are the rewards and punishments used to motivate the behaviour, (5) who receives information about the deviation from the standard, (6) what is the measured activity and can it be measured, and finally (7) what is the basic source of motivation for the activity. The answers to those questions, according to Lawler, will determine if the control system is likely inducing one of the three dysfunctional behaviours. While this might be true, this approach in no way is able to identify which of the behaviours is to occur or what the actor's reasoning is to engage in one form of action as opposed to another.

The second approach offered by Lawler (1976) is based on Vroom's Expectancy Theory. Lawler presents two similar models one for the extrinsic motivation model and one for the intrinsic motivation (Refer to Fig 1. below). Both models are adaptations of Vroom's model

Lawler (1976) believes that the actor's behaviour is governed by two key relationships. First is the actor's belief that the efforts he/she can master will lead to a successful performance ($E \rightarrow P$). That is to say that based on the actor's previous experiences, knowledge and self esteem actor beliefs that he/she can accomplish a given task. Second is the actor's belief that the performance will lead to the desired outcome ($P \rightarrow O$). In this case it is a belief that the actor is going to be rewarded for his/her successful performance. Furthermore Lawler (1976) argues that for the actor to perform the outcome, a reward has to have some positive valence associated with it. The reward must in some way satisfy actor's needs in order to have some positive valence. The major difference between the intrinsic and the extrinsic models lie in the relationship between the “performance” and the “reward”. In the intrinsic model the actor is always rewarded whenever he/she achieves the outcome. The reward comes from within for example as feelings of accomplishment. In addition to that in the intrinsic model the belief that effort will result in the performance has a positive effect on actor’s perception that preference will lead to outcome. Once again the explanation lies in the fact that the entire motivation comes from within the person and serves to satisfy goals, which cannot be satisfied by monetary means.

It the case of the extrinsic model the rewards are assigned by external party and thus, for the actor, the probability that the performance will lead to an outcome, which will result in a reward is not 100 %. This very reason can result, according to the author, in an actor’s lack of motivation to perform, or decreased perception of valence associated with the reward.

Thus generally it can be concluded that relationship between the control system and the reward system, and in particular the actor’s perception of the valence attached to the specific reward, is what motivates the actor to perform. This observation is very important and is very useful in understanding the mechanics of the studied phenomenon that is the notion and the role of valence that will be presented later in the theoretical formwork chapter. Unfortunately Lawler’s approach is only capable determine the outcome of very well defined and specific situations thus doesn’t explain the fundamental mechanism behind the influence the control system has on the behaviour. Lastly Lawler (1976) never really gives a definition of what the rigid bureaucratic behaviour is, he only shows the reasons and the outcome of such behaviours,

Economics literature offers one other model of behaviour based on the classic agency model. In this model the actor is characterized as risk averse (Gibbons, 1998) and though it is often not explicitly stated but motivated by the notion of economic profit. Those authors contribute actors’ behaviour to the incentive distortion, which can be defined as selection of performance indicators that “…do not further the true goal of organization.” (Courty et al., 2003). Furthermore Courty et al. (2003), and others such as Baker et al. (1994) or Prendergast (1999) suggest that actor action can further explained and predicted using principals of game theory and an assumption that at each stage of the game the actor is risk averse and motivated by the utility he/she derives form the reward. That explains why

Kerr (1995) and others suggest that only what is measured gets done. As a result the literature proposes to implement some combination of objective and subjective performance measures (Pendergeast 1998; Baker et al., 1994), and also suggest that the design and the implementation of the performance measurement system is a dynamic process (Courty et al., 2003) because the only way to know if the measure was accurate is to observe its effect on the actor.

Though all of those models are helpful in recognizing the potential problems and pitfalls of performance measurement design and implementation, by their own admission they fall short because they ignore important factor such as psychological and cognitive issues (Gibbons, 1998). Prendergast (1999) for example notices that most of the models are only theoretical and heavily reliant on assumptions. Furthermore Prendergast points out that many economical models are developed based on what is easily observable and measurable, however even that data is not a perfectly reliable since it can be very hard to isolate the effects of any particular policy within large and complex organizations. The major limitation of the economic models is that they assume that the actions the actor can engage in are mutually exclusive. This means that the actor can either do A or B when in fact this thesis will argue that there are degrees of doing A and B and that both actions often are interrelated and cannot be easily separated despite the fact only A, for example , contributes to a performance indicator..

Though not directly related to the notion of dysfunctional consequences of performance measures one other body of work can offer an insight into the problem. One of the issue, that were already motioned, is a fact that in many cases the unwanted consequences of the control systems are very difficult for that control system to detect and thus difficult to deal with. Duimering and Safaynei (1998) conducted a study on the role of the language and the formal structure on the task of reporting in organizations. They have found that the formal organizational structure " creates an elaborate system of statements in language that define in simple and positive terms what is officially supposed to be going on within the organization". These findings may be used to explain why negative feedback loop control systems could be unable to detect the problems, such as rigid (bureaucratic) behaviour, workarounds or falsification of data.

2.2 Workarounds in the Literature:

The phenomenon of workarounds is closely related to the problem of performance measure and its impact on ones behaviour. Actors often have to find a new and creative way of using the existing functionality of the system in order to achieve the goals that the system has set for them and/or their own personal goals (Petridas et al, 2004) Workarounds often occur as a result of problems created by some error or deficiency in the design of the system (Petrides et al, 2004), which often includes the way the control mechanisms are designed and implemented.

Though not exclusive restricted only to the domain of IT, workarounds have been found to be a systematic phenomenon in such areas as workflow technologies (Hayes, 2000) and ERP systems (Hamilton 1998) and often stem form the "restrictions arising from the functionality of the technology"(Hayes, 2000). The term 'workaround' also appears in the literature on military issues and applications. In this case the meaning of the word remains similar as in previous example, and reflects actions aimed at solving problems, which were not anticipated or planned for (Kingston, 2001; Parish et al, 1980). The examples above strengthen the conclusions of Suchman (1987) that there exist a number of different ways of attaining the same outcome and that the specific choice of the "path" depends on the circumstances of the situation.

Similarly to the problem of performance measure, existing literature acknowledges the phenomenon of "workarounds" but it seems to be focusing on the reasons for the existence of workarounds. For example Coate (1996) shows how hard it is for organizations to redesign or modify existing processes, thus giving a reason and need for the workarounds. From such examples it can be concluded that many systems, in particular IT systems, can be considered inflexible just as some of the performance indicators are. Furthermore, it has been noticed that many IT systems are design to meet only the most common needs of all the intended users, and thus often do not address the needs of some specific groups (Petrides, 2004). This situation is similar to the problem of performance indicators, which often do not capture the entire complexity of the performed task for the entire group

The inflexibility problem of IT systems and the issue of systems addressing only the generic needs were investigated by Hamilton (1998). In her study of workarounds in the IT industry she uses the principals of cybernetics in trying to explain what function do workarounds they play in the operations of organizations. Hamilton suggests like Suchman (1987) and Hayes (2000) that there is a gap in functionality between the design of the IT systems and the needs of individual users of the system and that the needs of the users cannot be fully determined and planned for ahead of time. That gap between the needs and available functionality is a reason why some users or group of users must seek alternative ways of achieving their goals. Using the notion of verity, meaning the possible number of the states of the environment, introduced by Ashby (1956), Hamilton goes on to explain the that a significant portion of external variety may not be captured by the existing IT structure and thus causing the problems to emerge. Hamilton (1998) argues that workarounds are the tools that provide the extra variety handling capability, which allows the system to operate and to adapt in order to reach its goal. Further explanation of the cybernetics principle of variety and its role in the understanding of the studied phenomenon will be presented in the theoretical framework chapter.

Though much more in-depth, Hamilton's model explains only the reasons for workarounds and once again shows that workarounds are a systematic phenomenon just like the reaction to the performance measures. She does not explain what the model of the behaviour is and what exactly drives the choices of one workaround versus another.

2.3 Evaluation of the Existing Approaches

Prior models present in the literature do not seem to focus on explaining the actual process, which they often call a" bureaucratic behaviour." In most cases various reasons are given for the occurrence of the dysfunctional behaviour but the actual investigation of that behaviour stops at the level of considering the motivation of the actor to do something and not the actual mechanics of making a decision. Thus it can be said that there is an exhaustive list of causes of potential problems and an accompanying list of effects (the dysfunctional behaviours), but not much linking the two. For example Lawler's model, which is based on the Expectancy Theory, is able to potentially provide the reasons why the employment officers in Blau's study behaved in the dysfunctional way but it cannot offer any theoretical explanation of how they have arrived at the choices that they made and why the control system was unable to detect that the displacement of the goals created by the selection of the performance indicator.

As was mentioned previously, many solutions, in form of new methods, are offered to the problems that were noticed and reported. However the new solutions, though they improve on the traditional approaches, once again ignore the fact that majority of the problems only become known as a result of an audit or a study. What does not seam to be addressed and explained, except in part by Duimering and Safayeni (1998), is the fact that the control systems in most cases were incapable of detecting that actors in the system had a different way of seeing the measures due to the ambiguity in the chosen indicators.

Finally not a lot of consideration has been given to the actors perception of the task at hand and his understanding of what the task is, should be and how the system perceives and evaluates this task, with the exception of Hamilton (1998). The majority of the discussion has been based around the control system itself and only in the case of Ansari (1977) has some consideration been given to the mutual impact that social (behavioural) and technical (control) systems have on each other and how both can influence each other's perceptions.

This study will build on existing models and approaches and will try developing a behavioural model, which could be used to explain the studied phenomenon in better detail and with greater accuracy. It is also the goal of this study to create a model, which could be applied more generally in various situations involving the design and effect of control systems.

The following chapter will draw on the findings mentioned in this chapter to create a theoretical framework, which will be concluded with a set of operational hypotheses.

Chapter 3: Theoretical Framework

3.1 The Role of Cybernetics

A common problem identified in both the literature on performance measures as well as in the literature on workarounds was the issue of systems, performance measure and IT, not being able to fully capture the complexities of tasks and situations for which they were designed for. As a result the actors in those systems often altered their behaviour to fit the requirements of the systems rather then completing their tasks, as they may have originally perceived them. In this section the concept of perceived complexity of tasks will be explained and discussed using the cybernetics notion of variety as proposed by Hamilton (1998). In addition the concepts of system and actor's point of view of the task will be defined and explained.

Complexities of tasks can be defined in terms of the number of the subtasks and the interdependencies between those subtasks as well as in terms of the number of different possible situations an actor might have to handle or respond to in order to complete the task or achieve the goal. In cybernetics, a field of science focused on "… control and communication in the animal and machine…" (Weiner 1948), of the number of possible situations or states a system can be in (Beer 1974) are referred to as variety. Thus it can be said that in the case of performance measure systems and IT systems described in literature review, those systems did not capture or did not perceive the entire variety of the tasks and situations at least not as much variety as the actors who operate within the systems perceived.

Hamilton (1998) concludes that workarounds are developed as a reaction to the variety-handling gap between the total variety occurring in the environment, which is perceived by the actor, and the system's variety handling capabilities. Her suggestion that workarounds are a tool that bridges that gap by increasing the total response variety and thus allowing system users (the actors) to achieve their goals, and fulfil their duties is consistent with one of the seminal laws in cybernetics, namely Ashby's Law of Requisite Variety. This law states that for any given system to be in the stable state (in control) the system needs to posses the same amount (or greater) of response to the possible states of the environment, as there are possible states of that environment. That is, a system needs to have a requisite verity in order to stay in control. Only verity can destroy variety (Ashby 1956). This law is illustrated by the following equation:

$$V_R \geq V_I + V_E$$

Response Variety ≥ Internal Varity generated by the system + External Variety generated by the environment.

In order to better understand the role this variety perception gap has in the understanding of the studied phenomenon, Ashby's approach should be used to investigate how those systems came to be designed.

According to Ashby's Law, in order to stay in balance a system must address all the external variety generated by the environment and internal variety generated by the actors. Failure to respond or to control the variety hinders a system's efforts to reach its goals. One way of achieving requisite variety is to reduce the internal variety. This can be achieved by implementing a system of rigid connections between the system's users (Beer, 1974). Those rigid connections take the form of rules, procedures and guidelines, which have an ability to constrain, otherwise unstructured, interactions between the actors and thus reduce the amount of variety system generates. Effort is made to ensure that every actor deals with a given disturbance in the same fashion, thus reducing internal variety. For example, rules and procedures on how to handle purchasing of office supplies reduces the variety of requisition forms a purchasing assistant must handle and thus enables the assistant to focus on the task of procurement. The rules, the procedures and the guidelines describe to the actor the system's view of how a task is to be performed and what the goal is. The description often involves a sequence and a manner in which a subtask or task must be performed. This sequence of activities (tasks or subtasks) charts a path to the goal as seen by the system, and it will be referred to as a "system-prescribed path".

Using Ashby's approach it can be said that the rules and the procedures describe to the actor how to respond to a given disturbance in order to reach system's goal. This variety-handling model is shown in the figure below.

Figure 1: Variety handling: the system's point of view

Response (rule to be applied)

Disturbance	1	2
A	X	
B		X

X- Systems desired response

The goal of a system is to address all disturbances with the appropriate responses. The reason for implementing all the rules and procedures is to reduce the variety the system generates. This is achieved by controlling the behaviour of the actors through those rules and procedures as described by Beer (1974). However, it is rather impossible to predict all the disturbances and create a rule for each single one of them. Furthermore this would be counterproductive, as the solution of one unique rule for one unique disturbance would not reduce the variety within the system. Some compromise between the need to reduce the variety and the need to properly respond to disturbances must be reached. Therefore all the plausible disturbances are grouped into categories of disturbances and then specific rules are created, which describe a single set of procedures that should be applied to the entire category of disturbances. The system then creates a control mechanism whose role is to ensure that the actor stays on the system-prescribed path (performs system-proscribed sequence of tasks in the system desired manner).

How the system, through the system's designer, chooses to group the disturbances and subsequently how it chooses to monitor the adherence to the system-prescribed path become crucial factors in the studied problem. Both of those issues contribute to the problem of systems not being able to capture the true complexities of situations. In many cases the categories of disturbances created by the system and the corresponding set of rules and indicators do not map very well onto the actor's perception of the task. Actors, due to their experience for example, may perceive significant differences between the disturbances within a single category as created by the system and as such they may have a different idea of how to handle each one of them.

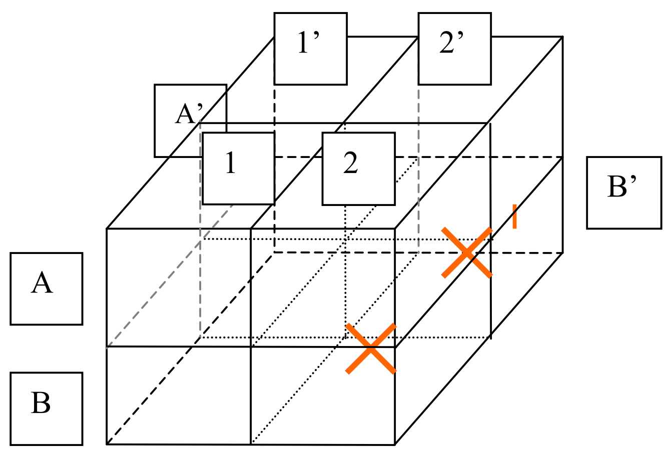
For example, a selection of inadequate performance indicators, which often is a result of yet another compromise between what can and what should be monitored as illustrated by Blau (1955) reduces the "systems perception" of the task complexity and subsequently its ability to ensure that the original goals of the systems are being achieved. The result is a situation where from the actor's point of view there are many numbers of paths leading to a number of solutions, which from the system's performance measure mechanism point of view are not distinguishable from the system-prescribed path and solution. Using Blau's study as an example, the task of matching appropriate people with the appropriate job offer is complex. However the metric used to evaluate their this task, the number of interviews arranged, not only did not capture the complexity of the task but also created a situation where the system would see no difference between the system-prescribed path of a good match between the candidate and the job and the alternative path of sending just about anyone to any interview.

The diagram below, using the above-described study as an example, helps to illustrates the model of the actor's point of view in handling the task variety with respect to the system's view of the task

Figure 2: Variety handling: the actor's point of view

<u>Description:</u>

Numbers 1 and 2 represent the responses as seen by the system, while 1' and 2' represent the additional responses as perceived by the actor. Similarly A and B are the disturbances as perceived by the system and A' and B' represent the additional variety captured by the actor and not by the system. In Blau's example A and B refer to different job candidates and their skills while 1 and 2 refer to appropriate course of action as defined by the rules. In this example goal of the system is to provide unemployed people with a job, the system-prescribed path for the actor is to match the applicant with the appropriate job. However due to the selection of performance indicator the actor noticed that the system is unable to distinguish between the action of sending the right person (disturbance B) to the job interview (response 2) or sending not fully qualified person (disturbance B') to the same job interview (response 2').



X – System desired solution
X'- Actors solution that is indistinguishable form the X in the system's view

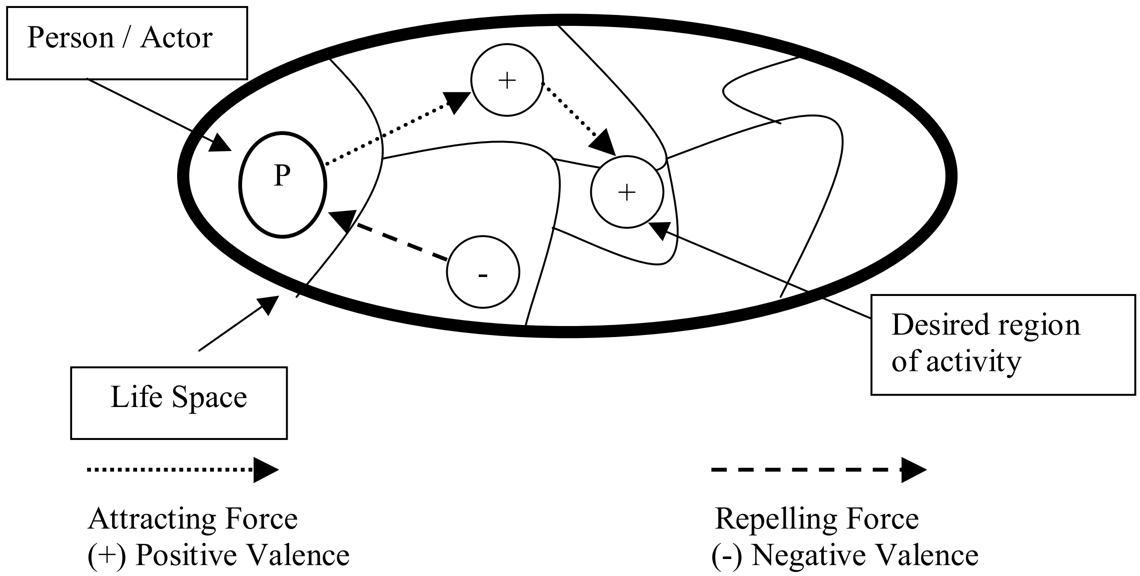
3.2 The Role of Valence

As was described in the previous section due to the fact that some system do not perceive the same amount of variety in the task as actors do, the actors while negotiating their way through the system, often arrive at a point in time when they have to make a decision on how to proceed. This choice situation is created by the actor's perception of existence of more paths then just the system-desired path.

In one of his models of behaviour based on the Vroom's Expectancy theory Lawler (1976) suggest that a valence perceived by the actor and associated with the outcome or the reward plays a key role in the process of choosing which behaviour to engage in. Lawler (1976) suggest that the actor's decision to perform in a certain way is an outcome of two factors. First, the actor must believe that he/she will be rewarded for the effort. Second, the actor must desire the reward in other word the rewords must be perceived as heaving positive valence. The notion of valence that Lawler (1976) uses in his model refer to a concept developed and studied by Kurt Lewin.

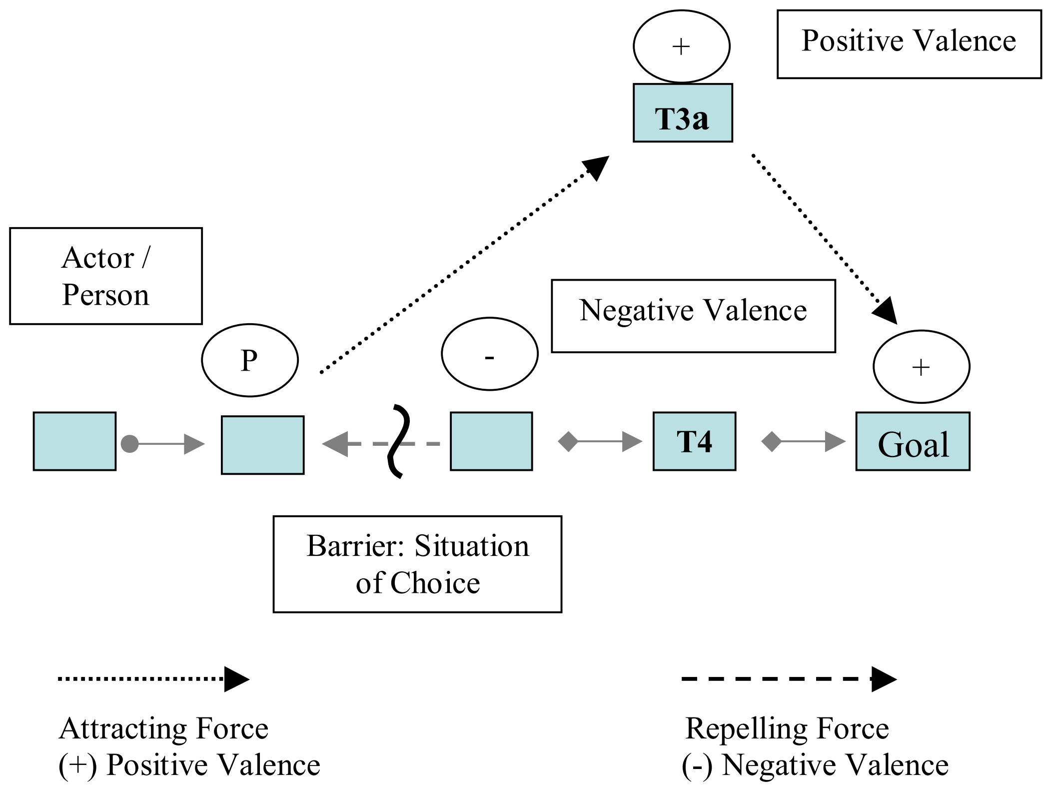
Kurt Lewin, a social psychologist and the creator of the Topological Psychology, proposed a theory of human behaviour based on the existence of a "Life space" which, "…is the total psychological environment which the person experiences subjectively"(Marrow, 1969). The life space, as described by Lewin (1936), is divided into various regions of activity and the person is occupying at least one of those regions at all times. (Refer to Figure. 3) For example, the life space of attending university might be divided into attending classes, writing exams, socializing, doing homework etc.
Each of the regions of activity has valence attached to it. The notion of valence can be compared to relative desire or attractiveness of associated with a given region. The regions of activity that are perceived to be desirable, for example receiving a good mark on the field report, have positive valence. On the other hand regains of activities that are not desired, writing exams or doing homework, have negative valence. If a given task or a region of activity has a positive valence one is attracted to it and depending on the magnitude of the valence will be compelled to move toward that region. Conversely if a region of activity has a negative valence, one is being repelled from that region. If such undesired region stands between the region this person is currently occupying and the desired region and this person perceives several additional regions of activity around him/her than such person will chose to proceed to the region that as a result of all the acting forces has the highest positive valence or the least negative valence. Figure 3 below illustrates this situation.

Figure 3: Life Space with Regions of activity.



This situation of movement and choices between the regions of activity is analogues to the situation of an actor moving along or embarking on the system-prescribed path and encountering a subtask, which he/she finds undesirable. The system of valences associated with: system-desired path, alternative paths perceived by the actor and the goal or objectives of the individual creates a form of a penetrable psychological barrier, which causes the actor to choose between the system-prescribed path and the alternative path or paths. Similarly to the case of movement between the regions of activity the actor, after consideration of all options involved, will choose a path that he/she perceives as having more positive valence than the others. Figure 4 helps to illustrate that situation.

Figure 4: System of valance forces perceived by the actor: Alternative-Path example



Legend:

T1, T2, T3, T4 – tasks along the system-prescribed path
T3a – alternative task creating alterative path

Although there are many reasons why an actor could have a feeling of negative valence towards system prescribed path, three major reasons capturing most situations will be described. First, the actor who perceives more verity in the system then the system can capture, can believe that he/she knows as better way of responding to a particular disturbance than the rules of the system dictate. In such situation the negative valence associated with the prescribed path comes from the feeling that a different, and in the mind of an actor, better path can lead to the same goal as the system-desired path. Thus not only the reward can be a source of valence but also the path itself. An easier path leading, even to a lesser, reward might be perceived as having higher valance then (less desirable) system-prescribed path leading to a greater reward. It is also important to note that such factors as ethics or idea of loyalty also can influence actor's perception of valence associated with a given path. The selection of an alternative path may be viewed as validation of ethics or breach of trust, which in turn can be a source of negative valence.

Second, the actor might be simply unable to follow the rules and procedures as prescribed by the system. The reasons can be many: for example the actor might not posses the necessary skills, the systems might ask for information that is not available or not in the possession of the actor (Lawler, 1976; Argyris, 1971) or the system might be designed inadequately, creating a situation when there is not enough functionality in the system itself to reach the stated goal (Hamilton, 1998). In any case the negative valence creates a psychological barrier that repels the actor from the system-prescribed path.

Third, a classic case of a goal displacement can occur as a result of the implementation of the performance measure. The control mechanism is often tied in with the reward mechanism. The actor, who feels a very strong valence towards the reward, might see the system-desired path as not the best one leading to the achievement of that goal. Combined with actors perception of alternative paths and his/hers knowledge of the performance indicator tied to the reward, the actor might be more attracted (feel stronger positive valence) to the path that results in a greater reward yet is indistinguishable from the system-prescribed path. This is evident in Blau's study of the NYC employment agency. The social workers started to arrange interviews despite the fact that they most likely had know would not result in the employment offer. This action clearly defeated the overall goals of this public employment agency, but in the eyes of the performance evaluation mechanism the actors did their job and were rewarded accordingly (Blau 1955).

In either case there is a potential for and actor to perceived conflict between the hers/his understanding of the task and the desire to fulfill the task in a manner the actor thinks it should be fulfilled and the actor's perception of how the system defines the task and wants it to be performed. This conflict may result in the perceived negative valence toward the system-prescribed path and encourage the actor to seek an alternative path.

3.3 The Role of Path Similarity in Perception of Valence

An actor, who perceives alternative paths and for some reason experiences a negative valence toward the system-prescribed path, must make a decision of which of the paths to choose. As was suggested in the previous section the in general the actor would choose a path that overall has the highest positive valence. An important factor in an actor's perception of valence for a given path, and consequently the selection of a path, is the perceived similarity of outcomes between the alternative paths and the system-prescribed path. One of the first steps in the path selection process is categorization of alternative paths into the ones that meet the minimum similarity standard and those which do not. In this process, the actor does not use his or her own perception of similarity but rather tries to categorize the paths using what he/she believes is the similarity as perceived by the system's performance measure mechanism. Therefore, this judgment of similarity of outcomes depends on the context provided not by actor per se but by the actor's model of performance measurement and knowledge performance indicators used.

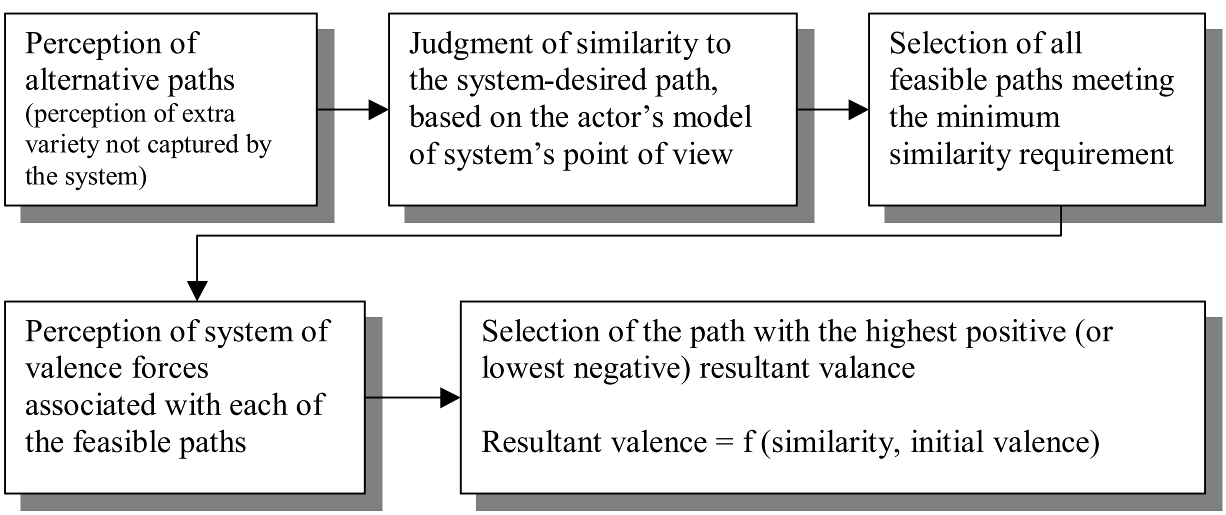
It has been indicated in several research papers that the perception of similarity between objects is very much context-dependent (Herdiman et al, 1989). Goldstone (1994) paraphrases Goodman (1972) who argues that any given object X is similar to another object Y only with respect to some property Z. "This object belongs to category A because it is similar to A items with respect to the property 'red.'"

Barsalou (1983) through empirical research has shown that even two seemingly not similar objects such as 'raccoon' and 'snake' or 'children ' and ' jewellery can be judged as being similar if an appropriate context of ' pets' or " things to retrieve from burning house," respectively, is supplied.

Tversky (1977) in his research on features of similarity found that the features upon which categories are created have two components to their measure: First the intensity (for example brightness of a colour); Second, the diagnosticity – the classificatory significance of feature. While the intensity is a function of the perception and cognition and "is relatively stable across the context", the diagnostocity changes with the context and can form a basis for new categories. In general, Tversky (1977) proposes that "similarity has two faces: casual and derivative," meaning that that similarity can serve as a basis for classification but it is also influenced by the imposed or "adopted" classification. Furthermore, it has has been shown that the level of knowledge about a given task as well as the goal of the task can influence the judgment of similarity (Suzuki et al, 1992).

Thus in the mind and perception of the actor two paths have similar outcomes if, for example, the actor knows that the performance measure mechanism cannot distinguish between the two outcomes even if the actor can as it was the case in Blau's (1955) study. It can be concluded therefore, that the resultant valence which determines the final decision to either take the alternative path or to remain on the system-prescribed path is a function of the initial valence experienced by the actor and the valence created by the judgment of similarity. That is to say that given two alternative paths of otherwise equal perceived valence, the actor will choose the path that he/she deems to be most similar to the system-prescribed path in the "eyes" of the system. Moreover, an actor given two alternative paths of the same similarity of outcomes would select the path with the higher positive valence. Figure 5 summarizes the path selection process.

Figure 5: The process of path selection



It can be argued, therefore, that the actor equates the entire idea of performance measurement as a process of similarity judgment that the system makes. The similarity judgment of how closely the outcome of a given task or sequence of tasks (path) matches with the system's desired outcome of the system-prescribed path. This view of performance measurement held by the actor , combined with hers/his knowledge of the specific performance indicators used by the system to perform the above described similarity judgment, influences the actor's choices of a manner in which she/he will perform the task. In other words which path the actor will choose.

3.4 The Thought Experiment

To further illustrate the mechanics of this phenomenon, imagine a simple thought experiment. A given system requires its user (the actor) to provide a red rectangle as an input. Let us also imagine that it is extremely difficult to measure both the dimensions of shape and colour at once. Therefore the system (or whoever is designing it) chooses the dimension of colour as a performance indicator. In another words the system will check only the colour of the objects and it will punish or reward the actor accordingly. Thus from the systems point of view there are two kinds of inputs (disturbances) red objects and other colour objects. It is important to point out that the systems does not disregard the shape per se, but rather it assumes that the actor will only consider rectangles. This assumption is illustrated in Blau (1955) where the metric of "number of interviews" was most likely chosen with the assumption that the right job candidates would be sent to the interview. There are also two responses from the system's point of view, to accept the object as an input when the object is red and to reject the object otherwise. Let us further imagine that the actor has only two available objects a red triangle and a brown rectangle. Additionally let us assume that the actor must provide the system with an input and that the reward that the system offers for the correct input has a positive valence for the actor

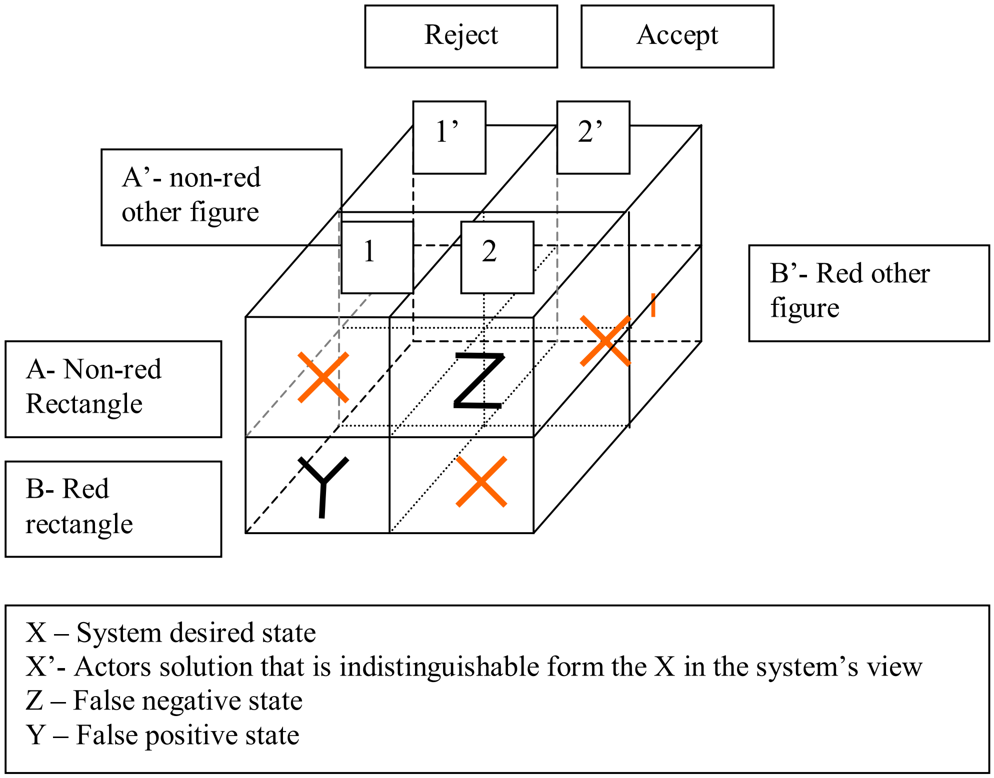
What emerges is a situation in which the system is not able to capture the full complexity (variety) of the task of selecting the colour and the shape of an object. In this case it is a result of the implementation of the performance measure mechanism, which reduced the disturbance verity from two dimensions to just one. The actor on the other hand clearly sees that the available objects don not only come in red or other colour but they also have a shape. Thus the actor perceives extra variety not captured and therefore not detectable by the system.

At this point in time the actor must make a choice on how to proceed. The actor must choose one of the objects as an input; however none of the objects is a red rectangle. As it was suggested before the actor is attracted to the reward (positive valance), and he/she also knows how the system evaluates the input. Thus what the actor will do is to compare the outcomes of each of the available paths to the outcome of a system-prescribed path. In this case the actor will perceive that in the system's point of view red triangle offers most similar outcome to the red rectangle as the performance measure controls only the colour and cannot distinguish between the shapes. The diagram below illustrates this gap in perception of variety and actors choice.

Figure 6: Though Experiment: Variety perception gap between system and actor

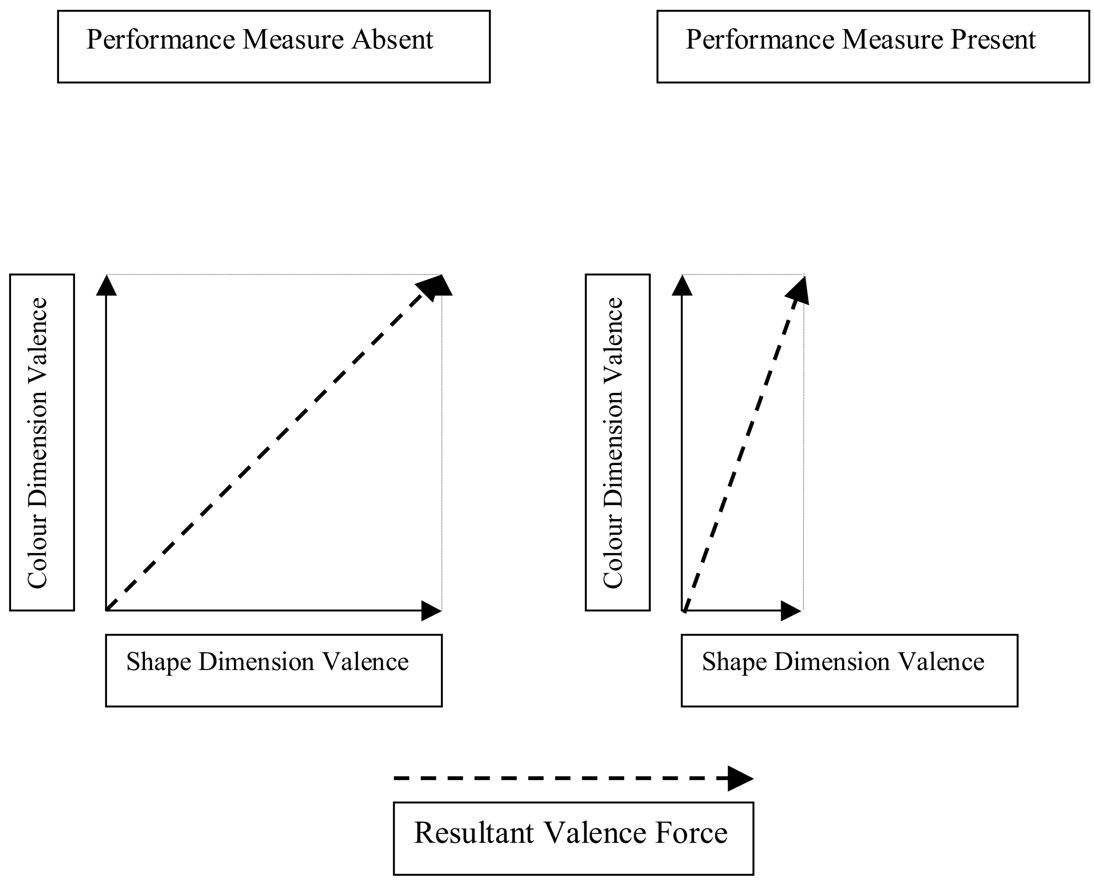
<u>Description:</u>

- A and B represent the disturbances as seen by the system. A is a non-RED input and B is a red input (system assumes that actor considers only rectangles)
- A' and B' represent additional disturbance perceived by the actor but not by the system. In this case it is a dimension of shape. A' is a non-Red non-Rectangle object while B' is a red non-rectangle.
- 1 and 2 represent the system responses, reject the non-red rectangle and accept the red rectangle respectively. In this particular example the actor is supposed to respond with 2 to disturbance B and with 1 to disturbance A
- 1' and 2' are additional responses as perceived by the actor, reject the object as an input and accept the object (red but not a rectangle) as an input respectively



To better illustrate the impact of the performance measure on actor perception of valence and hence the actor choices lets first consider the same system without the performance measure. In such case it is not unreasonable to assume that, for the actor, both features of colour and of shape would play an equal role in the selection of the input. This means that he actor would place equal weight on the colour and on the shape dimensions when selecting an object. In the case of this though experiment brown rectangle might have been chosen over the red triangle as brown can be considered closer to red the triangle to rectangle. When the performance measure mechanism is installed, however, and the reward is now closely attached to that performance, the dimension of colour becomes more important to the actor and the dimension of shape much less important. In Lewin's terms there is a stronger valence associated with the colour than shape for the actor. As a consequence the resultant force increase the valence associated with the colour of an object. The colour of the object also becomes the seminal feature with regards to which the actor will make the similarity judgment between the alternative path and the system-prescribed path. In the end the actor is compelled to select the red triangle a path that has the highest resultant valence. Figure 6 below helps to illustrate that impact of the performance measure on the perception of valence.

Figure 7: Impact of performance measure system on actor's perception of valence



The final choice of selecting the red triangle over the brown rectangle in the case of this experiment will allow the actor to reach his/her objective of being rewarded. It is important to notice, however, that in this thought experiment the final use the system has for the object is not specified. If the shape of the object did play an important role, despite the fact that the feature was not being monitored, then actor behaviour would be considered unwanted or dysfunctional just like the behaviour of the insurance adjustors described by

Irving et al (1986). Conversely, if the shape did not play such significant role for the system than actor's behaviour could be classified as a case of a workaround not unlike the cases presented by Hamilton (1998).

This vector representation is only a simplified and purified example of the impact of the performance measurement on the perception of valence associated with one of the object's features. It is not meant to suggest that these two dimensions, or any other features an object might have, are orthogonal in the mind of the actor and that they fallow Euclidian geometry.

3.5 Conclusions

The performance measurement mechanism has a crucial influence on actor's perception of the task to be performed and the actual performance thereof. All goal oriented systems, to remain stable, need to maintain the control by appropriately dealing with incoming variety, both internal and external. However, to be able to maintain that control some measures of performance must exist. It can be challenging to design a performance measurement mechanism, which could adequately capture the essence of a given task. This is because of the inherent complexities involved in many tasks, situations or environments.

As a result the performance measurement systems often do not capture or choose to ignore, certain aspects or dimensions of a task, which an actor perceives and my find to be important. Furthermore those overlooked features often contribute to actor's understanding of that task and to actor's perception of how the task is to be performed.

It can be, thus, said that the actor perceives the task at hand in a more holistic manner, with many dimensions each having some inherent weight assigned to it based on the actor's perception of a relative valence associated with each of those dimensions. All of the dimensions influencing the manner in which the actor understands the task and will perform that given task (the path the actor embarks on).

The performance measurement system, on the other hand, captures or is design to capture only a sub-set of the features and dimension of a given task. This may be because of the system's inability to capture all complexities of a given task or because the selected performance indicators are believed to capture adequately the nature of that task.

No matter the reason, however, the system often assumes that there are less sources of valence present, or that the only relevant sources of valence are the ones the system generates itself by the introduction of rewards and punishments.

It is likely, than, that a discrepancy exists between actor's perception of the task and the actor's belief (model) of how the performance measurement system perceives the same task .This discrepancy or a perception gap may lead to a situation where an actor perceives a conflict between his/hers understanding of how to perform the task (which path should be chosen) and his/ hers model of systems perception of that same task manifested by the system-prescribed path.

As suggested in section 3.3 the process of performance evaluation can be defined as similarity judgment, at each step of the path or the final outcome, between the actor's actual performance and the system-desired performance of the sequence of tasks comprising system-proscribed path.

The actor's knowledge of how the system judges the performance (i.e. judges similarity) combined with the actors knowledge of the fact that the system does not capture some dimensions/features of a given task, enables the actor to perceive alternative paths (alternative sequences of tasks) that are equivalent form the point of view of the system. It also enables the actor to choose the alternative path that satisfies the requirements of the performance measure yet is perceived by the actor as more attractive than the system proscribed path. Provided, of course that the actor's understanding of the performance measurement system is a correct one.

The introduction of the performance measure, however necessary it is, distorts than the original systems of relative valences associated with many dimensions/ features of a given task, creating now resultant valence force (as it is illustrated in Figure 7). Since, as this thesis argues, the way the task is performed (the sequence of the task creating means and end chains, where the outcome of one sub-task becomes the input to the next, and all together constituting a path) is a function of the resultant valence force perceived by the actor, the introduction of the performance measure changes how the actor's perceives and hence performs a given task. This is illustrated by the though experiment and the actors selection of the red triangle over the brown rectangle.

The change in the resultant valence force, changes the basis upon which the actor makes his/hers similarity judgments between the system-prescribed path and alternative paths. Thus, overall, the changes in valance are affecting the way the task is being performed and the outcome of that task. That is to say the manner in which the actor will perform the task is going to be influenced by both the system of perceived valences associated with the task itself, (actor's initial understanding of the task) and the system of valances associated with the incentives introduced by the performance measurement, not just the incentives alone as often assumed.

3.6 General Hypothesis:

The theoretical framework brings on some general assumptions and some general hypothesis that can be made with regards to the proposed explanation of actor's behaviour and the studied phenomenon.

Assumtion1: Independent of any performance measure actors would perceive the task at hand in a holistic manner assigning some weight to all of the features or the dimensions of that task based on his/hers perceived inherent valance associated with each of those features or dimensions.

Assumption 2: The introduction of the performance measure adjusts the existing system of perceived valences by adjusting the relative weights/ importance of existing salient features of that task or by introducing new sources of valence thus influencing the actual performance.

H1: The manner in which the actor performs a given task (the path the actor chooses to follow) is a function of resultant valence forces composed of the system of perceived valence forces associated with the salient features or dimensions of a given task.

H2: Changes (increases and decreases), influenced by the performance measurement, in the perceived relative valences associated with a given dimension of the task will result in the corresponding changes in the degree this dimension is used by the actor to evaluate the similarity of outcomes of hers/his actions and the system proscribed path.

It should be noted that implicit in this hypothesis is that the non measured dimensions also have an influence on the actor's performance.

Chapter 4: Methodology

4.1 Introduction

This experiment was developed to test general assumptions and hypothesis. The design of this experiment was largely influenced by the thought experiment described in the theoretical framework chapter.

In the experiment, subjects were asked to complete two or three tasks. Each task involved reading specific instructions and then ranking several objects on a page based on the instructions. The rankings of those objects by each participant were recorded in experimental booklets.

The description of the experiment, results and a discussion of the results are presented in the following sections of this chapter.

4.2 Subjects:

The 191 subjects that participated in this experiment were undergraduate students of the University of Waterloo. Seventy-two subjects were from Management Science 211 course (Organizational Behaviour) and 119 subjects were from Management Science 311 course (Organizational Design and Technology). Majority of the subjects were second and third year engineering students with the remainder being predominantly second and third year students from science and arts faculties. Three subjects were fourth year psychology students. As the subject matter or true purpose of the experiment was never revealed to them, this group can be described as naïve with a very limited knowledge of the actual concepts being tested.

All subjects were volunteers who received 3 bonus marks in their respective course for participating in this experiment

4.3 Stimulus Set:

The stimulus set consisted of a combination of six colours and six figures. Special care was taken to ensure that all figures are of roughly the same size so that the test objects can by judge only between two dimensions of shape and colour.

Two separate sates of stimuli were created. One set for the control group and another set for the experimental group. Both sets were assembled using the same shapes and colours but in a various and distinct combinations. All stimulus pages were prepared using Adobe Photoshop 6.0.1 CE.

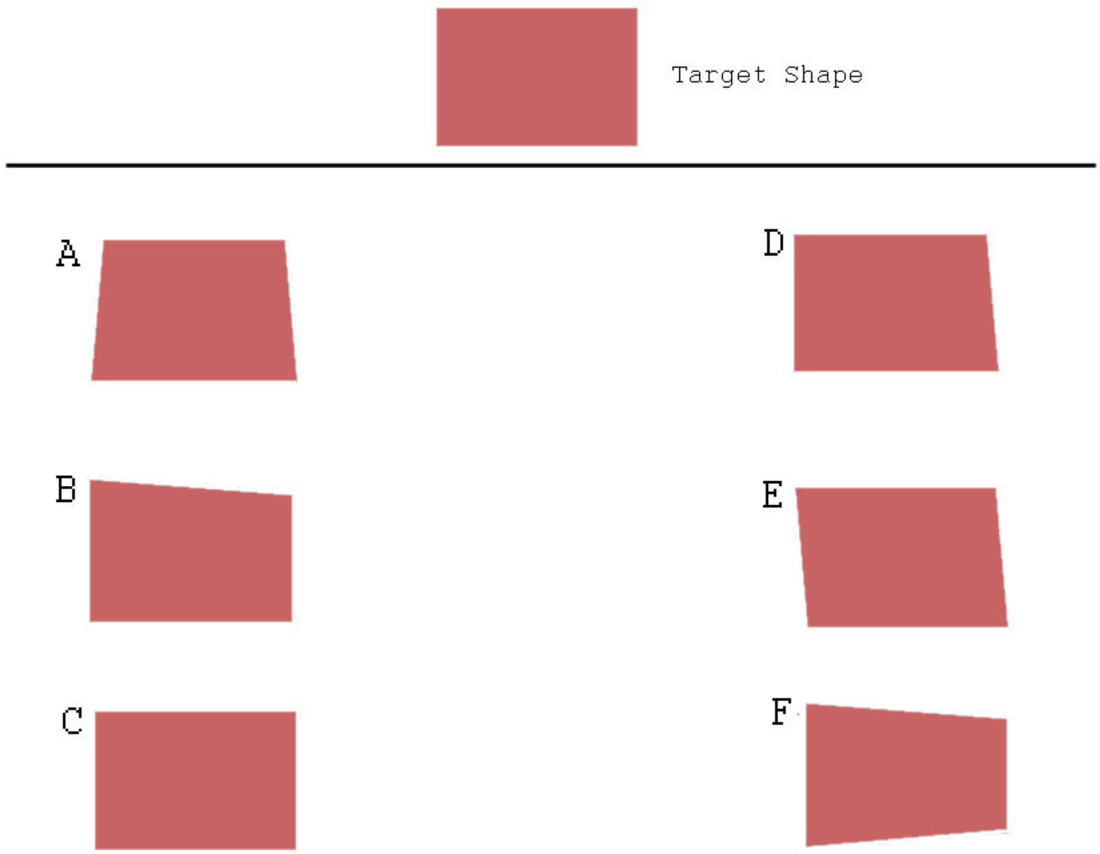
4.3.1 Pilot Run:

A small pilot exercise was conducted using an initial version of the experimental booklet containing both the graphical stimuli and the instruction sets. Groups of 6 graduate students from the department of Management Science were asked to do the entire experiment. Afterwards they were asked questions with regards to the clarity of the instructions, the quality of the graphic stimuli and general difficulty of completing the experimental tasks. The comments and suggestions gathered from the pilot exercise subjects' were implemented during the development of the final version of the stimulus. Some of the comments made by the pilot subjects will be presented in following sections.

4.3.2 Selection of shapes and colours:

The "target" shape, the rectangle, was inspired by the thought experiment. The remaining five shapes were created by changing various angles, of the original rectangle, by 5 degrees. This was done so that all 6 shapes look somewhat similar and none appears to be significantly bigger or smaller then the rest. For purpose of identification each shape received an alphabetical code from A to F (Refer to Figure below).

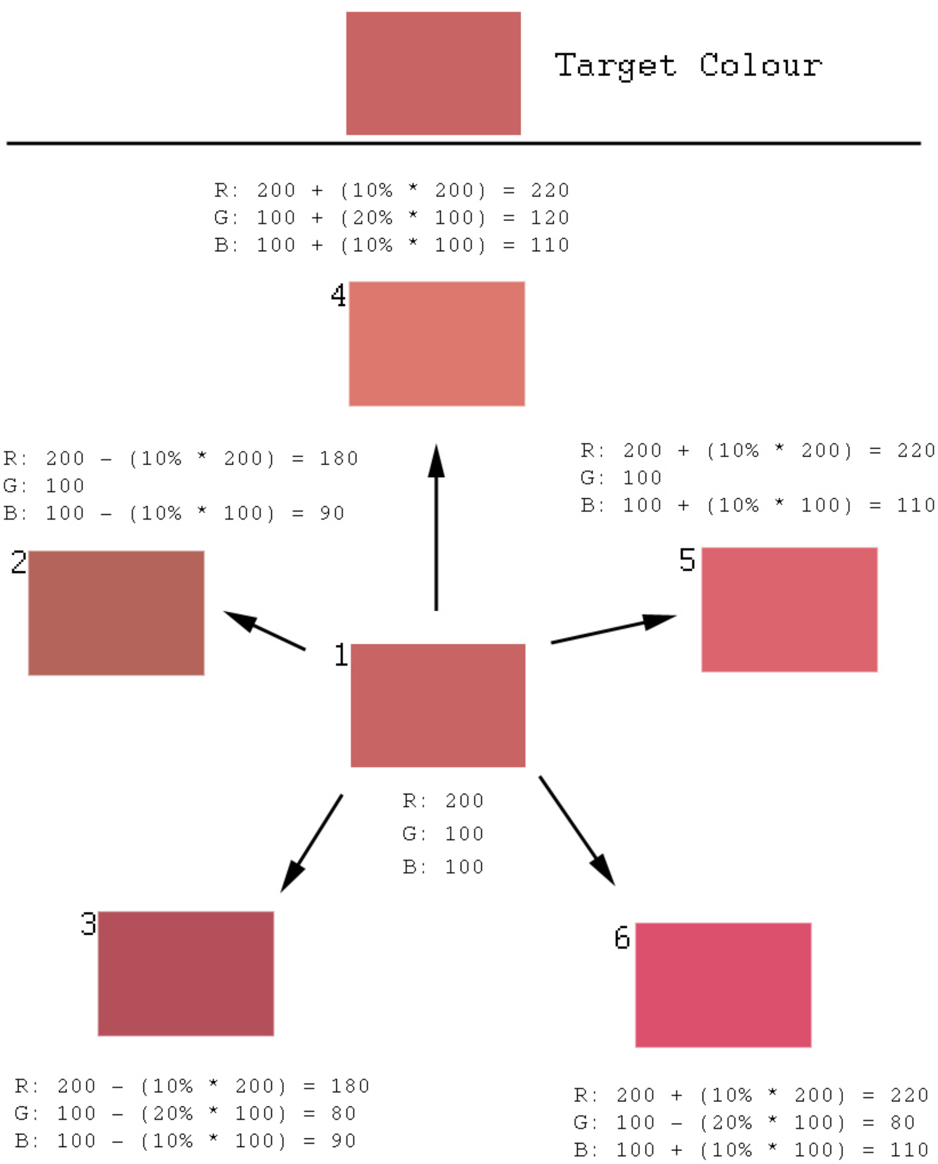
Figure 8: The Stimulus: Shapes



The “target” colour, red, was also inspired by the thought experiment. It was selected from the middle of the red color spectrum using Adobe Photoshop 6.0.1 CE. The remaining 5 colours were created by manipulating the Red Green and Blue (RGB) dimension. Manipulation was done in such way that, each colour differed from the “target“ colour by no more than 20 % in Red, 20% in Green and 20% in Blue dimension (Refer to Figure 9). This procedure allowed for creation of all six colours, which were similar to each other, and all appeared to be within the red light spectrum. Initially the idea was to create the colours, which would be of the same Euclidian distance in the RGB dimensions from the target colour. However some colours created that way appeared to be significantly outside of the general “red colour” category as reported by the pilot run subjects. For those colours the G value was changed to match the value target colour.

For the purpose of identification each colour was assigned a numeric code from 1 to 6. Colours 4, 5 and 6 are located 30 units in the Euclidian (in RGB dimensions) distance form the target colour and colours 2 and 5 are located 22.36 units from the target colour.

Figure 9: The Stimulus: Colours



In the thought experiment the input choices of red triangle and brown rectangle were perceptually in different groups/ categories of objects. The similarity of each shape and each colour was created so that all the objects fall in the same category of "red objects" or "rectangle-like objects". The reasoning behind making all the objects fall into one general category of shape and one general category of colour was to mimic real life situations where available alternatives (solutions etc.) are similar and usually fall into the same category. In addition, consideration was given to development of a believable and plausible background story (refer to section XXX for the story).

4.3.3 Control Group Stimulus Set:

For the purposes of the control group a page with six shapes of the same target colour and another page with six colours in the same target shape were prepared. Both pages featured the objects arranged in two columns, of three objects each, located in the middle of the page. In all cases the target shape or the target colour appeared on the top of that page in the same location every time (For illustrations refer to Appendix A). Changing the location of the shapes and colours on the respective pages created two versions of each page. The target shape/colour remained the same and in the same position in each version. This change of location allowed for control in case the objects location on the pages affected subject's perception of colour and shape.

4.3.4 Experimental Group Stimulus Set:

The Experiment Group's stimulus consisted of 36 objects, all combinations of the 6 shapes and 6 colours used for the control group. The target object, called the "perfect sample", which appears on the top of the page, is the same object (red rectangle) as the "target object" for the Control Group.

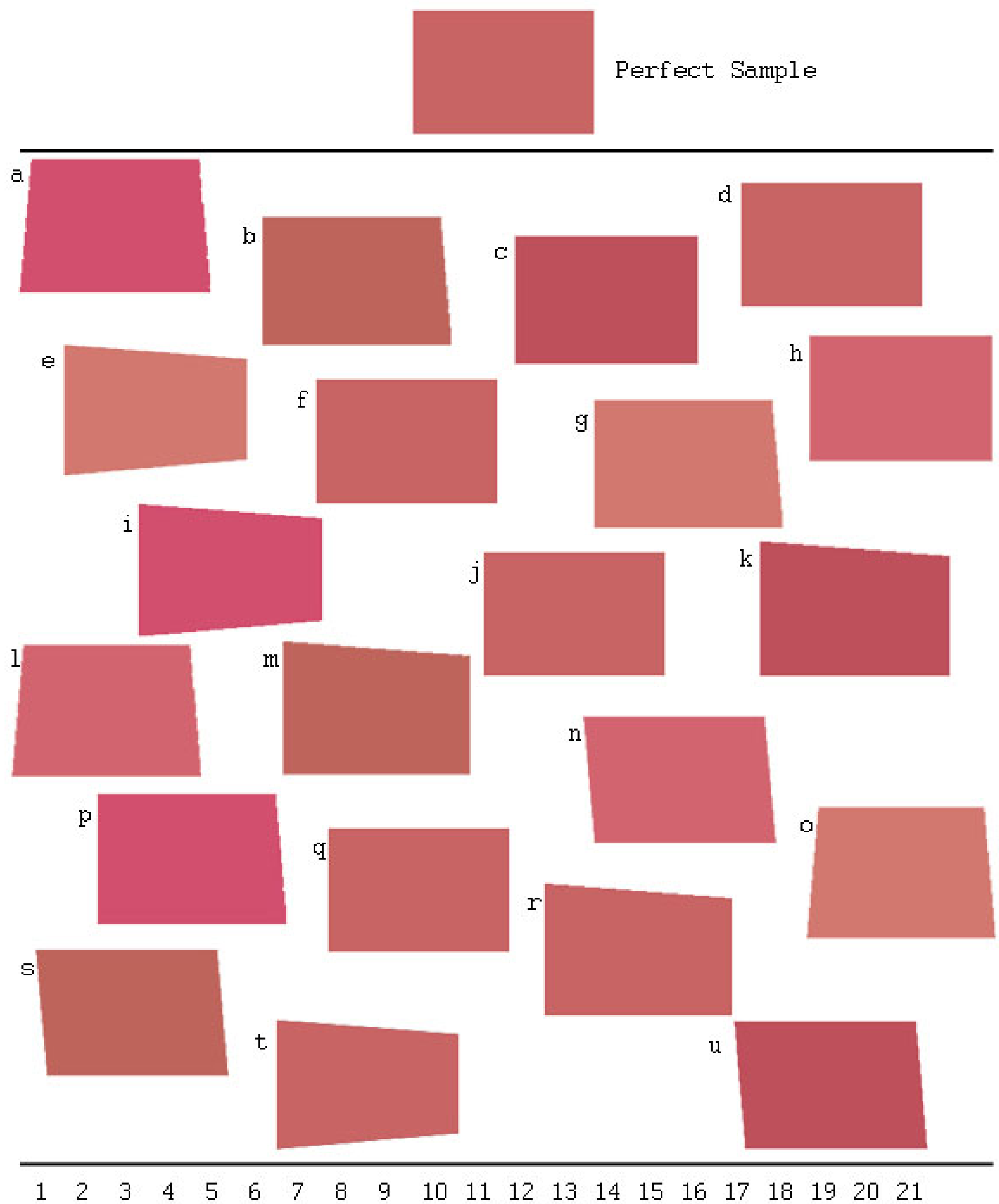
Originally all 36 objects were placed on one sheet of paper. This was done because the experiment was to be conducted in a classroom during the tutorials with many subjects at once. However the pilot study subject's had indicated that having 36 objects on one page was too many to focus on, and that the objects themselves became to small to notice crucial differences in colour. That is why the 36 objects were split into two equal groups A and B. Each set of 18 objects had equal representation of each shape colour combination. Since only set B contained the target object, it was decided to add 3 instance of the "target object" into each set. This created the final 2 sets of 21 objects:

- Stimulus Set A: consisting of 18 unique objects and 3 instances of "Perfect Sample"
- Stimulus Set B: consisting of 17 unique objects and 4 instances of "Perfect Sample"

Initially the objects on the page were displayed in rows and columns. After pilot run was performed the pilot run subjects reported that some objects' proximity to one another made them look significantly different from the "perfect sample" object. It was suggested that this visual effect was an outcome of the way the object were located on the page. The stimulus page was changed and all the objects were placed in a random scattered pattern. The subsequent test study indicated that previously reported visual effect was no longer present in the new stimulus pages.

Once again to control for potential location bias, two versions of each set were created by manipulating the location of the objects on the page. The following figure shows the sample experiment stimulus set in reduced size. For the entire set in actual size as seen by subjects during the experiment please refer to Appendix: A

Figure 10: Sample of experimental stimulus sets A reduced in size by 35%



4.4 Instruction set:

A page of specific instructions preceded each stimulus page and together they constituted a task. Each subject had to complete two or three different tasks as a part of the experiment (for the complete list of task combination refer to procedure sections of this chapter page 56).

Each instruction set was assigned an alphanumeric code. This section lists and explains all the instructions as they appeared in the experimental booklets used in experiment. It should be noted that the initial versions of some of the instruction sets had to be changed as a result of the findings of the pilot study. In some cases wording was changed and underlining of parts of the text was added. This was done in order to direct the attention of the subjects to important information contained within the instruction set. Information that was missed by the pilot study subjects while reading the instructions was added or clarified.

The general premise behind the instruction sets is based on the theory formulated in the theoretical framework chapter. As it was illustrated by the thought experiment, the control system has a potential to influence the outcome of the resultant valence force. This influence arises from the fact that performance measurement mechanism increases the valence in the direction of the feature chosen as a performance indicator. At the same time, the performance measurement may decrease the perceived valence associated with any other feature of the object, which might be vital, in reaching the goal, but is not being measured. As a result the actor adjusts his or hers behaviour according to the new valence.

Each instruction set is designed to either detect the original system of valences, for the two experimental dimensions of colour and shape, or to influence the valence on one of the dimensions (refer to theoretical framework chapter for additional explanation and illustration)

4.4.1 Instruction T1

This instruction set was created for the control group and has two versions “a” and “b” for colour and shape respectively. This task was created to get the control similarity ranking of six colours and six shapes with respect to the “target” colour and “target” shape.

Instruction T1a:

On the next page you will find 6 colours (do not turn the page yet). Your task is to rank these 6 colours in order of similarity to the target colour shown on the top of the next page.

Instruction T1b:

On the next page you will find 6 shapes (do not turn the page yet). Your task is to rank these 6 shapes in order of similarity to the target shape shown on the top of the next page

4.4.2 Instruction T2

This instruction was created to get the naïve similarity ranking of the 21 objects on a page. No explanation of similarity is given and it is up to the subject to determine how much emphasis to put on each dimension (shape vs. colour).

On the next page you will find 21 objects (do not turn the page yet). Your task is to rank these 21 objects in order of similarity to the “Target Object” shown on the top of the next page.

Two objects can be assigned the same rank if you think they are equally similar to the "Target Object"

For the purpose of identification this task was called "T2- Naïve ranking"

4.4.3 Instructions T3 and T4

The story of a toy factory was introduced as a background for the task. The stimulus objects looked like children's playing blocks and it was not unlikely that a machine with some form of human supervision would have produced such blocks. The purpose of the story was to test the effect of the performance measure mechanism on subject's perception of the task. In this case an idea of quality is introduced to the story. The quality is defined in terms of the two dimensions of the objects (colour and shape), and the over all instructions aim at equating the idea of quality with similarity to the "perfect sample"

The last part of the instruction was developed to place valence on the colour (T3) or shape (T4) dimension. The subjects were supposed to judge the similarity to the "perfect sample" more on either colour or shape. It was expected that in the extreme case some subjects would totally disregard one of the dimensions.

In addition, this section was designed to ensure that each subject takes the task seriously. At the beginning of the experiment subjects were told that they will receive up to 3% bonus marks depending on the quality of their answers. If their answers were of poor quality they could receive no bonus marks at all. Such a system induces real valences associated with the experimental task. Although it was not revealed until the end of all experimental runs, all subjects were, in fact, awarded 3% bonus marks regardless of their answers.

Instruction T3:

You are an employee in a toy factory that assembles block sets for kids. The machine you work with is fully automated and it is set up to produce one type of block called red rectangles as shown on the next page. The machine is not perfect, *however, and the quality of its output deviates a little bit in terms of the colour and the shape of each individual block.*

Your task is to rank the 21 blocks you will see on the next page, in order of highest quality, by comparing them to the perfect sample at the top of the next page. Two blocks can be assigned the same rank if you think they are of equal quality.

To evaluate your performance an expert has used a colour sensor to come up with an objective quality ranking by comparing the colour of each of the blocks to the perfect sample. The number of bonus marks received for participating in today's experiment will depend on how closely your ranking matches the expert's ranking.

Instruction T4:

You are an employee in a toy factory that assembles block sets for kids. The machine you work with is fully automated and it is set up to produce one type of block called red rectangles as shown on the next page. The machine is not perfect, however, and the quality of its output deviates a little bit in terms of the colour and the shape of each individual block.

Your task is to rank the 21 blocks you will see on the next page, in order of highest quality, by comparing them to the perfect sample at the top of the next page. Two blocks can be assigned the same rank if you think they are of equal quality.

To evaluate your performance an expert has used a shape sensor to come up with an objective quality ranking by comparing the shape of each of the blocks to the perfect sample. The number of bonus marks received for participating in today's experiment will depend on how closely your ranking matches the expert's ranking.

For the purpose of identification this tasks were named “T3 – Valence on colour” and “T4- Valence on shape.”

4.4.4 Instruction T5

This instruction set was design to investigate the effects of the “toy factory” story on the experiment task of ranking. Equal valence was placed on both dimensions of the objects. Similarly to T2 this allowed for testing of which feature of the object will be dominant in the judgment of similarity to the “perfect sample”. Similar to T3 and T4, the last paragraph of instructions was designed to ensure that subject takes the test seriously and gives their best effort.

> *You are an employee in a toy factory that assembles block sets for kids. The machine you work with is fully automated and it is set up to produce one type of block called red rectangles as shown on the next page. <u>The machine is not perfect, however, and the quality of its output deviates a little bit in terms of the colour and the shape of each individual block.</u>*
>
> *Your task is to rank the 21 blocks you will see on the next page, in order of highest quality, by comparing them to the perfect sample at the top of the next page. Two blocks can be assigned the same rank if you think they are of equal quality.*
>
> *<u>To evaluate your performance an expert has used a colour and shape sensor to come up with an objective quality ranking by comparing the colour and shape of each of the blocks to the perfect sample</u>. The number of bonus marks received for participating in today’s experiment will depend on how closely your ranking matches the expert’s ranking.*

For the purpose of identification this task was named: “T5 – Valence on both”

4.4.5 Instructions T6 and T7

This set of instructions was design to show that the subjects are capable of perceiving the task from the point of view of the performance measure mechanism and that they are capable of mimicking the actions of that control mechanism. Once again the last part of the instruction was designed to ensure that the subjects took each task seriously.

Instruction T6:

> *Now imagine a slightly different situation. There is a toy factory that produces block sets for kids. In this factory there is a machine that is fully automated and its task is to produce one type of blocks called red rectangles as shown on the next page. The machine is not perfect, however, and the quality of its output deviates a little bit in terms of the colour and the shape of each individual block.*
>
> *To ensure the highest quality in the finished toy sets a colour sensor was installed to monitor the output of that machine.*
>
> *Today the sensor is malfunctioning. Your task is to rank order the 21 objects (you will find on the next page) as you think the colour sensor would have ranked them. Two objects can be assigned the same rank if you think they are of equal quality.*
>
> *To evaluate your performance the 21 objects on the next page have already been, historically, rated by the colour sensor before the malfunction occurred. The number of bonus marks you receive for participating in this experiment will depend on how closely your ranking matches the sensor’s ranking.*

Instruction T7:

Now imagine a slightly different situation. There is a toy factory that produces block sets for kids. In this factory there is a machine that is fully automated and its task is to produce one type of blocks called red rectangles as shown on the next page. The machine is not a perfect, however, and the quality of its output deviates a little bit in terms of the colour and the shape of each individual block.

To ensure the highest quality in the finished toy sets a shape sensor was installed to monitor the output of that machine.

Today the sensor is malfunctioning. Your task is to rank order the 21 objects (you will find on the next page) as you think the shape sensor would have ranked them. Two objects can be assigned the same rank if you think they are of equal quality.

To evaluate your performance the 21 objects on the next page have already been, historically, rated by the shape sensor before the malfunction occurred. The number of bonus marks you receive for participating in this experiment will depend on how closely your ranking matches the sensor's ranking.

For the purpose of identification this tasks were named: "T6 – Point of View (POV) of colour sensor" and "T7- Point of View (POV) of shape sensor"

4.4.6 Instructions T8 and T9

These instructions were created to further test the impact of the "toy factory" story on the subjects perception of the stimulus and subsequently the ranking thereof. For this tusk subjects were asked to rank the objects by colour (T8) and shape (T9), both instruction sets were followed by the experimental stimulus set. This instruction set was introduced in the second run of the experiment performed in Management Science 211 course. These two tests were developed to test subjects' ability to rank the experimental stimulus by only one dimension (colour or shape) ignoring the other one.

Instruction T8:

On the next page you will find 21 objects (do not turn the page yet). Your task is to rank these 21 objects in order of similarity to the "Target Object" based on the colour of the "Target Object" shown on the top of the next page.

Two objects can be assigned the same rank if you think they are equally similar to the "Target Object".

To evaluate your performance an expert has used a colour sensor to come up with an objective similarity ranking by comparing the colour of each of the blocks to the "Target Object". The number of bonus marks received for participating in today's experiment will depend on how closely your ranking matches the expert's ranking.

Instruction T9:

On the next page you will find 21 objects (do not turn the page yet). Your task is to rank these 21 objects in order of similarity to the "Target Object" based on the shape of the "Target Object" shown on the top of the next page.

Two objects can be assigned the same rank if you think they are equally similar to the "Target Object".

To evaluate your performance an expert has used a shape sensor to come up with an objective similarity ranking by comparing the shape of each of the blocks to the "

Target Object". The number of bonus marks received for participating in today's experiment will depend on how closely your ranking matches the expert's ranking.

For the purpose of identification this tasks were named: "T8- Naïve ranking by colour" and "T9- Naïve ranking by shape"

4.5 Operational Hypothesis:

Since, in general, valence a person perceives cannot be captured or easily measured the general hypothesis testing was be broken down into a series of operational hypothesis, which involve the outcome, and comparison of the experimental tasks. The notion of valence was operationalized using the incentive system (extra marks in the course) design to reward certain behaviours as described in the instruction set section.

OH1: When the valence is placed on the colour of the object then the ranking of the stimulus becomes closer (more correlated) to the ranking produced by the colour control group.

It is predicted that T8 will have the highest correlation to the colour control ranking followed by T6, T3, T2 and T5 (in no particular order), T4, T9 and T7 also in no particular order.

OH2: When the valence is placed on the shape of the object than the ranking of the stimulus becomes closer (more correlated) to the ranking produced by the shape control group.

It is predicted that T9 will have the highest correlation to the shape control ranking followed by T7, T4, T2 and T5 (in no particular order), T3, T8 and T6 also in no particular order.

It should be noted that for OH1 and OH2 the predictive ranking orders implies that actor perceives that task in a holistic way. That is to say that despite the fact that a performance indicator is associated with only one dimension, the other dimension does not stop influencing the manner in which the actor chooses to perform the task. This leads to a situation where what gets done is not only what is being measured.

OH3: When the valence is placed on both sensors (T5), it is expected that the colour and shape correlation to the respective colour and shape control will be the same as in the naïve ranking (T2).

It is predicted that the "toy factory" story should have no effect on the ranking of the objects if both features of the objects are being equally emphasized by the performance measurement system"

Operational Assumption (OA) 1: The locations of the objects on the page do not affect the ranking of that object.

For purpose of testing this assumption, two versions of each stimulus set were created. The only difference between the version one and two of the same stimulus was the location of the objects on the page.

4.6 Procedure:

The stimulus set and the instructions set (tasks) were combined into 16 unique experimental booklets. Booklets 1a and 1b were prepared for the control group and involved 3 tasks; the remaining booklets were prepared for the experimental group and involved two tasks. Every booklet had a general instruction page as its cover page (Refer to Appendix A).

The subjects were tested in groups during their regular tutorial times. The experiment was conducted in two major runs, first during Management Science 311 tutorials and second during Management Science 211 tutorial. Each subject was asked to fill out an experimental booklet; the booklets were distributed in such way that neighbours seated beside each other would not work on the same booklet. Before the experiment commenced everyone was instructed, by the experimenter, with regards to the general description of the experiment, their expected behaviour during the experiment and the potential of earning up to 3 bonus marks for participating with good quality answers. Students were also advised that there was no time limit for this experiment. In order to ensure that all groups had received exactly the same pre-experiment instructions a short greeting and instruction script was developed (Refer to Appendix: A).

Subjects were asked to rank order the stimulus as per instructions and fill out a short survey located at the last page (Refer to Appendix: A for the survey). Subjects were also told that once they had completed a task they were allowed to start the following task at their convenience. Each subject was also informed that during the time they were working on a given task they were able to change their answers (ranking) however once they moved on to the next task they were no longer allowed to change any previous answers (rankings). The following tables show all the booklet combinations used in both experiment runs.

Table 1: Booklet Combinations Used for Data Collection in Run 1 (Msci 311)

	Task1	Task2	Task 3	n
Booklet 1a	T2 + ESA_V1	T1a + Shape_A	T1b + Colour_A	9
Booklet 1b	T2 + ESB_V1	T1b + Shape_B	T1a + Colour_B	10
Booklet 2a	T3 + ESA_V1	T6 + ESA_V2		8
Booklet 2b	T3 + ESB_V1	T6 + ESB_V2		6
Booklet 3a	T4 + ESA_V1	T7 + ESA_V2		8
Booklet 3b	T4 + ESB_V1	T7 + ESB_V2		6
Booklet 4a	T5 + ESA_V2	T6 + ESA_V1		8
Booklet 4b	T5 + ESB_V1	T6 + ESB_V2		5
Booklet 5a	T5 + ESA_V2	T7 + ESA_V1		9
Booklet 5b	T5 + ESB_V1	T7 + ESB_V2		5
Booklet 6a	T2 + ESA_V2	T3 + ESA_V1		8
Booklet 6b	T2 + ESB_V1	T3 + ESB_V2		6
Booklet 7a	T2 + ESA_V1	T4 + ESA_V2		7
Booklet 7b	T2 + ESB_V1	T4 + ESB_V2		6
Booklet 8a	T2 + ESA_V1	T5 + ESA_V2		6
Booklet 8b	T2 + ESB_V2	T5 + ESB_V1		6

Table 2: Booklet Combinations Used for Data Collection in Run 2 (Msci 211)

	Task 1	Task2	n
Booklet 6a	T2 + ESA_V1	T3 + ESA_V2	7
Booklet 6b	T2 + ESB_V1	T3 + ESB_V2	11
Booklet 7a	T2 + ESA_V1	T4 + ESA_V2	8
Booklet 7b	T2 + ESB_V1	T4 + ESB_V2	10
Booklet 8a	T2 + ESA_V1	T5 + ESA_V2	6
Booklet 8b	T2 + ESB_V1	T5 + ESB_V2	8
Booklet 9a	T2 + ESA_V1	T8 + ESA_V2	15
Booklet 9b	T2 + ESB_V1	T9 + ESB_V2	15

Legend:
ESA _V1- Experimental Stimulus set A, version One ESA_V2 - Experimental Stimulus set A, Version Two ESB_V1- Experimental stimulus set B, version One ESB_V2- Experimental stimulus B, version Two Shape_A/B - Control groups shape stimulus versions A and B Colour_A/B - Control group colour stimulus versions A and B

Chapter 5: Results

The analysis of the data was performed using Microsoft Excel 2000 and SPSS 12.0.1 for Windows. The description of the results in this chapter is divided into two major parts: control group results and experimental group results respectively.

5.1 Introduction to Analysis

In the first part of the analysis, the control rankings for colour and shape as well as a predictive control ranking using both shape and colour will be devised. This was obtained using the control group data.

In the second part, control rankings will be used to test the responses of the experimental group. The responses of each subject will be tested against the control ranking using Spearman Correlation rank “Rho”. Two kinds of analysis were conducted using this method. First, a within booklet (within-subject) analysis compared how the subjects changed their rankings of particular objects in response to a different instruction set. Second, a between task (across-subjects) analysis was conducted comparing the correlations to the control rankings between tasks, for example: T2 to T3, T2 to T4, etc.

In the fourth part, a simple linear regression analysis between the average of the ranks for each task and control ranks is discussed in order to corroborate the results of the between task (across-subject) analysis

5.2 Control Group Results

First the average of the ranks, for versions one and two (the same objects different location on the page) for both the colour and the shape control stimulus sets were calculated. Those averages for versions One and Two were found to be significantly correlated (Pearson Correlation test), scoring .865 with $p= 0.013$ and .995 with $p< 0.001$ for colour and shape respectively. Both results support operational assumption OA1 and indicate that there is no strong evidence that the location of the objects on the page had an effect on the ranking of those objects. As a result, versions one and two have been combined into a single set for the purpose of further analysis.

The control ranking for colour and shape were obtained identically by: first computing the average rank for each colour and shape; second, computing the mode, the median and the standard deviation of the rank for each colour and shape; third, comparing the results of the first two steps and assigning the rank to each colour and shape respectively. The main criteria of assigning the rank was the average rank computation, in case of average rank being the same or statistically the same the value of median and mode were taken to the consideration.

5.2.1 Colour Control Ranking

First the average rank, mode, median and standard deviation were computed for all six colours across all the observations; the results are presented in the Table 3 below

Table 3: Colour Ranking by Control Group (n=19)

Colour	Average Rank	Median	Mode	Std. Dev	Rank Assigned
2	3.684	4	4	1.455	3
3	3.974	4	5	1.207	5
4	3.868	4	5	1.104	4
5	2.395	2	2	1.231	2
6	5.395	6	6	1.062	6
1(X)	1.421	1	1	0.692	1

(X) Indicates the target colour

The average of the ranks for colours 2, 3 and 4 were found not to be statistically different from one another (Refer to Appendix B). Median and mode in conjunction with the numerical value of the average of the ranks were used to determine the rank for each of the colours.

Using the rank obtained for each individual colour the Colour Control Ranking (CCR) for the experimental stimulus set A and set B were created by assigning the rank of the colour to any object of that colour. The complete Colour control Ranking is shown below in Table X:

Table 4: Colour Control Ranking (CCR) for Stimulus A and B

CCR for Stimulus A

Stim A	Control Colour
1A	1
1C(X)_(1)	1
1C(X)_(2)	1
1C(X)_(3)	1
1D	1
1E	1
2A	3
2C	3
2F	3
3A	5
3D	5
3F	5
4B	4
4C	4
4E	4
5B	2
5D	2
5F	2
6B	6
6C	6
6E	6

CCR for Stimulus B

Stim B	Control Colour
1B	1
1C(X)_(1)	1
1C(X)_(2)	1
1C(X)_(3)	1
1C(X)_(4)	1
1F	1
2B	3
2D	3
2E	3
3B	5
3C	5
3E	5
4A	4
4D	4
4F	4
5A	2
5C	2
5E	2
6A	6
6D	6
6F	6

(X)-(i) – indicates the target object and its instance on the page

5.2.2 Shape Control Ranking

Just as in the previous case the average of the ranks, median, mode and standard deviation were computed across all the observations and than the appropriate rank were assigned. The results are shown in the Table 5 below.

Table 5: Shape Ranking by Control Group (n=19)

Shape	Average Rank	Median	Mode	Std. Dev	Rank Assigned
A	4.579	5	5	0.607	5
B	3.684	3.3	3	0.820	3
C(X)	1.000	1	1	0.000	1
D	2.053	2	2	0.229	2
E	3.632	4	4	0.761	4
F	5.895	6	6	0.315	6

(X)- indicates the target shape

The average of the ranks for shapes B and E were found not to be statistically different (Refer to Appendix: B). The mode and median results were used to decide their assigned rank.

Using the rank obtained for each individual shape the Shape Control Ranking (SCR) for the experimental stimulus set A and set B were created by assigning the rank of the shape to any object of that shape. The complete Shape Control Ranking is shown below in Table 6.

Table 6: Shape Control Ranking (SCR) for Stimulus A and B

SCR for Stimulus A		SCR for Stimulus B	
Stim A	Control Shape	Stim B	Control Shape
1A	5	**1B**	3
1C(X)_(1)	1	**1C(X)_(1)**	1
1C(X)_(2)	1	**1C(X)_(2)**	1
1C(X)_(3)	1	**1C(X)_(3)**	1
1D	2	**1C(X)_(4)**	1
1E	4	**1F**	6
2A	5	**2B**	3
2C	1	**2D**	2
2F	6	**2E**	4
3A	5	**3B**	3
3D	2	**3C**	1
3F	6	**3E**	4
4B	3	**4A**	5
4C	1	**4D**	2
4E	4	**4F**	6
5B	3	**5A**	5
5D	2	**5C**	1
5F	6	**5E**	4
6B	3	**6A**	5
6C	1	**6D**	2
6E	4	**6F**	6

(X)-(i) – indicates the target object and its instance on the page

5.2.3 Predictive Control Ranking

Similarity score was obtained by multiplying each object's average shape rank by its average colour rank. For example to obtain the similarity score of object 1A the average rank for colour 1 (1.421052632) was multiplied by the average rank for shape A (4.578947368), yielding the similarity score for object 1A of 6.50693. This procedure was repeated for all the remaining objects in each stimulus set, than the data were imported into SPSS. The ranks values were generated for each object using the "rank case" function in SPSS. The ranking and the similarity scores are shown in the Table 7 below.

Table 7: Predictive Control Ranking (PCR) for Stimulus A and B

Stim A	Average	RANK Control Both	Stim B	Average	RANK Control Both
1A	6.507	10	**1B**	5.235	7
1C(X)_(1)	1.421	2	**1C(X)_(1)**	1.421	2.5
1C(X)_(2)	1.421	2	**1C(X)_(2)**	1.421	2.5
1C(X)_(3)	1.421	2	**1C(X)_(3)**	1.421	2.5
1D	2.917	4	**1C(X)_(4)**	1.421	2.5
1E	5.161	8	**1F**	8.377	10
2A	16.870	16	**2B**	13.573	15
2C	3.684	5	**2D**	7.562	8
2F	21.717	20	**2E**	13.380	14
3A	18.195	17	**3B**	14.640	17
3D	8.157	11	**3C**	3.974	6
3F	23.424	21	**3E**	14.431	16
4B	14.252	15	**4A**	17.713	18
4C	3.868	6	**4D**	7.940	9
4E	14.048	13	**4F**	22.803	19
5B	8.823	12	**5A**	10.965	12
5D	4.916	7	**5C**	2.395	5
5F	14.116	14	**5E**	8.697	11
6B	19.875	19	**6A**	24.702	20
6C	5.395	9	**6D**	11.073	13
6E	19.591	18	**6F**	31.801	21

(X)-(i) – indicates the target object and its instance on the page

5.3 Experimental Group Results

This section reports on the analysis, and the subsequent results, performed on the data gathered from the experimental group. First the two versions of each experimental stimulus (A and B) were tested to determine whether or not the location of the object on the page had a significant impact on the ranking of that object. The data was analyzed using the Spearman correlation method of comparing the ranked data and using simple linear regression in order to test the operational hypothesis.

5.3.1 Experimental Stimulus Testing:

Two versions of experimental stimulus A and B were created in order to control for any bias or effect that a location of the object on a page could have on the ranking of that object. In order to test the operational assumption OA1 the average of the ranks of each task was computed. Then the data was imported into SPSS and case ranks were assigned for each test for each version, next those ranks for a given task for stimulus A (or B) version one and version two were compared using Spearman Rank Correlation method. A sample procedure is shown below in Table 8. This procedure was repeated for all the tasks for which versions one and two of a given experimental stimulus set appeared with the same instruction set. For the complete results refer to the Appendix H.

Table 8: Testing the Two Versions of Stimulus A for T3

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
10.81	10	6.43	7
2.06	1	1.43	3
2.63	3	1.29	2
2.38	2	1.14	1
6.44	4	3.43	4
7.88	8	5.86	5.5
13.13	16	10.14	11
7.75	7	5.86	5.5
14.69	19	11.00	14.5
15.06	20	15.00	18
11.44	13	11.71	16
14.44	18	16.29	19
10.88	11	10.86	12.5
6.56	5	7.86	10
11.56	14	10.86	12.5
8.94	9	7.43	8
7.38	6	7.57	9
12.38	15	13.86	17
14.25	17	27.00	21
11.06	12	11.00	14.5
15.25	21	16.57	20

Spearman Correlation:	**0.903**
Sig	$p < 0.001$

It has been found that for all tasks, for which the calculation was possible to be performed, versions one and two of the same experimental stimulus set were highly correlated with each other. As a result versions one and two of the same stimulus A or B, were combined to be used in subsequent analysis.

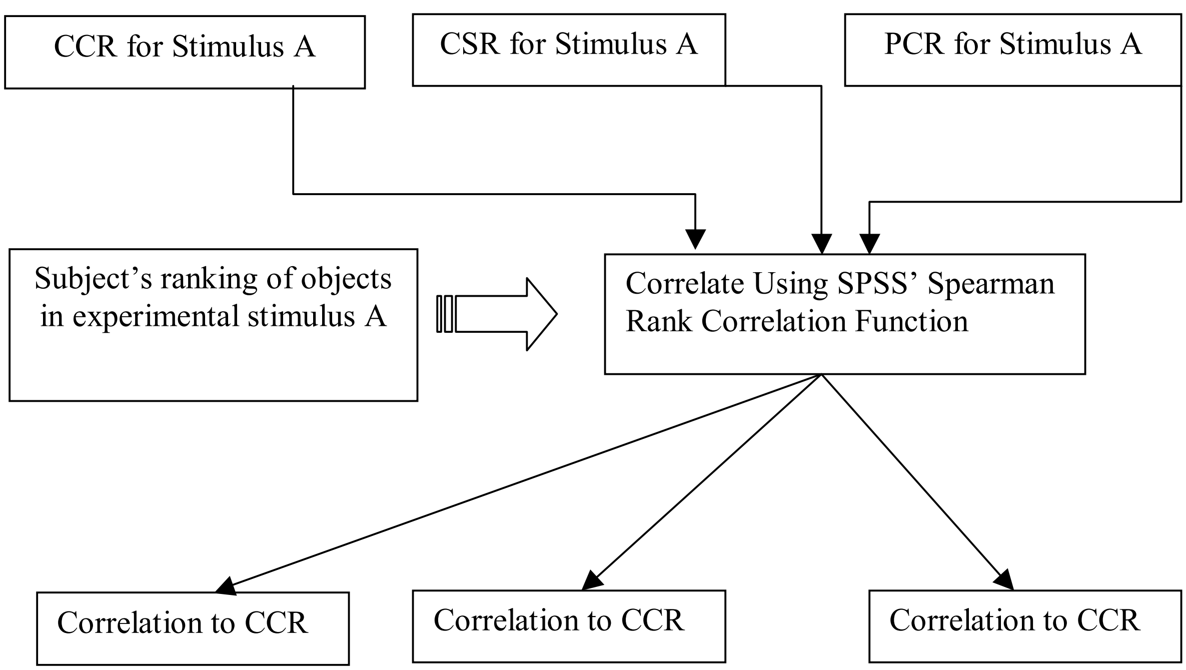
5.3.2 Results of Within Subject (within-booklet) Analysis:

Using SPSS, each subject's ranking for each task recorded in the experimental booklets was compared to respective (stimulus set A/ stimulus set B) Control Colour Ranking (CCR), Control Shape Ranking (CSR) and the Predictive Control Ranking (PCR).

The Spearman Rank Correlation statistic "Rho" was used to compare the subject rankings to the respective control rankings. The closer the statistic "Rho" is to 1 the stronger the correlation (similarity) the closer Rho is to 0 the weaker the correlation.

This approach resulted in obtaining three separate Spearman correlation coefficients "Rho" for each subject for each task, the correlation to the CCR, the correlation to the CSR and the correlation to the PCR. An illustration of this procedure for task, involving experimental stimulus A, is shown below. The same procedure was applied to experimental stimulus B with respect to control rankings for stimulus B.

Figure 11: Obtaining Spearman Correlation Ranks respective to the control rankings



In addition to the correlation value Rho, the significance levels for each calculation were also computed. However due to a very large size of the SPSS' output a summary table for selected values of t-Test is presented here.

Table 9: Selected Spearman Rank Correlation Values and Their Significance

Selected "Rho Value	Corresponding Significance for n = 21
0.336	0.136
0.355	0.114
0.373	0.096
0.379	0.9
0.41	0.65
0.472	0.031
0.512	0.018
0.595	0.004
0.652	0.002
0.715 and up	$p < 0.001$

Since the "Rho" calculations were performed by correlating the subject's ranking of the experimental stimulus A to the control rankings of stimulus A and correlating subject's ranking of experimental stimulus B to the control rankings of stimulus B for a given task, the booklets that contained the same combination of tasks but different stimulus sets (A or B) could be combined into a single data set. For example, the Experimental Booklet 4a contained task T4 and Task T7 followed by experiment stimulus sets A versions One and experiment set A version Two respectively. Booklet 4b also contained tasks T4 and T7 but followed by experiment stimulus sets B versions One and Two. The above-described method of obtaining Rho's allowed for combining the results, obtained from those two booklets, into a single data set from now one referred to as Experimental Booklet 4.

Once the data was arranged by the booklets average correlations (the average of "rhos" from each subject) to CCR, SCR and PCR, for each pair of tasks in each booklet, were computed. The results of this procedure are reported in the Table 10. For the entire data set arranged by the booklets refer to the Appendix C.

Table 10: Within Subject Average Correlations to Control Rankings by Booklet

	Correlation to ->	CCR	SCR	PCR	CCR	SCR	PCR
Booklet 2	Task ->	**T3**			**T6**		
(n)= 14	Avg. Correlation ->	0.623 **	0.650 ***	0.773**	0.818**	0.337***	0.670**
Booklet 3	Task ->	**T4**			**T7**		
(n)= 14	Avg. Correlation ->	0.440	0.840	0.828	0.390	0.852	0.802
Booklet 4	Task ->	**T5**			**T6**		
(n)= 13	Avg. Correlation ->	0.584***	0.768***	0.876***	0.814***	0.289***	0.634***
Booklet 5	Task ->	**T5**			**T7**		
(n)= 14	Avg. Correlation ->	0.506**	0.833	0.851*	0.355**	0.914	0.824*
Booklet 6	Task ->	**T2**			**T3**		
(n)= 32	Avg. Correlation ->	0.528**	0.713*	0.775	0.661**	0.624*	0.742
Booklet 7	Task ->	**T2**			**T4**		
(n)= 30	Avg. Correlation ->	0.526**	0.777	0.810	0.454**	0.844	0.812
Booklet 8	Task ->	**T2**			**T5**		
(n)= 26	Avg. Correlation ->	0.544	0.727	0.784	0.533	0.767	0.796
Booklet 9	Task ->	**T2**			**T8**		
(n)= 15	Avg. Correlation ->	0.438***	0.721***	0.820***	0.726***	0.106***	0.432***
Booklet 10	Task ->	**T2**			**T9**		
(n)= 15	Avg. Correlation ->	0.519	0.864	0.818	0.450	0.907	0.816

Legend:

T2 - Naïve Ranking
T3- Valence on Colour Sensor
T4- Valence on Shape Sensor
T5- Valance on Both Sensors
T6 - Point of View of Colour Sensor
T7- Point of View of Shape Sensor
T8 - Naïve Ranking by Colour
T9 - Naïve Ranking by Shape

Significance of difference between the control rankings of the first and second task in a given booklet
*** p < 0.001
** p < 0.05
* p < 0.1

In order to test whether or not the "within subject" differences in correlation to the control ranking were significantly different form one another a series of T-Tests were performed. The results are presented in the Appendix: D. The more conservative T-Test for samples of equal variance was used for the cases where the variance was of the same order of magnitude (for example 0.034 vs. 0.016). In few cases the t-Test for samples of unequal variance had to be used when samples variances differed by at least one order of magnitude (for example 0.005 vs. 0.044).

It can be observed that, for example, results for the Booklet 2 are consistent with the operational hypothesis. The correlation to the colour control has increased and the correlation to the shape control has decreased, as the valence on the colour dimension has increased. The correlation to CCR increased from 0.623 in T3 to 0.816 in T6 (this difference has statistical significance of $p = 0.006$). The results of the Booklet 8 show that the correlations to the colour and shape controls for tests T2 and T5 are not statistically different with significance for CCR of $p = 0.406$ and significance for SCR of $p = 0.274$. This result is also consistent with the operational hypothesis.

The results for the Booklet 3 show that in the case of test T4 and T7 the correlation to the shape colour control has decreased as predicted, but the correlation to the shape control did not increase significantly ($p = 0.434$). All the other booklets were analyzed in the same manner with regards to the operational hypothesis. For the results of T-Tests for the data in Table 10 refer to Appendix: D

5.3.3 Results of between task (across-subject) analyses:

For this analysis an average correlation ranks were computed for each of the tasks T2 through to T9 using the Spearman correlation parameters obtained in the previous analysis. Those average correlations were obtained by first combining the individual correlations to CCR, SCR and PCR by task (for T2 then T3 and so on) and than computing the average correlations by task to each control ranking. The summary table of the results is shown below.

Table 11: Average Spearman Correlation "Rho" for the Between Task Analysis

Task	Correlation to CCR	Correlation to SCR	Correlation to PCR
T2	0.5320	0.7391	0.7895
T3	0.6498	0.6316	0.7518
T4	0.4492	0.8427	0.8168
T5	0.5382	0.7846	0.8297
T6	0.8159	0.3140	0.6523
T7	0.3719	0.8844	0.8133
T8	0.7196	0.1014	0.4351
T9	0.4563	0.8978	0.8146

The T-Test statistic was, once again, used to determine if the average Spearman correlations for CCR and SCR (Rho) were statistically different from one another between the tasks. The same logic as before, for the use of T-Test for samples with equal and unequal variance was applied. For the complete set of T-Tests please refer to the Appendix: E

It can be observed that, as predicted, the correlation to colour control and shape control for test T2 and T5 are not statistically different $p = 0.419$ and 0.120 respectively. It can also be observed that test T3 has a higher correlation to colour control than the base tests T2 and T5 ($p < 0.001$ and $p=0.002$ respectively), but lower than test T6 or test T8, which is consistent with the hypotheses. Furthermore we can observe that test T9 has the highest correlation to shape control ranking (SCR) while T8 has the lowest. Similarly T4 has a higher correlation to the shape control ranking than T3 ($p < 0.001$), and T3 has a higher correlation to the colour control ranking than T4 ($p <0.001$); a result that is also consistent with the experimental predictions.

All other comparisons between tests were conducted in the same manner with regards to control rankings and hypotheses. The results of T-Test can be found in Appendix: E

5.3.4 Summary of within subject and between task findings

The results in most cases indicate that the experimental perdition were correct. It can be observed that as the instructions change and therefore the valence placed on a given dimension changes, the subjects respond in the direction predicted by the operational hypothesis. As a valence in the task is placed on the colour, a higher correlation to the colour control can be observed. Moreover, as the valence is placed on the shape, a higher correlation to the shape control can be observed.

It is also important to notice, however, that in both modes of analysis (within booklet and within task) the results for tasks T7, T4 and T9 in terms of correlation to shape control ranking, were found not be statistically significantly different. This result indicates that there could have been a prior bias towards shape. This theory seems to be corroborated by the results of tasks T2 and T5. Both tasks place equal valance on each dimension thus it was expected that the correlation to colour control and shape control should be very similar. The results however show that the correlation to shape control is higher than to colour control suggesting that there exists prior bias toward shape.

It can also, be concluded that the pattern in ranking changes, which comes as a result of various performance measurement conditions, can be observed not only for the within subject analysis but also for the between tests (across the subjects) analysis.

5.3.5 Results of Average Rank Regression Analysis

Introduction:

In this section the average ranks for each of the objects in each task and in each of the experimental stimulus sets A and B are investigated. For this analysis the experimental stimulus set A and set B had to be considered separately. This is because stimulus sets A and B contain different objects and the purpose of this analysis was to investigate average rank of a particular object in a given task.

For example, task T4-A (task T4 followed by one of the versions of experimental stimulus A) and Task T4-B could not have been combined to obtain average of the ranks for a given object because sets A and B contain different objects. Conversely, versions One and Two for each of the experimental sets A/B can be combined as they contain the same sets of objects only arranged differently on a page.

Results of Computations:

First, the average of the ranks was calculated for a given object in a given task within a given experimental set A and B. Then the data were imported into SPSS and the case rank function used to assign ranks to the objects based on the average of the ranks. Please refer to the Appendix: F and Appendix: G for the results.

The second step in this analysis was to perform a simple linear regression analysis between each of the tests in each experiment stimulus set A and B and the CCR and SCR for corresponding test sets. The ranked order of each of the task sets was considered to be the dependent variable while CCR and SCR rankings where assumed to be the independent variables.

The purpose of this regression analysis was to determine if there is a relationship, and if so what kind of relationship, between ranks assigned to an object by the control group and the ranks assigned to that same object by the experimental group. In addition to the above, this analysis was conducted to corroborate the results of the "within subject analysis".

The regression results are summarized in the summary tables below. For the detailed results and full-scale regression line graphs please refer to Appendix: F and Appendix: D

Table 12: Summary Regression Table for Stimulus A

	Colour Control			Regression	Shape Control			Regression
Test	Slope	Intercept	R Square	Significance	Slope	Intercept	R Square	Significance
T2	1.90	5.04	0.32	0.008	2.90	1.88	0.75	P < 0.001
T3	2.48	3.21	0.55	P < 0.001	2.39	3.48	0.51	P < 0.001
T4	1.58	6.05	0.22	0.032	3.08	1.33	0.84	P < 0.001
T5	1.87	5.13	0.31	0.009	2.87	1.99	0.73	P < 0.001
T6	3.00	1.56	0.80	P < 0.001	1.21	7.20	0.13	0.108
T7	1.31	6.88	0.15	0.079	3.16	1.05	0.89	P < 0.001
T8	3.22	0.87	0.93	P < 0.001	0.75	2.73	0.04	0.363

Legend:

T2 - Naïve Ranking
T3 - Valence on Colour Sensor
T4 - Valence on Shape Sensor
T5 - Valence on Both Sensors
T6 - Point of View of Colour Sensor
T7- Point of View of Shape Sensor
T8 - Naïve Ranking by Colour
T9 - Naïve Ranking by Shape

Table 13: Summary Regression Table for Stimulus B

Test	Colour Control Slope	Intercept	R Square	Regression Significance	Shape Control Slope	Intercept	R Square	Regression Significance
T2	2.14	4.26	0.41	0.002	3.00	1.56	0.80	P < 0.001
T3	2.33	3.67	0.49	P < 0.001	2.83	2.11	0.71	P < 0.001
T4	1.82	5.27	0.30	0.011	3.21	0.92	0.92	P < 0.001
T5	2.06	4.54	0.38	0.003	3.04	1.44	0.82	P < 0.001
T6	2.94	1.76	0.77	P < 0.001	1.06	7.68	0.10	0.163
T7	1.66	5.78	0.25	0.022	3.21	0.92	0.92	P < 0.001
T9	1.65	5.82	0.24	0.024	3.22	0.89	0.92	P < 0.001

<u>Legend:</u>

T2 - Naïve Ranking
T3- Valence on Colour Sensor
T4- Valence on Shape Sensor
T5- Valance on Both Sensors
T6 - Point of View of Colour Sensor
T7- Point of View of Shape Sensor
T8 - Naïve Ranking by Colour
T9 - Naïve Ranking by Shape

As a final step of this analysis, the regression lines for each of the tasks, for both experimental stimulus sets A and B were sorted by their slope angle in the ascending order. The results of that manipulation are shown in the table below.

Table 14: Tasks According to the Slope Angle of Their Regression Lines in the Ascending Order

Order	Experimental Stimulus A		Experimental Stimulus B	
	Colour Control	Shape Control	Colour Control	Shape Control
1	T7	T8	T9	T6
2	T4	T6	T7	T3
3	T5	T3	T4	T2
4	T2	T5	T5	T5
5	T3	T2	T2	T4 and T7
6	T6	T4	T3	
7	T8	T7	T6	T9

The steeper the angle of the regression of the line the higher the predictive value of the CCR / SCR in explaining the experimental stimulus sets ranking. As the residuals are closer together and align in a clearer pattern, the steeper the regression line becomes. Conversely, the greater the dispersion of the residual points and the more chaotic the pattern the flatter the regression line indicating that a given independent variable has much weaker relationship with the dependent variable set.

5.3.6 Summary of Findings for the Regression Analysis

The results for this analysis are consistent with the experimental predictions and also corroborate the findings of the previous analysis method. As the valence imposed by the task shift from shape to colour, it can be observed that the slope of the regression line changes accordingly. It increases with regards to the shape control as the valence in the task is placed on the shape and decreases for the colour control for the same task and vice versa. It should be noted that slope lines angles for tasks T4, T7 and T9 are very close or even the same at the times. This indicates, as previously suggested, that a bias toward shape exists.

5.4 Results of the survey study

It should be noted that no rigorous analysis of the survey was conducted; only a brief overview of the comments was performed.

A number of subjects indicated, in the surveys, that they thought the shape was more important because they believe that the toy block should be stackable or because they believed that for the purpose of quality control it is easier to change the colour of the block than its shape. Comments such as that suggest that there could have been a prior bias toward shape on top of any bias introduced by the "toy factory" background story. Furthermore those types of comments indicate that the instruction set had enough ambiguity to avoid the solution to be too obvious and the instruction set to be too suggestive. On the other hand this also indicates that there is a considerable amount of noise in the data, making the detection of the signal harder. It is also important to notice that nothing was done about the noise in the data and the outliers.

Chapter 6: Discussion of the Results

6.1 Summary of the Results

6.1.1 Within subject and between task analysis

The analysis can be divided into two distinct groups. The first group called "within subject" analysis, focused on comparing individual subject's responses to the Colour Control Ranking (CCR), Shape Control Ranking (SCR) and Predictive Control Ranking (PCR). The "within subject" responses have been analyzed by the experiment booklet and than by the task. In both cases, the results had shown that as the focus of the task shifted from the colour dimension to the shape dimension, the correlation to the respective colour control and shape control rankings have changed accordingly in the predicted direction. In other words, as the valence, induced by the instruction set, increased on a given feature (colour or shape), so did the correlation of the experiment set rankings to the respective colour or shape control ranking.

It is important to notice that the number of observations used for T-Test, were low at times ($n < 15$). In such cases just a few outliers could throw off the results of the T-Test in the direction opposite to the hypothesis.

It has been observed that in both cases, the naïve ranking task T2 and the task T5 (equal valence placed on both features), subjects ranked the objects in a very similar way, placing greater weight on the shape of the object rather than on its colour. T-Test indicated that the average correlation values for those two tasks were not statistically different.

The fact that the shape influenced subjects' ranking of the objects more then the colour did, was also indicated by the other results of the within subject analysis. Results of the booklet and task analysis revealed that the average correlations to, the control sets, for tasks T7 (Point of view of the shape sensor) and T9 (naïve ranking by shape) were not statistically different form correlation of task T4 (Valence on shape) especially for the Shape

Control Ranking correlations. This is in contrast to tasks T8 (naïve ranking by colour), T6 (Point of view of colour sensor) and T3 (valence on colour) where all the average correlations to the control rankings were statistically different, and much more pronounced.

6.1.2 Average Ranking Regression Analysis

The results of the regression were consistent with the previously reported results. This analysis had focused on regression analysis of the relationship between the average ranks for each test (for each stimulus) and the predictive rankings CCR and SCR. The results showed that for Colour Control Ranking the angle of the regression line was the steepest (the strongest relationship) for the test that placed the highest emphasis on colour dimension and weakest for the tasks that placed the higher valence on shape. It has also been found that the strength of the relationship (as indicated by the steepest angle) placed the test in the order predicted by the operational hypothesis.

In addition, it can be observed that for the tasks that induced valence on shape (T9, T7, T4) the differences between the angles of the slopes of the regression lines are much smaller than the differences between the slope angles of task, which induced valance on colour (T8, T6, T3). This finding is consistent with the results of the "within subject" analysis, which found that average correlations for T9, T7 and T4 were not statistically different.

Furthermore, the results indicate that tasks T5 and T2 have very similar slopes and intercepts, which means that in both cases the objects were rank very similarly despite the difference in the instruction sets (T2 – no story, T5- toy factory story). Once again these results are consistent with the findings of the first analysis.

6.2 Support of Hypotheses

The overall results seem to support the hypothesis. Operational hypothesis OH1 is supported by the fact that tasks, which placed the strongest valence on colour T3 and T6 had the strongest correlation to the colour control as compared with other the tasks especially the tasks T2 and T5, which placed equal valence on both features. Those findings were also confirmed by the regression analysis. The tasks T3 and T6, which focused on the colour, had the steepest angles of their regression lines, and smallest residual errors indicating a strong positive relationship of those tasks with the Colour Control Ranking. Furthermore, task T6 had a stronger correlation to CCR and steeper angle of the regression line than task T3

The operational hypothesis OH2 is supported by the fact that the tasks T4 and T7 had the highest correlation to the Shape Control Ranking and both the regression lines had the steepest angles for SCR, once again indicating that the ranking of the objects for those tasks was indeed done more on the basis of shape than colour.

It can be observed however that the T7's and T9's correlation to the Shape Control and the angle of the regression lines were not much steeper than those for task T4 and in some cases the results showed no statistical difference between the results for those tasks.

The results seem to indicate that subjects placed higher weight on the shape then on the colour when ranking the objects. It can be observed that, for any task, the correlation to the predictive ranking PCR is much more sensitive to the changes in changes in the shape control ranking (SCR) than it is to any changes in Colour Control Ranking. Furthermore, the results show that when the same valence is placed on both dimension, like in task T5, or when there are not explicit instructions, as in task T2, than in both cases the average correlation to shape control is larger then the one for colour. Also, regression results suggest that more variance in the ranking is explained by the shape control ranking than by the colour. This even further strengthens the argument that there was a pre-existing bias toward shape. This bias is making the actual signal more difficult to detect than in the case of the colour. Nevertheless the results still support the general hypothesis.

The results also support operational hypothesis OH3. Despite the bias that was introduced by the background story, the rankings for tasks being present T2 and T5 are very similar and often statistically not different. Once again both modes of analysis consistently show that the correlations to Colour Control Ranking and Shape Control Ranking for both tasks are very similar as indicate the similar slopes of the regression lines, and similar average Spearman rank correlations to the control rankings. It is also important to mention again that in the case of this experiment subjects paid more attention to the shape of the objects rather than the colour. This fact does not go against the general theory proposed by this thesis; it is not consistent however with the experimental prediction of both dimensions being regarded as equal. This could be because of the way the stimulus was prepared. It has been mention by the subject in the debriefing and surveys that the colours were very similar and with 21 objects on a single page it was simply easier to focus on the shape rather than the colour.

The results of both analysis for tasks T9 and T8 are also consistent with the proposed theory. In both cases subjects could clearly perceive the two separate dimensions of shape and colour and were able to rank the experimental stimulus accordingly. A highest correlation of T8 and T9 to Colour Control Ranking and Shape Control Ranking respectively as well as high correlation of T8 to T6 and T9 to T7 indicates that the outcome results produced by the "toy factory" story are not unique to the background story itself. Additionally the results of T8 and T9 support the assumption that the background story itself did not introduce a very large bias.

The operational assumption OA1 is supported by the results of the Spearman rank correlation tasting performed on all the tasks that had both version One and Two of given experimental stimulus present. The results support the experimental prediction that the location of an individual object on a page does not affect the ranking of that object by the subjects.

It is important to note that the magnitude of correlation coefficient that has been used as a support for OA1 (the rankings of versions one and two of the same stimulus set being the same) is comparable in some cases to the magnitude of correlation coefficient used in "within subject" analysis to support that the average correlations are different. However the T-Tests were used in all cases to investigate the significance of all numbers and in case of the "within subject" analysis the results are systematic across all performed tests. Furthermore if the location of the object on the page did play a major role then it would have only introduced noise to the data set making the signal even harder to detect.

6.3 Limitations of the Study

One of the most significant limitations of this study was the fact that there was no scientific of measure of differences between the shapes and the colures that could have been used to create the stimulus sets. An effort was made to keep the shapes and the colours similar, for the shape the angles were changed by 5% only, and for the colour the RGB dimensions were manipulated in such way so the colours are roughly the same distance away from the target colour (in the RGB dimension). However, because no reliable measure of "rectangle-ness" for shapes or "redness" for colours were present it was impossible to predict the control or experimental rankings ahead of time. This made the task of evaluating the subject's responses more difficult. And some undetectable bias related to the choice of the objects (colour/shape) could have been operating as a result.

Another challenging task in this study was to write the instructions so that the task is described properly. The instruction set had to convey that there is a discrepancy between the stated goal and the performance measure mechanism. However it had to be done in a way not to make the task too obvious and tip the hand or to settle and make it too difficult to detect that discrepancy. Furthermore the instruction and the background story had to be written so that the actual experimental task, the ranking, the subjects were asked to perform makes sense and is reasonable. Additional challenges came from the fact that for the purpose of the hypothesis testing the instruction set had to communicate in a reasonable fashion that the quality of the object should be equated with similarity with the similarity to the object called "perfect sample". The overall goal was to make sure that here is a just enough ambiguity to the outcome so that the subjects don't miss the phenomenon studied or that the subjects are not lead into performing exactly what the study is testing.

This study was designed as an abstraction of reality with two very clear dimensions the colour dimension and the shape dimension. In real life, environments are much more complex with many dimensions to each situation.

The actions of the individuals in real life are not only motivated by the task at hand or by the performance evaluation alone. This study does not take into account all the external elements issues such as emotions, loyalties etc. that influence people and their decisions.

Furthermore, in real life systems the actors often are part of a feedback loop where they can learn and adjust their responses accordingly. The experiment on the other hand was designed as a one-off task, more of a snapshot in time rather than a continuous process observation.

In the experimental study the 3 % bonus mark was the really the only source of valence. Conversely, in real life the stakes are much higher and hence the valence forces are usually much stronger. Additionally, there are also more sources of valence than in this experimental study, making the real life situations much more complex and hence making the available choices less clear-cut.

One other limitation of this study is that the results of the experiment could be explained in the alternative manner. One might say they the results are consistent with the instructions given to the subjects and all the subjects did were following these instructions.

While such an explanation is plausible, it is not so. The instructions were design to mimic the real world where most of the jobs come with explicit description of the tasks to be performed and with the explicit mention of potential rewards and punishments associated with the performance. Yet the behaviours described by this thesis still occur and it is not always the case that what gets measured gets done. Furthermore the subjects in this experiment actually did not follow the instructions to the letter as it was shown in the analysis section. As a matter of fact the non measured features still influenced subject's choices despite instructions telling them otherwise.

Chapter 7: Conclusions, Contributions and Future Directions

7.1 Conclusions

The actors often see problems as multidimensional, even if the system due to its design sees the same problem as less complex. The actor and the performance measure mechanism often might have a different understanding of the priorities or sequences of the sub-tasks. Actors are capable of perceiving and exploiting this perception gap. As a result when actors negotiate their way through the system, and especially when they choose a path to the goal their selection is heavily influenced by their model of the performance measure mechanism and the valences this mechanism induces. The similarity between the alternatives becomes a function of what the control system/ performance mechanism sees as similar not what could be categorized as similar or dissimilar in the absence of the control mechanism or in the presence of entirely different mechanism.

This may have various impacts on the system itself. As the performance measurement system may influence the behaviour of the actor in both a positive and negative fashion. On one hand this enables people to achieve good job evaluations despite the job not been done properly i.e. does not lead to the achievement of system's goal. On the other hand, however this can allow for problem solving where the system's performance measure or control mechanism by its very design becomes an obstacle in reaching the goal.

7.2 Outline of Contributions:

This study offers a theoretical explanation of performance measurement influencing actor's behaviour. The theoretical framework and the empirical findings help to answer, at least in part, why people behave as they do and how the control system and in particular performance measurement influences their behaviour.

As well this study offers a more precise explanation of the cognitive processes involved in the actor's perception of the task and actor's decision-making process. This research offers a conceptualization of the actor's point of view in terms of the actor's perception of the complexity not being captured by the system, perception of alternative paths stemming from the additional variety, and finally perception of the system of valence forces stemming from the existence of alternatives and influencing the actor's decision of a path. This study expands on Lawler's (1976) use of valence in decision-making bringing the meaning and function of the valence closer to the concept of valence introduced by Lewin (1936)

In addition, this thesis introduces and implements the idea of similarity judgment in the decision process. It is proposed that the similarity judgment between the system-prescribed path and alternative paths is correlated to actor's perception of valences associated with those paths and that the performance measurement provides the very basis upon which the actor judges that similarity. Furthermore, the valence an actor perceives toward a given path is also influenced by above mentioned similarity judgment. This thesis also proposes to look at the performance measurement activity itself as a task of similarity judgment between the actual actions performed by the actor and the system's model of the desired performance.

This research also offers an experimental explanation and illustration of the impact of the performance measurement on an actor's behaviour. This is in a contrast to most of the existing literature, which offers general explanations based on case studies. Those case studies, though useful in highlighting the problem in a particular instance, are very difficult to judge because of numerous other external factors affecting the situations.

A better understanding of the impact on the performance measure on the actor's behaviour, presented in this thesis, also helps to better understand the mechanisms behind the phenomenon of workarounds. Both problems (workarounds and influence of performance measurement) have been identified and acknowledge in the past however, the existing literature did not focus on the mechanisms behind those problems, offering only descriptions of the overall effects of those phenomena.

The findings of this study can be applied into more general cases of system design problems, then previous models present in the literature. For example, this study could shed some light and offer additional explanations and insight for the general theory behind Suchman's study of situated actions. The complexities of the situation and the disturbances generated by the environment often cannot be captured in a plan. Just like in the case of performance measurement systems planning activities are usually done to prepare for the most common and likely disturbances as seen by the person designing the plan and not necessarily as a person (an actor) who is going to use it. A case that is very similar to the relationship between the performance measurement system and the responses of an actor in that system.

7.3 Future Research:

One potential avenue for a future research would be to repeat this experiment, but beforehand conducting a "Just Noticeable Difference" study on the stimulus sets. This would allow for greater control for the experiment and possibly help prevent any bias associated with the choice of shape or colour.

Another potential direction for this kind of research would be to devise a very similar experiment but choose some other object (stimulus set) with some other two or more well-defined features (dimensions), which could be controlled for. The more complex stimulus would allow for additional manipulation and could reveal the extent to which the valence associated with the task itself influences the way in which the actor performs. Furthermore a task could be devised which would be comprised of sever subtasks allowing for more accurate testing of the sequence the actor selects based on provided conditions. Those kind of follow-up studies could be used to validate previous findings and further prove that the studied phenomenon has generalizeable properties.

The sources of valence as well as the stakes have been proven to be a potential limitation of this study. A more detailed study could be devised where smaller groups of subjects could be exposed to various scenarios (similar to the one described in this study), and their behaviour could be tracked across those different tasks. The subjects in this proposed study could be rewarded monetarily after each event. This could increase the stakes (strength of valence) as well as mimic a feedback loop situation by allowing the subject to learn.

In addition, this research raises some interesting questions with regards to human perception. Some future study could focus on the role the valence and perception of similarity has in the perception of other purposeful social actions. Lastly, it would also be of interest to go back to some previous studies in this general area and investigate if the methods proposed in this thesis will suggest something new in terms of the decisions and analysis of those decisions made in those previous studies.

References:

Agarwal, Sanjeev (1999). "Impact of Job Formalization and Administrative Controls on Attitudes of Industrial Salespersons", *Industrial Marketing management*, Vol. 28, pp. 359-368

Ansari, Shahid L. (1977). "An Integrated Approach to Control System Design", *Accounting, Organizations and Society*, Vol. 2, No. 2, pp. 101-112

Argyris Chris (1971). "Management Information Systems: The Challenges to Rationality and Emotionality", *Management Science*, Vol.17, No. 6, Application Series, pp B275-B292

Ashby, W.R. (1956). *An Introduction to Cybernetics*, John Wiley & Sons, New York, U.S.A.

Baker G., Gibbons R., Murphy K.J. (1994)." Subjective Performance Measures in Optimal Incentive Contracts", *The Quarterly Journal of Economics*, Vol. 109, No. 4, pp. 1125-1156

Barsalou, L.W. (1983). "Ad hoc categorization", *Memory and Cognition*, Vol. 11, pp. 211-227

Beer, Stafford (1973). *Designing Freedom*, CBC Publications, Toronto, Canada

Blau, Peter M. (1955). *The Dynamics of Bureaucracy*, The University of Chicago Press, Chicago, U.S.A.

Coate, L.E. (1996). "Beyond re-engineering: Changing the organizational paradigm", *Organizational Paradigm Shift*, National Association of College and University Business Officers, Washington, D.C., U.S.A.

Courty Pascal, Marschke Gerald (2003). "Dynamics of Performance Measurement Systems", *Oxford Review of Economic Policy*, Vol. 19, No. 2, pp. 268-284

Duimering P. Robert, Safayeni Frank (1998). "The Role of Language and Formal Structure in the Construction and Maintenance of Organizational Images" Int. Studies of Mgt. & Org., Vol. 28, No. 3, pp. 57-85

Feller, Irwin (2002). "Performance Measurement Redux", *American Journal of Evaluation*, Vol. 23, No. 4, pp. 435-452

Ghalayini, Alaa M., Noble, James S. (1996). "The changing basis of performance measurement", *International Journal of Operations & Productions Management*, Vol. 16, No. 8, pp. 63-80

Gibbons Robert (1998). “Incentives in Organizations”, *Journal of Economic Perspectives*, Vol.12, No. 4, pp. 115-132

Goldstone Robert L. (1994). “The role of similarity in categorization: providing a groundwork”, *Cognition*, Vol. 52, pp. 125-157

Hamilton, Tiffany G. (1998). *The interactions of information technology and organizations: An exploratory study of the use of workaround*s, Unpublished Master’s Thesis. University of Waterloo, Ontario, Canada

Hayes Naill (2000). “Work-arounds and Boundary Crossing in a High Tech Optronics Company: The Role of Co-operative Workflow Technologies”, *Computer Supported Cooperative Work*, Vol. 9, pp. 435-455

Herdiman, P.T., Dufresne, R., Mestre, J.P. (1989). “The relationship between problem categorization and problem solving among experts and novices”, Memory and Cognition, Vol. 17, pp. 627-638

Irving, R.H., Higgins, C.A., Safayeni, F.H. (1986). ”Computerized Performance Monitoring Systems: Use and Abuse”. Communications of the ACM, Vol. 29, No. 8, pp.794-801

Jensen Pieter (2004). “Performance Measurement in Governmental Organizations: A Contingent Approach to Measurement and Management Control”, *Managerial Finance*, Vol. 30, No. 8, pp. 54-68

Kaplan, R., Norton, D. (1992). “The balanced scorecard as a strategic management system”, *Harvard Business Review*, January-February, pp. 61-66

Kerr, Steven (1995).”On the folly of rewarding A, while hoping for B”, The Academy of Management Executives, Vol. 9, No. 1, pp. 7-14

Kingston J. (2001). “High Performance Knowledge Bases: four approaches to knowledge and acquisition, representation and reasoning for workaround planning” *Expert Systems with Applications*, Vol. 21, pp. 181-190

Lawler III, Edward E. (1976). “Control Systems in Organizations”, *Handbook of Industrial and Organizational Psychology*, Dunnette, M.D. Editor, Rand McNally College Publishing Company, Chicago, IL, U.S.A., pp. 1247-1291

Lewin, Kurt (1936). *Principles of Topological Psychology*, McGraw Hill Book Company, New York, U.S.A.

March, J.G., Simon, H.A. (1958). *Organizations,* John Wiley & Sons, Inc. New York, U.S.A.

Marrow, Alfred Jay (1969). *The practical theorist: the life and work of Kurt Lewin,* Basic Books, New York, U.S.A.

Merton, Robert K. (1936). "The Unanticipated Consequences of Purposive Social Action", *American Sociological Review*, Vol. 1, No 6, pp. 894-904

Merton, Robert K. (1952). "Bureaucratic Structure and Personality", Reader in Bureaucracy, Free Press, Glencoe, Illinois, U.S.A.

Neely, Andy, Richards Huw, Mills John, Platts Ken, Bourne Mike (1997)." Designing performance measures: a structured approach" *International Journal of Operations & Production Management*, Vol. 17, No. 11, pp.1131-1152

Norreklit, Hanne (2000). "The balance on the balanced scorecard-a critical analysis of some of its assumptions", *Management Accounting Research*, Vol. 11, pp. 65-88

O'Mara, Charles E., Hyland Paul W., Chapman Ross L. (1998). "Performance measurement and strategic change", *Managing Service Quality*, Vol. 8, No. 3, pp. 178-182

Ouchi, William G. (1977). "The Relationship Between Organizational Structure and Organizational Control", *Administrative Science Quarterly*, Vol. 22, No. 1, pp. 95-113

Parish Gary, Sollfrey William (1980). "A preliminary analysis of the effect of work-arounds on space system performance and procurement requirements – a proposal", Rand Corporation, Prepared for the U.S.A.F.

Perrin, Burt (1998). "Effective Use and Misuse of Performance Measurement" *American Journal of Evaluation*, Vol. 19, No. 3, pp. 367-379

Petrides Lisa A., McCelland Sara I., Nodine Thad R. (2004). "Costs and benefits of the workaround: inventive solution or costly alternative", *The international Journal of Education Management*, Vol. 18, No. 2, pp. 100-108

Prendergast Canice (1999). "The provision of incentives in firms", *Journal of Economic Literature*, Vol. 37, No.1, pp. 7-63

Ridgway, V.F. (1956). "Dysfunctional Consequences of Performance Measurement", *Administrative Science Quarterly*, Vol. 1, No. 2, pp.240-247

Suchman, Lucy A. (1987). *Plans and situated actions, Cambridge*, University Press, New York, U.S.A.

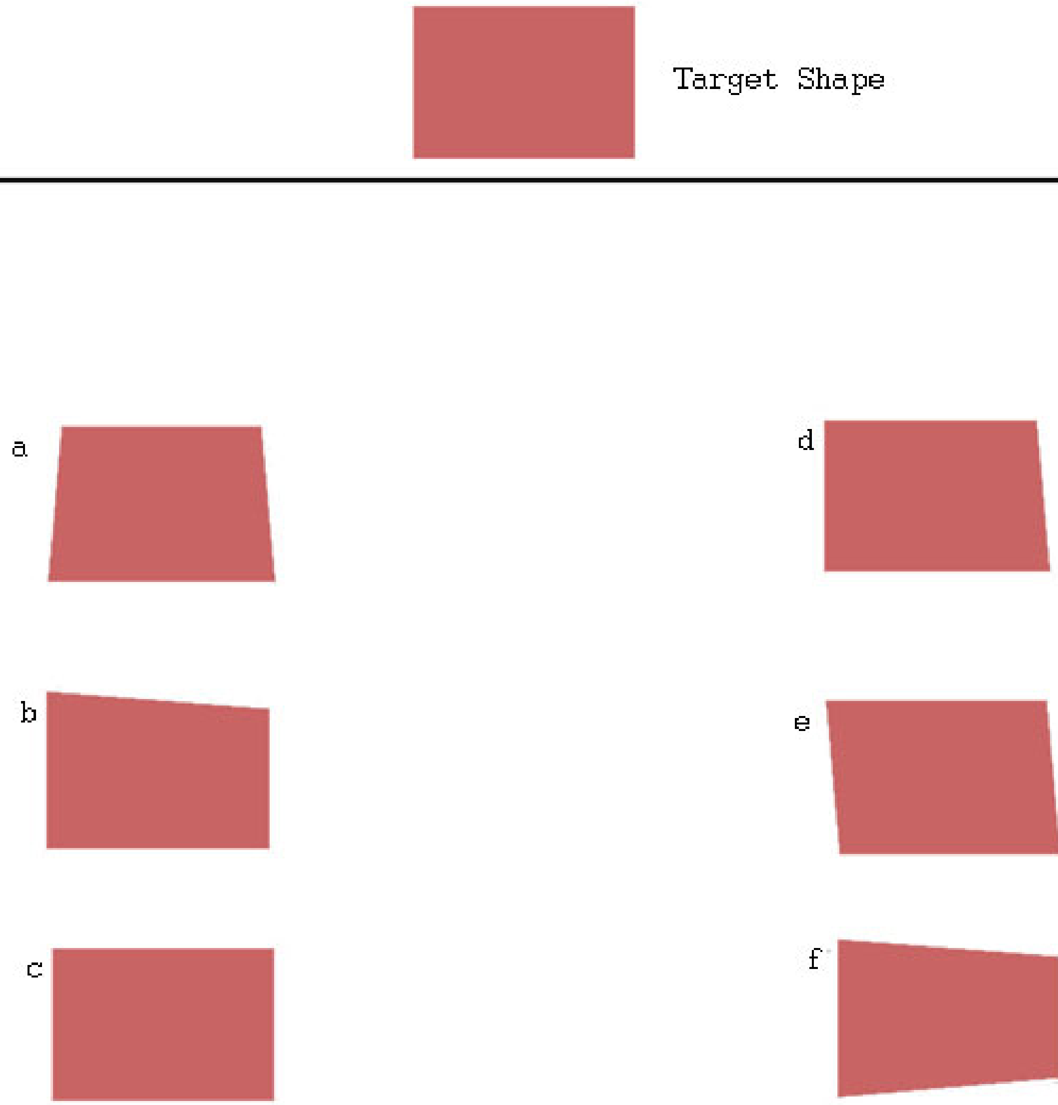
Suzuki, H., Ohnishi, H., Shingemasu, K, (1992). " Goal directed processing in similarity judgment", *Proceedings of Fourteenth Annual Conference of the Cognitive Science Society*, pp 343-348, Erlbaum, Hillsdale, NJ, U.S.A.

Tversky Amos (1977). “Features of Similarity”, *Psychological Review*, Vol. 84, No. 4, pp. 327-352

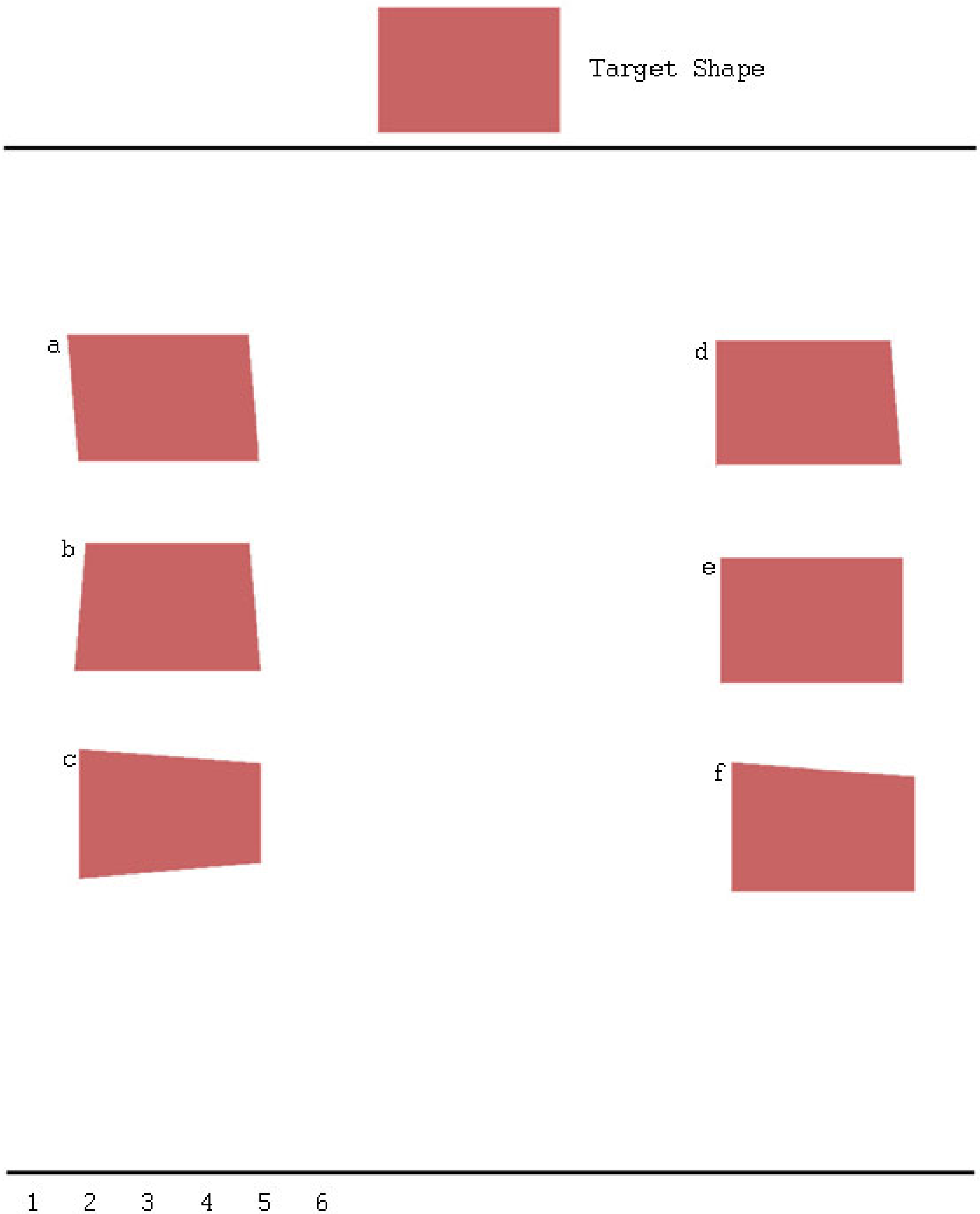
Weiner, Norbert (1948). *Cybernetics or control and communication in the Animal and the Machine*, The Technology Press, John Wiley & Sons, Inc. New York, U.S.A.

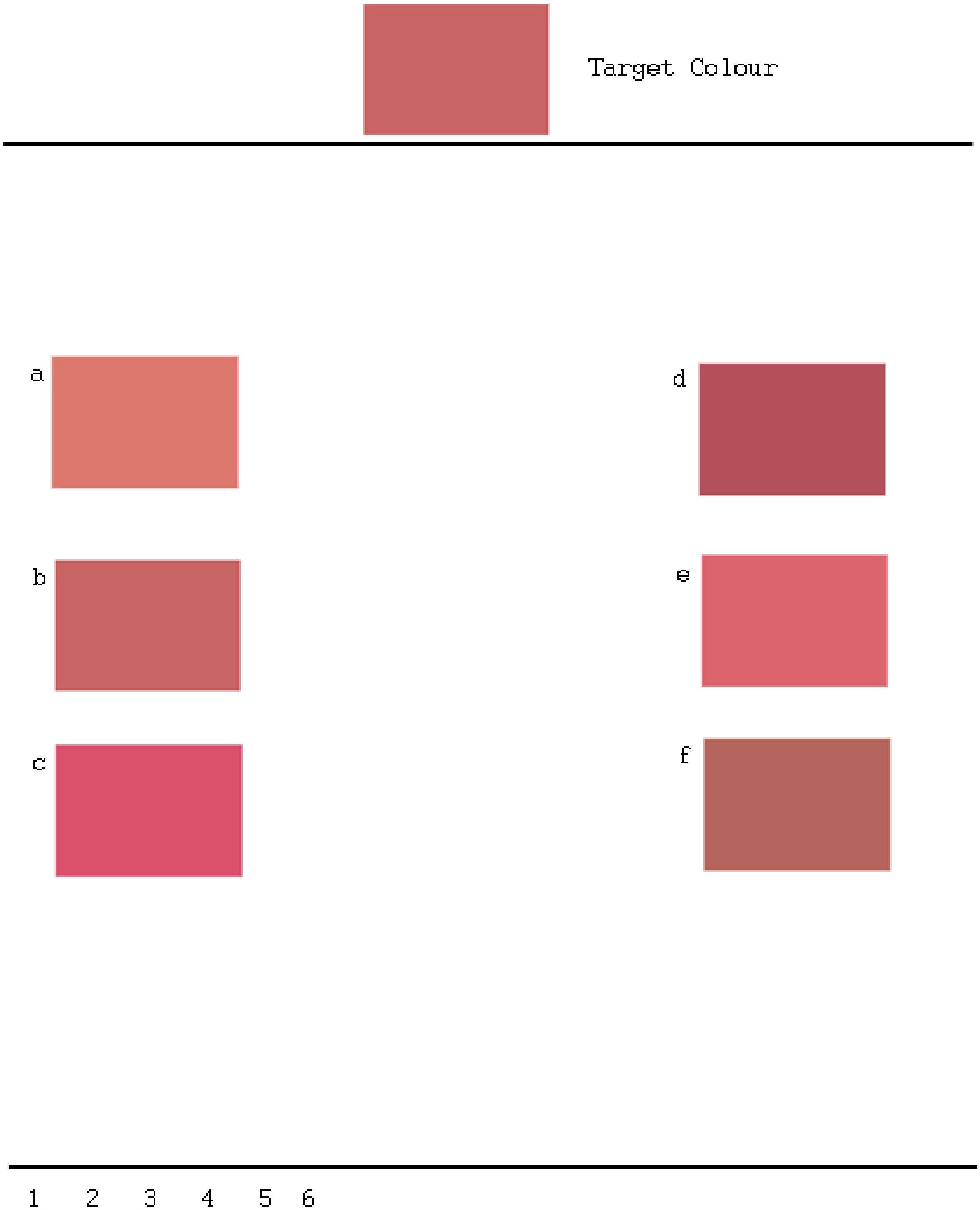
Appendix A: Graphic Stimulus and Additional Booklet Materials.

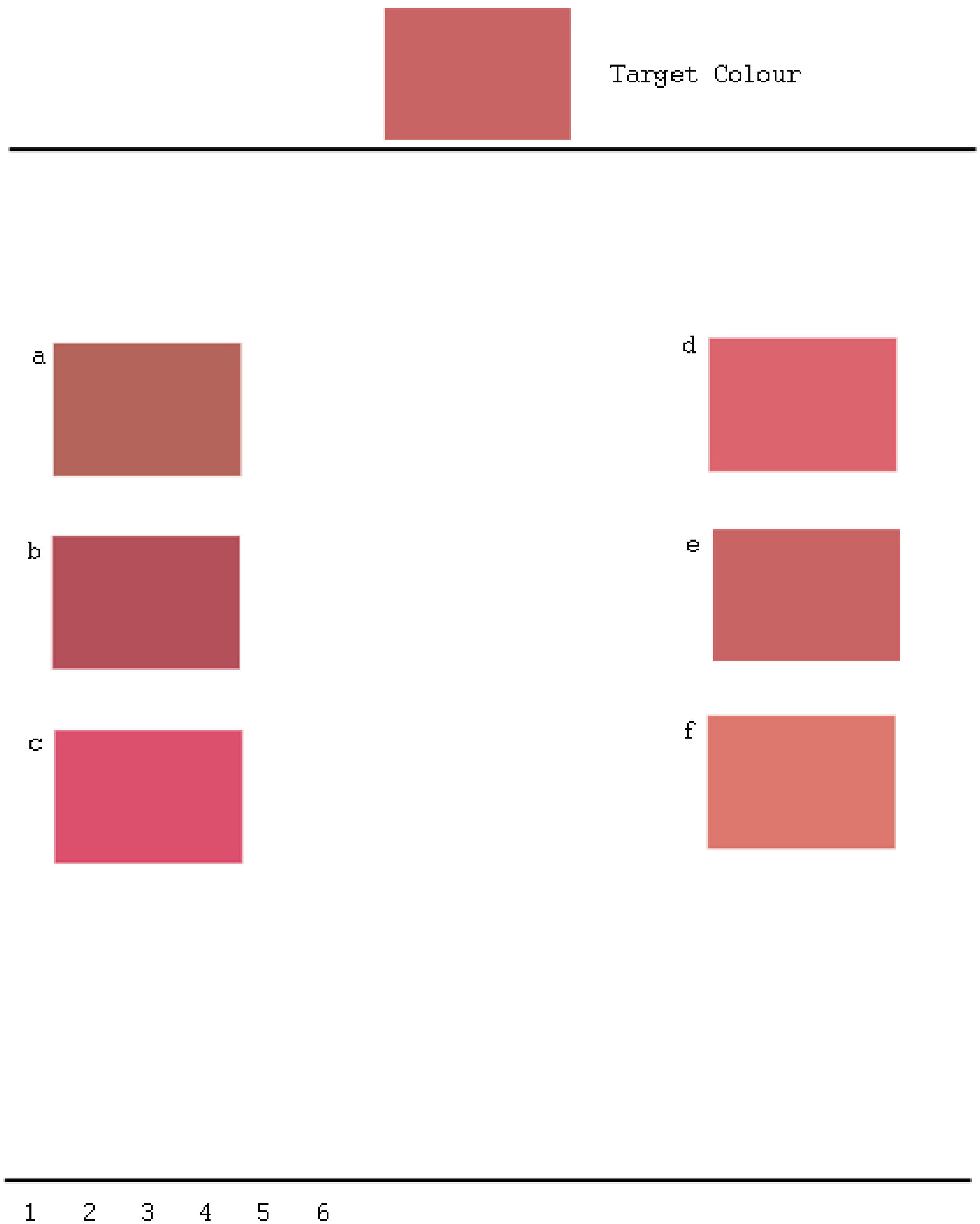
- Page 87: Control groups shape stimulus versions A
- Page 88: Control groups shape stimulus versions B
- Page 89: Control group colour stimulus versions A
- Page 90: Control group colour stimulus versions B
- Page 91: Experimental Stimulus set A, version One
- Page 92: Experimental Stimulus set A, version Two
- Page 93: Experimental Stimulus set B, version One
- Page 94: Experimental Stimulus set B, version Two
- Page 95: General Instructions set (cover page for all booklets)
- Page 96: Survey Page (last page for all booklets)
- Page 97: Instructions for the TA

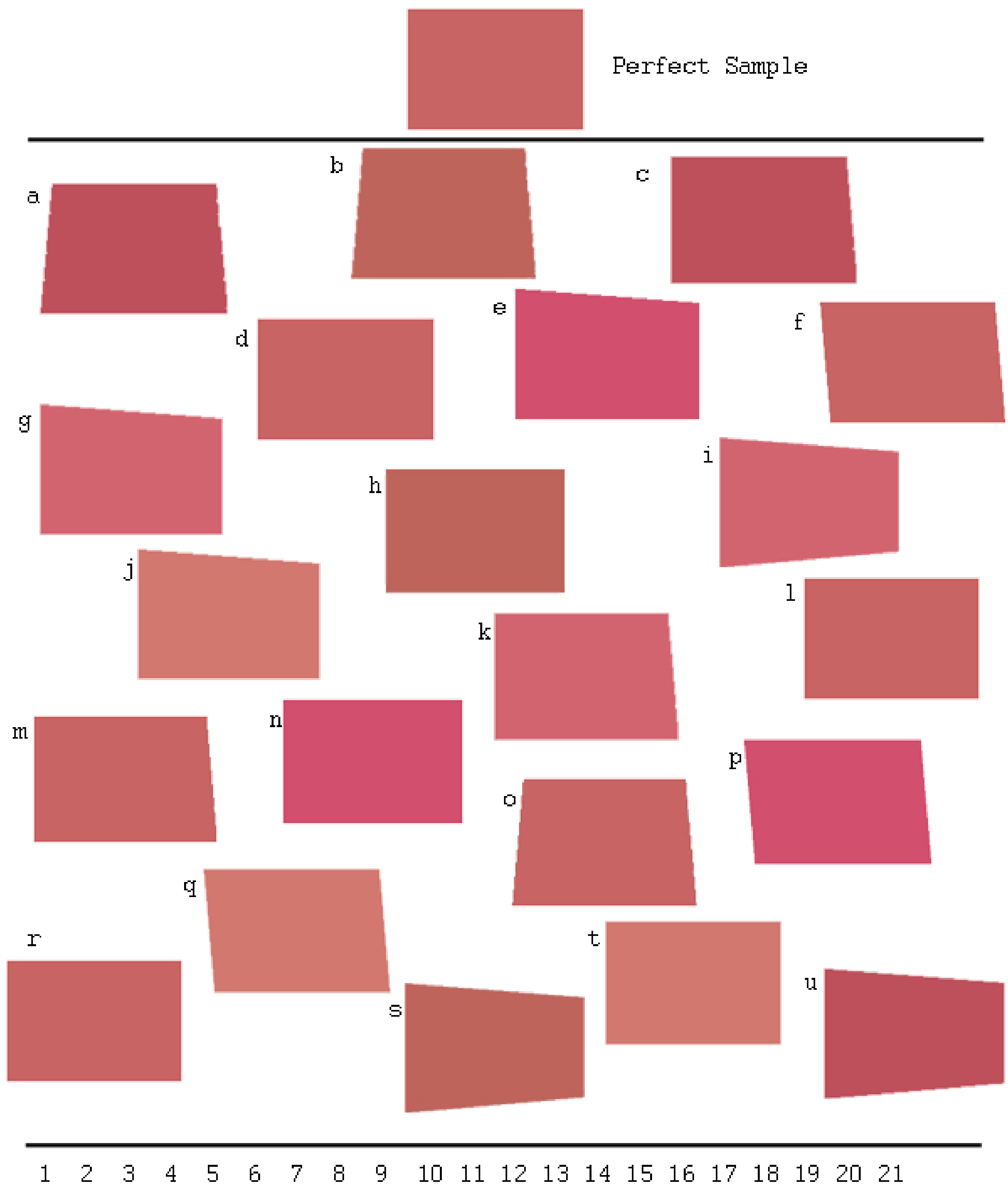


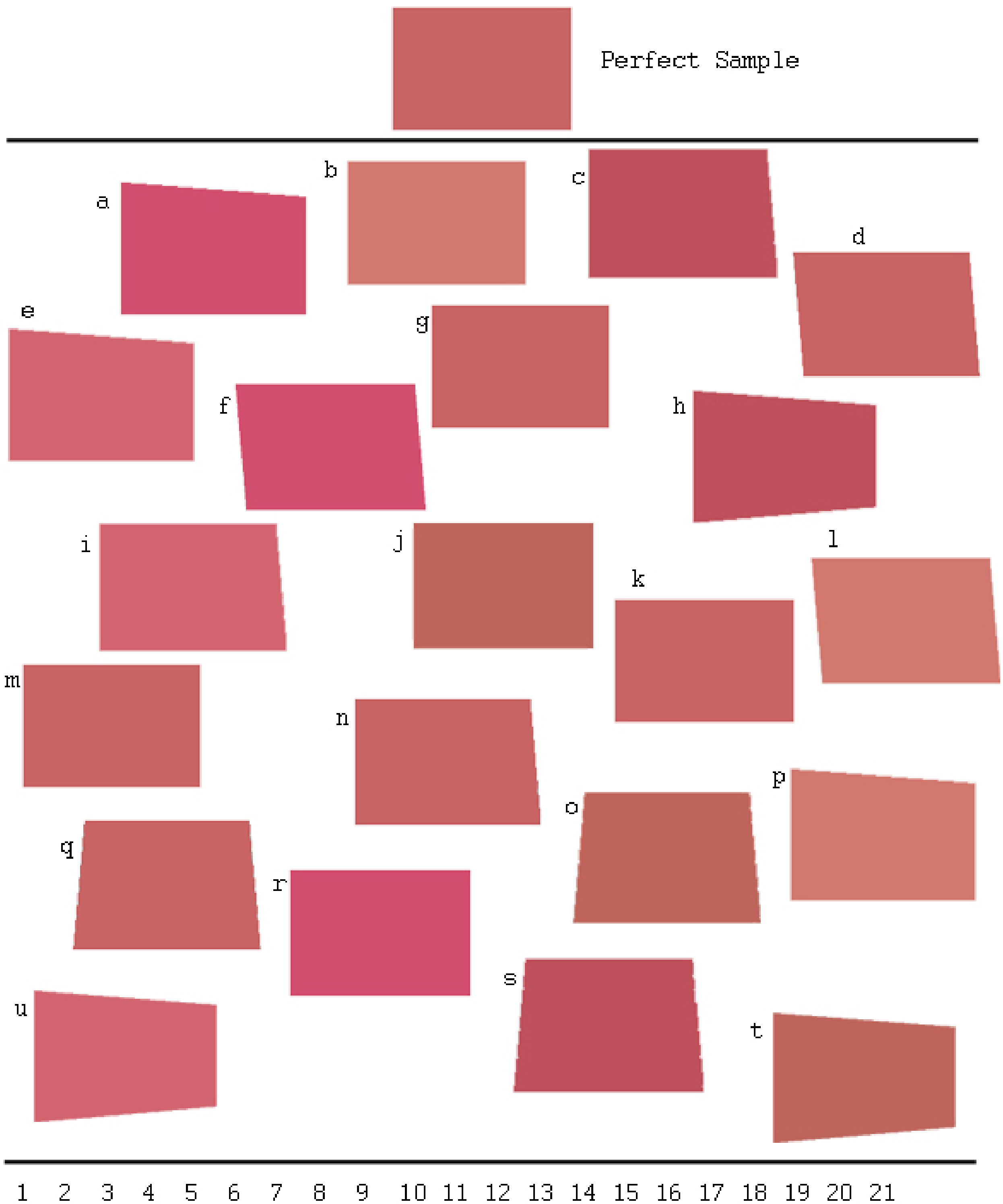
1 2 3 4 5 6

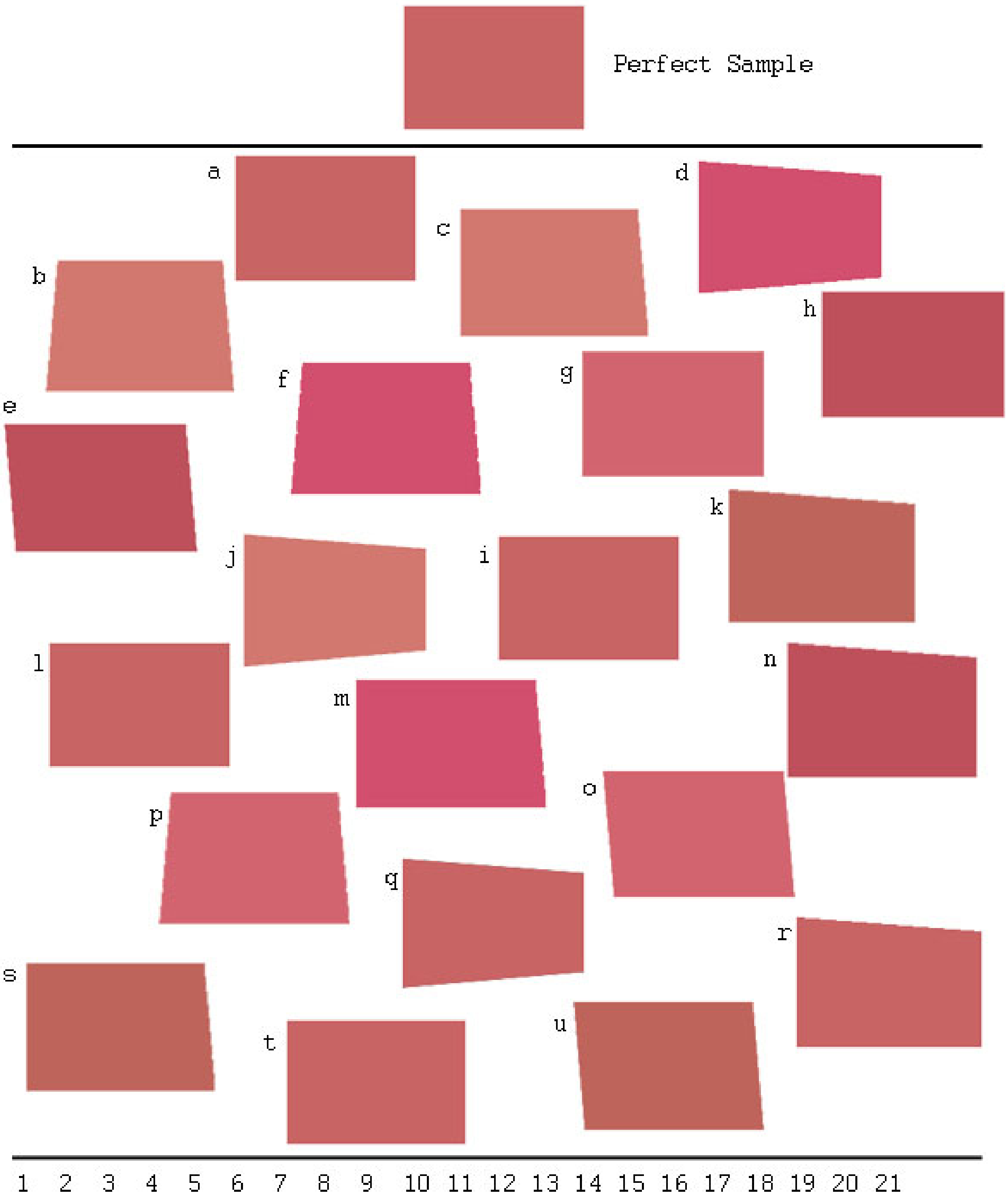

Target Shape
a
b
c
d
e
f
1 2 3 4 5 6

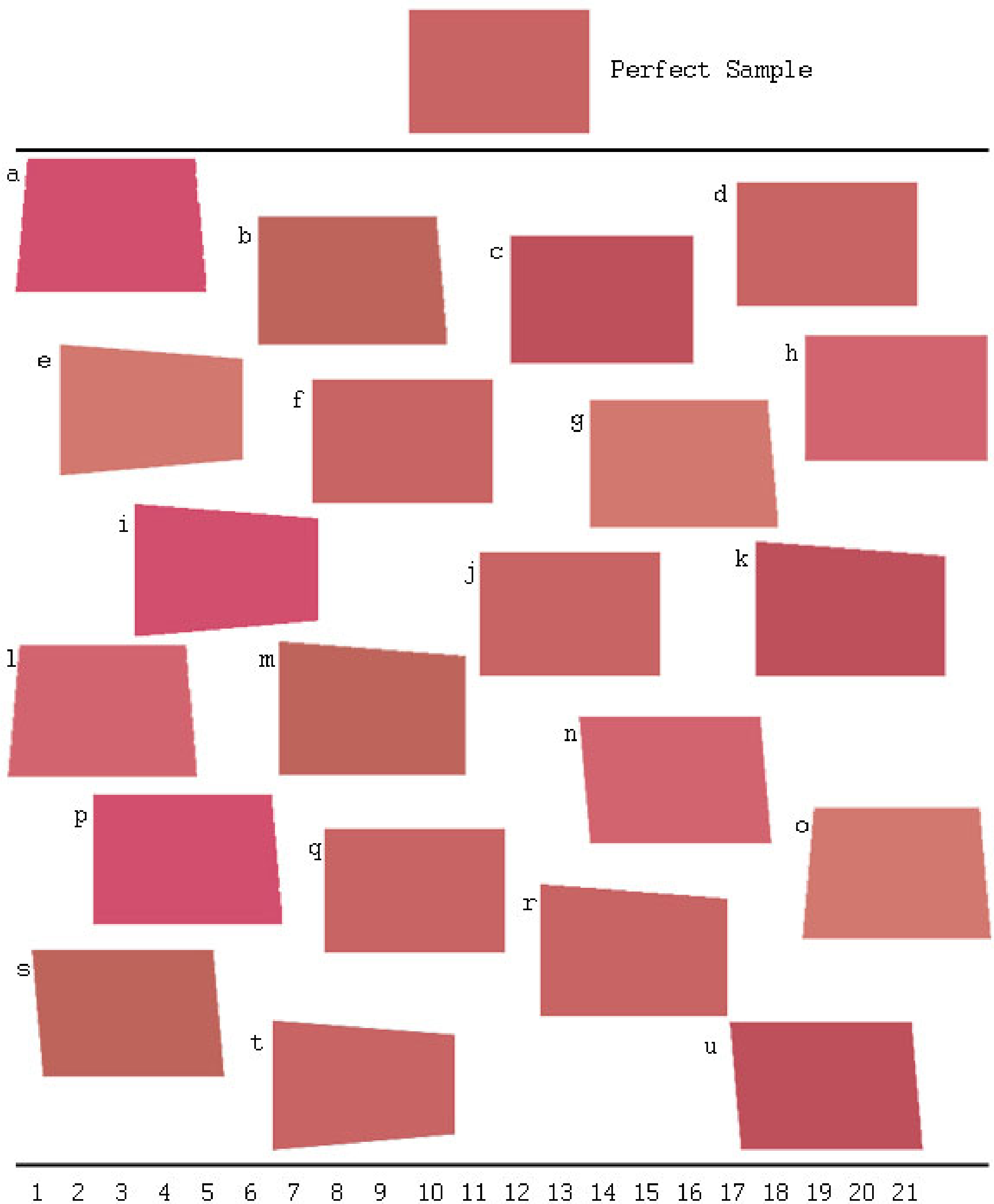

Target Colour
a
b
c
d
e
f
1 2 3 4 5 6


Target Colour
a
b
c
d
e
f
1 2 3 4 5 6


Perfect Sample
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21


Perfect Sample
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21


Perfect Sample
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21


Perfect Sample
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Management Science 311: Human Perception Experiment
Conducted by Mike Bobinski under the supervision of Professor Rob Duimering
The Department of Management Science at the University of Waterloo

Name: ____________________________

Student ID #: _______________________

Sex: Female Male

General Instructions:

Please read all the instructions carefully, **do not** start until instructed to do so by the person in charge of the experiment. Should you have any questions please wait until all the instructions are communicated and than raise your hand. **Please do not** ask questions out loud. When you finish please remain seated until all the students have completed the experiment

The experiment consists of two or three tasks and it should take you about 20 min. Please do not skip ahead until you have fully completed each task. Once you have moved ahead to the next task, please do not go back and change your answers for an earlier task

This experiment is design to test certain aspects of human perception and it is not designed to trick you in any way. Please follow instructions and be open and honest in your judgments. Please read all instructions twice through.

Please do not look over at what your neighbour is doing, as they are doing a **different** experimental task and it is important that you provide us with your own opinion

The Number of course bonus marks you will receive for participating (i.e., 0, 1, 2, or 3 marks) will depend on how well you perform on each task. If for some reason you do not feel comfortable participation in this experiment, you may quit at any time and earn the extra course marks by completing an alternate assignment related to organizational research (please see Prof. Duimering for details).

Thank you, very much for taking part in this experiment, if you have any questions about the participation in this study contact Mike Bobinski at mpbobins@engmail.uwaterloo.ca

This project satisfies all University of Waterloo's guidelines defining ethical research

Questionnaire:

In the first tasks which feature did you pay more attention to?

a) The Shape b) The Colour

Briefly explain why:

In the second tasks which feature did you pay more attention to?

a) The Shape b) The Colour

Briefly explain why:

Did you have any difficulties completing the tasks? Yes No

If so briefly explain why:

Thank You for participating in the experiment.

Instructions for the TA conduction the Experiment

1. Please sit the one sit apart (if it is possible)
2. Please remain quite during the experiment, if you finish your work ahead of the other students remain seated until the experiment booklets are collected
3. Please read all the instructions very carefully, should you have any questions rise your hand and wait until you are attended to, Please do not ask your questions out loud
4. Once you have completed a task or a section, do not go back and change your answers as it will render your test invalid, and you will have to complete another task, as described in genera instruction (first page of the booklets) to get the bonus marks.
5. After the results are analyzed, Mike will debrief you all, during one of the regular lectures.
6. Should you have any questions or concerns please contact Mike at mpbobins@engmail.uwaterloo.ca
7. Thank You vary much for participating

[TA] – Please make sure they are all stead and that they all read the general instruction on the first page of the booklet. Once you confirm that they all read the instructions you may tell them to start the experiment.

Please make sure that they do not go back and change their answers once they have completed a task.

If someone finishes ahead of time please collect their booklet, however the students needs to remain quite and seated not to disturb others

The experiment should take between 20 to 25 minutes but there is no time limit.

Appendix B: T-Test Results for the Comparison of Control Group Stimulus Sets

- Page 99: t-Test comparison of averages of the ranks between versions A and B of Control groups shape stimulus.
- Page 100: t-Test comparison of averages of the ranks between versions A and B of Control groups colour stimulus.

Shape A

	Var1	*Var2*
Mean	4.5556	4.6
Variance	0.2778	0.4889
Observations	9	10
Pooled Variance	0.3895	
Hyp. Mean Diff.	0	
df	17	
t Stat	-0.1550	
P(T<=t) one-tail	0.4393	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.8787	
t Critical two-tail	2.1098	

Shape C (Target Shape)

	Var1	*Var2*
Mean	1	1
Variance	0	0
Observations	9	10
Pooled Variance	0	
Hyp. Mean Diff.	0	
df	17	
t Stat	65535	
P(T<=t) one-tail	#NUM!	
t Critical one-tail	2	
P(T<=t) two-tail	#NUM!	
t Critical two-tail	2	

Shape E

	Var1	*Var2*
Mean	3.4444	3.8
Variance	0.7778	0.4000
Observations	9	10
Pooled Variance	0.5778	
Hyp. Mean Diff.	0	
df	17	
t Stat	-1.0181	
P(T<=t) one-tail	0.1615	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.3229	
t Critical two-tail	2.1098	

Shape B

	Var1	*Var2*
Mean	3.7778	3.6
Variance	0.6944	0.7111
Observations	9	10
Pooled Variance	0.7033	
Hyp. Mean Diff.	0	
df	17	
t Stat	0.4614	
P(T<=t) one-tail	0.3252	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.6504	
t Critical two-tail	2.1098	

Shape D

	Var1	*Var2*
Mean	2.1111	2
Variance	0.1111	0.0000
Observations	9	10
Pooled Variance	0.0523	
Hyp. Mean Diff.	0	
df	17	
t Stat	1.0576	
P(T<=t) one-tail	0.1525	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.3051	
t Critical two-tail	2.1098	

Shape F

	Var1	*Var2*
Mean	5.7778	6
Variance	0.1944	0.0000
Observations	9	10
Pooled Variance	0.0915	
Hyp. Mean Diff.	0	
df	17	
t Stat	-1.5989	
P(T<=t) one-tail	0.0641	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.1283	
t Critical two-tail	2.1098	

Colour 2

	Var1	*Var2*
Mean	3.5556	3.8
Variance	3.0278	1.5111
Observations	9	10
Pooled Variance	2.2248	
Hyp. Mean Diff.	0	
df	17	
t Stat	-0.3567	
P(T<=t) one-tail	0.3629	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.7257	
t Critical two-tail	2.1098	

Colour 3

	Var1	*Var2*
Mean	3.9444	4
Variance	2.2778	0.8889
Observations	9	10
Pooled Variance	1.5425	
Hyp. Mean Diff.	0	
df	17	
t Stat	-0.0974	
P(T<=t) one-tail	0.4618	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.9236	
t Critical two-tail	2.1098	

Colour 4

	Var1	*Var2*
Mean	3.7222	4
Variance	1.4444	1.1111
Observations	9	10
Pooled Variance	1.2680	
Hyp. Mean Diff.	0	
df	17	
t Stat	-0.5369	
P(T<=t) one-tail	0.2991	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.5983	
t Critical two-tail	2.1098	

Colour B

	Var1	*Var2*
Mean	3.1667	1.7
Variance	1.6250	0.4556
Observations	9	10
Pooled Variance	1.0059	
Hyp. Mean Diff.	0	
df	17	
t Stat	3.1827	
P(T<=t) one-tail	0.0027	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.0054	
t Critical two-tail	2.1098	

Colour 6

	Var1	*Var2*
Mean	4.8333	5.9
Variance	1.7500	0.1000
Observations	9	10
Pooled Variance	0.8765	
Hyp. Mean Diff.	0	
df	17	
t Stat	-2.4797	
P(T<=t) one-tail	0.0120	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.0239	
t Critical two-tail	2.1098	

Colour 1 (Target Colour)

	Var1	*Var2*
Mean	1.2222	1.6
Variance	0.1944	0.7111
Observations	9	10
Pooled Variance	0.4680	
Hyp. Mean Diff.	0	
df	17	
t Stat	-1.2019	
P(T<=t) one-tail	0.1229	
t Critical one-tail	1.7396	
P(T<=t) two-tail	0.2459	
t Critical two-tail	2.1098	

Appendix C: Results of Spearman Rank Correlations Arranged by Booklets

Booklet 2a and Booklet 2b

	T3 - Valance on Colour Sensor			T6- POV Colour Sensor		
Subject	CCR	SCR	PCR	CCR	SCR	PCR
1	0.576	0.727	0.848	0.878	0.295	0.684
2	0.633	0.565	0.722	0.854	0.265	0.638
3	0.587	0.445	0.611	0.947	0.234	0.678
4	0.327	0.899	0.869	0.938	0.325	0.758
5	0.934	0.25	0.691	0.88	0.243	0.683
6	0.824	0.153	0.571	0.807	0.252	0.642
7	0.593	0.613	0.81	0.887	0.363	0.719
8	0.222	0.891	0.73	0.228	0.778	0.65
9	0.653	0.847	0.879	0.891	0.273	0.651
10	0.547	0.798	0.786	0.807	0.355	0.642
11	0.718	0.859	0.916	0.899	0.395	0.734
12	0.876	0.409	0.727	0.702	0.327	0.601
13	0.49	0.982	0.882	0.953	0.288	0.707
14	0.745	0.664	0.785	0.775	0.33	0.588
Average	0.623214	0.650143	0.773357	0.817571	0.337357	0.669643

Booklet 3a and Booklet 3b

	T4- Valance on Shape Sensor			T7- POV Shape sensor		
Subject	CCR	SCR	PCR	CCR	SCR	PCR
1	0.605	0.831	0.915	0.742	0.615	0.862
2	0.167	1	0.806	0.131	0.881	0.752
3	0.47	0.59	0.755	0.147	0.944	0.767
4	0.159	0.773	0.871	0.921	0.143	0.626
5	0.2	0.983	0.817	0.351	0.871	0.894
6	0.597	0.766	0.904	0.24	0.977	0.845
7	0.381	0.833	0.848	0.23	0.926	0.829
8	0.645	0.553	0.745	0.37	0.994	0.828
9	0.552	0.951	0.883	N/A	N/A	N/A
10	0.51	0.948	0.824	0.534	0.861	0.791
11	0.701	0.744	0.848	0.355	0.991	0.82
12	0.47	0.982	0.871	0.274	0.983	0.784
13	0.442	0.982	0.864	0.462	0.947	0.846
14	0.256	0.819	0.644	0.321	0.946	0.782
Average	0.439643	0.839643	0.828214	0.390615	0.852231	0.802

Booklet 4a and 4b

Subject	T5- Valence on Both Sensors CCR	SCR	PCR	T6-POV Colour Sensor CCR	SCR	PCR
1	0.751	0.586	0.799	0.852	0.312	0.678
2	0.388	0.853	0.889	0.627	0.342	0.565
3	0.562	0.849	0.953	0.982	0.147	0.642
4	0.478	0.91	0.913	0.863	0.07	0.503
5	0.607	0.7	0.909	0.762	0.29	0.607
6	0.416	0.927	0.9	0.746	0.287	0.65
7	0.644	0.72	0.875	0.85	0.186	0.59
8	0.788	0.441	0.77	0.879	0.226	0.794
9	0.442	0.946	0.829	0.927	0.428	0.716
10	0.661	0.732	0.81	0.615	0.524	0.578
11	0.771	0.775	0.915	0.71	0.275	0.578
12	0.46	0.697	0.881	0.815	0.292	0.588
13	0.626	0.846	0.938	0.956	0.377	0.748
Average	0.584154	0.767846	0.875462	0.814154	0.288923	0.633615

Booklet 5a and Booklet 5b

Subject	T5- Valence on Both Sensors CCR	SCR	PCR	T7- POV Shape Sensor CCR	SCR	PCR
1	0.635	0.522	0.71	0.411	0.926	0.865
2	0.565	0.877	0.93	0.167	1	0.806
3	0.512	0.862	0.88	0.406	0.904	0.851
4	0.623	0.667	0.852	0.224	0.983	0.829
5	0.419	0.827	0.831	0.343	0.934	0.824
6	0.288	0.863	0.786	0.293	0.917	0.833
7	0.293	0.976	0.866	0.272	0.982	0.844
8	0.359	0.922	0.893	0.32	0.948	0.888
9	0.359	0.97	0.897	0.167	1	0.806
10	0.427	0.985	0.878	0.364	1	0.83
11	0.435	0.948	0.835	0.364	1	0.83
12	0.633	0.911	0.894	0.91	0.242	0.675
13	0.841	0.557	0.824	0.355	0.991	0.82
14	0.688	0.78	0.841	0.367	0.972	0.833
Average	0.5055	0.833357	0.851214	0.3545	0.914214	0.823857

Booklet 6a and Booklet 6b

	T2- Naïve Ranking			T3 - Valance on Colour Sensor		
Subject	CCR	SCR	PCR	CCR	SCR	PCR
1	0.742	0.501	0.771	0.752	-0.03	0.409
2	0.328	0.983	0.874	0.89	0.163	0.603
3	0.817	0.117	0.597	0.697	0.637	0.853
4	0.45	0.796	0.826	0.782	0.559	0.834
5	0.278	0.965	0.86	0.252	0.969	0.802
6	0.496	0.663	0.778	0.317	0.974	0.856
7	0.407	0.895	0.871	0.148	0.242	0.248
8	0.431	0.891	0.881	0.335	0.964	0.865
9	0.265	0.982	0.843	0.324	0.983	0.871
10	0.486	0.781	0.877	0.807	0.59	0.871
11	0.591	0.597	0.787	0.939	0.081	0.546
12	0.386	0.862	0.86	0.763	0.628	0.848
13	0.504	0.593	0.717	0.851	0.325	0.689
14	0.681	0.583	0.809	0.914	0.333	0.714
15	0.3	0.807	0.823	0.499	0.888	0.935
16	0.492	0.97	0.878	0.756	0.76	0.913
17	0.625	0.673	0.801	0.822	0.502	0.721
18	0.607	0.919	0.892	0.471	0.931	0.868
19	0.742	0.768	0.896	0.747	0.658	0.878
20	0.333	0.978	0.797	0.489	0.97	0.958
21	0.492	0.779	0.776	0.456	0.949	0.965
22	0.707	0.654	0.787	0.632	0.903	0.917
23	0.521	0.949	0.869	0.613	0.87	0.878
24	0.637	0.757	0.798	0.644	0.218	0.286
25	0.532	0.867	0.835	0.7	0.811	0.858
26	0.396	0.509	0.509	0.595	0.944	0.942
27	0.639	0.509	0.626	0.616	0.624	0.616
28	0.511	0.97	0.865	0.963	0.407	0.509
29	0.902	0.569	0.806	0.833	0.564	0.68
30	0.871	0.432	0.752	0.948	0.407	0.523
31	0	0	0	0.76	0.645	0.714
32	0.711	0.498	0.722	0.851	0.482	0.585
Average	0.5275	0.713031	0.774469	0.661438	0.623469	0.742344

Booklet 7a and Booklet 7b

	T2- Naïve Ranking			T4- Valance on Shape Sensor		
Subject	CCR	SCR	PCR	CCR	SCR	PCR
1	0.353	0.93	0.871	0.24	0.867	0.776
2	0.27	0.983	0.84	0.508	0.866	0.86
3	0.847	0.074	0.528	0.238	0.948	0.784
4	0.558	0.796	0.868	0.255	0.98	0.854
5	0.637	0.127	0.449	0.709	-0.018	0.397
6	0.46	0.917	0.908	0.477	0.799	0.867
7	0.342	0.85	0.823	0.306	0.859	0.807
8	0.599	0.707	0.832	0.65	0.735	0.898
9	0.273	0.8	0.678	0.426	0.905	0.838
10	0.337	0.95	0.898	0.368	0.936	0.905
11	0.224	0.933	0.84	0.316	0.972	0.859
12	0.316	0.978	0.864	0.29	0.982	0.844
13	0.613	0.508	0.726	0.337	0.938	0.831
14	0.759	0.463	0.738	0.757	0.483	0.741
15	0.521	0.825	0.884	0.471	0.772	0.857
16	0.395	0.823	0.768	0.575	0.863	0.86
17	0.48	0.982	0.875	0.364	1	0.83
18	0.665	0.87	0.886	0.519	0.971	0.857
19	0.486	0.982	0.881	0.486	0.982	0.881
20	0.683	0.732	0.873	0.353	0.967	0.798
21	0.716	0.641	0.805	0.683	0.645	0.812
22	0.627	0.795	0.832	0.364	1	0.83
23	0.479	0.952	0.867	0.631	0.766	0.81
24	0.656	0.781	0.838	0.434	0.946	0.817
25	0.897	0.506	0.793	0.562	0.947	0.892
26	0.559	0.944	0.868	0.548	0.93	0.857
27	0.593	0.926	0.9	0.498	0.961	0.875
28	0.482	0.771	0.732	0.705	0.717	0.822
29	0.562	0.934	0.855	0.422	0.982	0.84
30	0.398	0.836	0.781	0.119	0.621	0.446
Average	0.526233	0.7772	0.810033	0.4537	0.844067	0.8115

Booklet 8a and Booklet 8b

	T2- Naïve Ranking			T5- Valence on Both Sensors		
Subject	CCR	SCR	PCR	CCR	SCR	PCR
1	0.444	0.803	0.875	0.643	0.774	0.906
2	0.684	0.298	0.609	0.563	0.674	0.828
3	0.238	0.954	0.793	0.457	0.884	0.855
4	0.339	-0.135	-0.005	0.004	-0.099	0.006
5	0.698	0.506	0.732	0.452	0.915	0.896
6	0.625	0.678	0.858	0.644	0.654	0.786
7	0.797	0.556	0.864	0.687	0.633	0.8
8	0.465	0.785	0.866	0.617	0.661	0.784
9	0.333	0.978	0.866	0.329	0.982	0.827
10	0.677	0.647	0.855	0.697	0.52	0.744
11	0.354	0.875	0.903	0.351	0.893	0.843
12	0.262	0.871	0.73	0.339	0.899	0.779
13	0.562	0.785	0.781	0.647	0.763	0.838
14	0.506	0.97	0.883	0.482	0.941	0.827
15	0.468	0.934	0.866	0.657	0.866	0.871
16	0.677	0.772	0.842	0.293	0.866	0.752
17	0.506	0.566	0.586	0.52	0.618	0.71
18	0.931	0.384	0.747	0.802	0.604	0.837
19	0.57	0.891	0.853	0.527	0.841	0.812
20	0.47	0.858	0.785	0.5	0.971	0.863
21	0.645	0.807	0.864	0.522	0.931	0.824
22	0.683	0.82	0.897	0.548	0.891	0.883
23	0.478	0.982	0.878	0.813	0.786	0.955
24	0.424	0.79	0.743	0.541	0.882	0.838
25	0.772	0.78	0.951	0.61	0.713	0.778
26	0.54	0.756	0.766	0.606	0.87	0.833
Average	0.544154	0.727346	0.784154	0.532731	0.766654	0.795192

Booklet 9

Subject	T2- Naïve Ranking CCR	SCR	PCR	T8 - Naïve Ranking by Colour CCR	SCR	PCR
1	0.526	0.758	0.893	0.748	0.168	0.483
2	0.429	0.847	0.917	0.712	0.057	0.407
3	0.366	0.733	0.815	0.587	0.011	0.313
4	0.641	0.52	0.728	0.72	-0.074	0.31
5	0.423	0.706	0.769	0.664	0.161	0.459
6	0.313	0.858	0.795	0.794	0.138	0.468
7	0.736	0.583	0.874	0.783	-0.036	0.473
8	0.167	0.1	0.806	0.564	0.149	0.288
9	0.436	0.903	0.903	0.743	0.167	0.474
10	0.149	0.938	0.939	0.82	0.167	0.518
11	0.32	0.928	0.928	0.758	0.171	0.5
12	0.429	0.797	0.797	0.76	0.168	0.49
13	0.38	0.76	0.76	0.7	0.1	0.394
14	0.54	0.776	0.776	0.802	0.149	0.492
15	0.709	0.604	0.604	0.739	0.09	0.416
Average	0.4376	0.720733	0.820267	0.726267	0.105733	0.432333

Booklet 10

Subject	T2- Naïve Ranking CCR	SCR	PCR	T9 - Naive Ranking by Shape CCR	SCR	PCR
1	0.459	0.948	0.859	0.364	1	0.83
2	0.41	0.966	0.804	0.336	0.898	0.728
3	0.379	0.911	0.816	0.47	0.92	0.846
4	0.576	0.931	0.892	0.478	0.841	0.758
5	0.828	0.472	0.745	0.875	0.571	0.85
6	0.382	0.891	0.778	0.382	0.891	0.778
7	0.47	0.958	0.858	0.474	0.982	0.877
8	0.302	0.732	0.593	0.381	0.982	0.847
9	0.483	0.938	0.836	0.355	0.97	0.786
10	0.373	0.955	0.796	0.442	0.962	0.827
11	0.595	0.871	0.877	0.442	0.974	0.838
12	0.833	0.757	0.905	0.364	0.949	0.83
13	0.436	0.96	0.796	0.355	0.914	0.776
14	0.758	0.715	0.841	0.652	0.824	0.885
15	0.506	0.949	0.881	0.382	0.891	0.778
Average	0.523643	0.857571	0.815571	0.456286	0.897786	0.814571

Appendix D: t-Test Results for the “Within Subject “Analysis

- Majority of the t-Tests are computed assuming that the two samples have equal variance. In some cases the t-Test have been computed assuming that the two samples have unequal variance all those these are indicted by double asterisk (**)

Booklet 2 **T3-T6**

Colour Control

	Var1	Var2
Mean	0.6232	0.8176
Variance	0.0386	0.0337
Observations	14	14
Pooled Variance	0.0362	
Hyp. Mean Diff.	0	
df	26	
t Stat	-2.7035	
P(T<=t) one-tail	0.0060	
t Critical one-tail	1.7056	
P(T<=t) two-tail	0.0119	
t Critical two-tail	2.0555	

Shape Control

	Var1	Var2
Mean	0.6501	0.3374
Variance	0.0659	0.0184
Observations	14	14
Pooled Variance	0.0421	
Hyp. Mean Diff.	0	
df	26	
t Stat	4.0311	
P(T<=t) one-tail	0.0002	
t Critical one-tail	1.7056	
P(T<=t) two-tail	0.0004	
t Critical two-tail	2.0555	

Predictive Ranking Control

	Var1	Var2
Mean	0.7734	0.6696
Variance	0.0108	0.0024
Observations	14	14
Pooled Variance	0.0066	
Hyp. Mean Diff.	0	
df	26	
t Stat	3.3806	
P(T<=t) one-tail	0.0011	
t Critical one-tail	1.7056	
P(T<=t) two-tail	0.0023	
t Critical two-tail	2.0555	

Booklet 3 **T4-T7**

Colour Control

	Var1	Var2
Mean	0.4396	0.3906
Variance	0.0331	0.0523
Observations	14	13
Pooled Variance	0.0423	
Hyp. Mean Diff.	0	
df	25	
t Stat	0.6190	
P(T<=t) one-tail	0.2707	
t Critical one-tail	1.7081	
P(T<=t) two-tail	0.5415	
t Critical two-tail	2.0595	

Shape Control

	Var1	Var2
Mean	0.8396	0.8522
Variance	0.0212	0.0554
Observations	14	13
Pooled Variance	0.0376	
Hyp. Mean Diff.	0	
df	25	
t Stat	-0.1686	
P(T<=t) one-tail	0.4337	
t Critical one-tail	1.7081	
P(T<=t) two-tail	0.8675	
t Critical two-tail	2.0595	

Predictive Ranking Control

	Var1	Var2
Mean	0.8282	0.8020
Variance	0.0053	0.0044
Observations	14	13
Pooled Variance	0.0049	
Hyp. Mean Diff.	0	
df	25	
t Stat	0.9743	
P(T<=t) one-tail	0.1696	
t Critical one-tail	1.7081	
P(T<=t) two-tail	0.3392	
t Critical two-tail	2.0595	

Booklet 4 **T5-T6**

Colour Control

	Var1	Var2
Mean	0.5842	0.8142
Variance	0.0192	0.0136
Observations	13	13
Pooled Variance	0.0164	
Hyp. Mean Diff.	0	
df	24	
t Stat	-4.5799	
P(T<=t) one-tail	0.0001	
t Critical one-tail	1.7109	
P(T<=t) two-tail	0.0001	
t Critical two-tail	2.0639	

Shape Control

	Var1	Var2
Mean	0.7678	0.2889
Variance	0.0209	0.0140
Observations	13	13
Pooled Variance	0.0175	
Hyp. Mean Diff.	0	
df	24	
t Stat	9.2421	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.7109	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0639	

Predictive Ranking Control

	Var1	Var2
Mean	0.8755	0.6336
Variance	0.0032	0.0067
Observations	13	13
Pooled Variance	0.0049	
Hyp. Mean Diff.	0	
df	24	
t Stat	8.7639	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.7109	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0639	

Booklet 5 **T5-T7**

Colour Control

	Var1	*Var2*
Mean	0.5055	0.3545
Variance	0.0269	0.0318
Observations	14	14
Pooled Variance	0.0294	
Hyp. Mean Diff.	0	
df	26	
t Stat	2.3301	
P(T<=t) one-tail	0.0139	
t Critical one-tail	1.7056	
P(T<=t) two-tail	0.0278	
t Critical two-tail	2.0555	

Shape Control

	Var1	*Var2*
Mean	0.8334	0.9142
Variance	0.0228	0.0386
Observations	14	14
Pooled Variance	0.0307	
Hyp. Mean Diff.	0	
df	26	
t Stat	-1.2210	
P(T<=t) one-tail	0.1165	
t Critical one-tail	1.7056	
P(T<=t) two-tail	0.2331	
t Critical two-tail	2.0555	

Predictive Ranking Control

	Var1	*Var2*
Mean	0.8512	0.8239
Variance	0.0030	0.0023
Observations	14	14
Pooled Variance	0.0027	
Hyp. Mean Diff.	0	
df	26	
t Stat	1.4011	
P(T<=t) one-tail	0.0865	
t Critical one-tail	1.7056	
P(T<=t) two-tail	0.1730	
t Critical two-tail	2.0555	

Booklet 6 **T2-T3**

Colour Control

	Var1	*Var2*
Mean	0.5275	0.6614
Variance	0.0381	0.0481
Observations	32	32
Pooled Variance	0.0431	
Hyp. Mean Diff.	0	
df	62	
t Stat	-2.5812	
P(T<=t) one-tail	0.0061	
t Critical one-tail	1.6698	
P(T<=t) two-tail	0.0122	
t Critical two-tail	1.9990	

Shape Control

	Var1	*Var2*
Mean	0.7130	0.6235
Variance	0.0593	0.0880
Observations	32	32
Pooled Variance	0.0737	
Hyp. Mean Diff.	0	
df	62	
t Stat	1.3197	
P(T<=t) one-tail	0.0959	
t Critical one-tail	1.6698	
P(T<=t) two-tail	0.1918	
t Critical two-tail	1.9990	

Predictive Ranking Control

	Var1	*Var2*
Mean	0.7745	0.7423
Variance	0.0277	0.0377
Observations	32	32
Pooled Variance	0.0327	
Hyp. Mean Diff.	0	
df	62	
t Stat	1	
P(T<=t) one-tail	0.2401	
t Critical one-tail	1.6698	
P(T<=t) two-tail	0.4802	
t Critical two-tail	1.9990	

Booklet 7 **T2-T4**

Colour Control

	Var1	*Var2*
Mean	0.5262	0.4537
Variance	0.0292	0.0259
Observations	30	30
Pooled Variance	0.0275	
Hyp. Mean Diff.	0	
df	58	
t Stat	1.6939	
P(T<=t) one-tail	0.0478	
t Critical one-tail	1.6716	
P(T<=t) two-tail	0.0957	
t Critical two-tail	2.0017	

Shape Control

	Var1	*Var2*
Mean	0.7772	0.8441
Variance	0.0545	0.0432
Observations	30	30
Pooled Variance	0.0488	
Hyp. Mean Diff.	0	
df	58	
t Stat	-1.1723	
P(T<=t) one-tail	0.1229	
t Critical one-tail	1.6716	
P(T<=t) two-tail	0.2459	
t Critical two-tail	2.0017	

Predictive Ranking Control

	Var1	*Var2*
Mean	0.8100	0.8115
Variance	0.0111	0.0126
Observations	30	30
Pooled Variance	0.0118	
Hyp. Mean Diff.	0	
df	58	
t Stat	-0.0522	
P(T<=t) one-tail	0.4793	
t Critical one-tail	1.6716	
P(T<=t) two-tail	0.9585	
t Critical two-tail	2.0017	

Booklet 8 **T2-T5**

Colour Control

	Var1	*Var2*
Mean	0.5442	0.5327
Variance	0.0291	0.0296
Observations	26	26
Pooled Variance	0.0293	
Hyp. Mean Diff.	0	
df	50	
t Stat	0.2405	
P(T<=t) one-tail	0.4055	
t Critical one-tail	1.6759	
P(T<=t) two-tail	0.8109	
t Critical two-tail	2.0086	

Shape Control

	Var1	*Var2*
Mean	0.7273	0.7667
Variance	0.0625	0.0477
Observations	26	26
Pooled Variance	0.0551	
Hyp. Mean Diff.	0	
df	50	
t Stat	-0.6038	
P(T<=t) one-tail	0.2744	
t Critical one-tail	1.6759	
P(T<=t) two-tail	0.5487	
t Critical two-tail	2.0086	

Predictive Ranking Control

	Var1	*Var2*
Mean	0.7842	0.7952
Variance	0.0334	0.0287
Observations	26	26
Pooled Variance	0.0311	
Hyp. Mean Diff.	0	
df	50	
t Stat	-0.2258	
P(T<=t) one-tail	0.4111	
t Critical one-tail	1.6759	
P(T<=t) two-tail	0.8223	
t Critical two-tail	2.0086	

Booklet 9 **T2-T8**

Colour Control

	Var1	*Var2*
Mean	0.4376	0.7263
Variance	0.0298	0.0054
Observations	15	15
Pooled Variance	0.0176	
Hyp. Mean Diff.	0	
df	28	
t Stat	-5.9579	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.7011	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0484	

Shape Control

	Var1	*Var2*
Mean	0.7207	0.1057
Variance	0.0449	0.0065
Observations	15	15
Pooled Variance	0.0257	
Hyp. Mean Diff.	0	
df	28	
t Stat	10.5037	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.7011	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0484	

Predictive Ranking Control

	Var1	*Var2*
Mean	0.8203	0.4323
Variance	0.0082	0.0057
Observations	15	15
Pooled Variance	0.0069	
Hyp. Mean Diff.	0	
df	28	
t Stat	12.7582	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.7011	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0484	

Booklet 10 **T2-T9**

Colour Control

	Var1	*Var2*
Mean	0.5193	0.4501
Variance	0.0281	0.0202
Observations	15	15
Pooled Variance	0.0242	
Hyp. Mean Diff.	0	
df	28	
t Stat	1.2193	
P(T<=t) one-tail	0.1164	
t Critical one-tail	1.7011	
P(T<=t) two-tail	0.2329	
t Critical two-tail	2.0484	

Shape Control

	Var1	*Var2*
Mean	0.8636	0.9046
Variance	0.0192	0.0113
Observations	15	15
Pooled Variance	0.0152	
Hyp. Mean Diff.	0	
df	28	
t Stat	-0.9095	
P(T<=t) one-tail	0.1854	
t Critical one-tail	1.7011	
P(T<=t) two-tail	0.3708	
t Critical two-tail	2.0484	

Predictive Ranking Control

	Var1	*Var2*
Mean	0.8185	0.8156
Variance	0.0060	0.0021
Observations	15	15
Pooled Variance	0	
Hyp. Mean Diff.	0	
df	28	
t Stat	0.1239	
P(T<=t) one-tail	0.4511	
t Critical one-tail	1.7011	
P(T<=t) two-tail	0.9022	
t Critical two-tail	2.0484	

Appendix E: T-Tests Results for the “Between Task” Analysis

- Majority of the t-Tests are computed assuming that the two samples have equal variance. In some cases the t-Test have been computed assuming that the two samples have unequal variance all those these are indicted by double asterisks (**)

T2-T3 Correlation to CCR

	Var1	*Var2*
Mean	0.5320	0.6498
Variance	0.0317	0.0446
Observations	88	46
Pooled Variance	0.0361	
Hyp. Mean Diff.	0	
df	132	
t Stat	-3.4082	
P(T<=t) one-tail	0.0004	
t Critical one-tail	1.6565	
P(T<=t) two-tail	0.0009	
t Critical two-tail	1.9781	

T2-T3 Correlation to SCR

	Var1	*Var2*
Mean	0.7391	0.6316
Variance	0.0581	0.0798
Observations	88	46
Pooled Variance	0.0655	
Hyp. Mean Diff.	0	
df	132	
t Stat	2.3100	
P(T<=t) one-tail	0.0112	
t Critical one-tail	1.6565	
P(T<=t) two-tail	0.0224	
t Critical two-tail	1.9781	

T2-T3 Correlation to PCR

	Var1	*Var2*
Mean	0.7895	0.7518
Variance	0.0234	0.0293
Observations	88	46
Pooled Variance	0.0254	
Hyp. Mean Diff.	0	
df	132	
t Stat	1.2987	
P(T<=t) one-tail	0.0981	
t Critical one-tail	1.6565	
P(T<=t) two-tail	0.1963	
t Critical two-tail	1.9781	

T2-T4 Correlation to CCR

	Var1	*Var2*
Mean	0.5320	0.4492
Variance	0.0317	0.0275
Observations	88	44
Pooled Variance	0.0303	
Hyp. Mean Diff.	0	
df	130.0000	
t Stat	2.5751	
P(T<=t) one-tail	0.0056	
t Critical one-tail	1.6567	
P(T<=t) two-tail	0.0111	
t Critical two-tail	1.9784	

T2-T4 Correlation to SCR

	Var1	*Var2*
Mean	0.7391	0.8222
Variance	0.0581	0.0473
Observations	88	44
Pooled Variance	0.0545	
Hyp. Mean Diff.	0	
df	130	
t Stat	-1.9271	
P(T<=t) one-tail	0.0281	
t Critical one-tail	1.6567	
P(T<=t) two-tail	0.0562	
t Critical two-tail	1.9784	

T2-T4 Correlation to PCR

	Var1	*Var2*
Mean	0.7895	0.8168
Variance	0.0234	0.0102
Observations	88	44
Pooled Variance	0.0190	
Hyp. Mean Diff.	0	
df	130	
t Stat	-1.0743	
P(T<=t) one-tail	0.1423	
t Critical one-tail	1.6567	
P(T<=t) two-tail	0.2847	
t Critical two-tail	1.9784	

T2-T5 Correlation to CCR

	Var1	*Var2*
Mean	0.5320	0.5382
Variance	0.0317	0.0262
Observations	88	53
Pooled Variance	0.0296	
Hyp. Mean Diff.	0	
df	139	
t Stat	-0.2058	
P(T<=t) one-tail	0.4186	
t Critical one-tail	1.6559	
P(T<=t) two-tail	0.8372	
t Critical two-tail	1.9772	

T2-T5 Correlation to SCR

	Var1	*Var2*
Mean	0.7391	0.7846
Variance	0.0581	0.0343
Observations	88	53
Pooled Variance	0.0492	
Hyp. Mean Diff.	0	
df	139	
t Stat	-1.1783	
P(T<=t) one-tail	0.1203	
t Critical one-tail	1.6559	
P(T<=t) two-tail	0.2407	
t Critical two-tail	1.9772	

T2-T5Correlation to PCR

	Var1	*Var2*
Mean	0.7895	0.8297
Variance	0.0234	0.0165
Observations	88	53
Pooled Variance	0.0208	
Hyp. Mean Diff.	0	
df	139	
t Stat	-1.6025	
P(T<=t) one-tail	0.0557	
t Critical one-tail	1.6559	
P(T<=t) two-tail	0.1113	
t Critical two-tail	1.9772	

T2-T6 Correlation to CCR

	Var1	Var2
Mean	0.5320	0.8159
Variance	0.0317	0.0231
Observations	88	27
Pooled Variance	0.0297	
Hyp. Mean Diff.	0	
df	113	
t Stat	-7.4858	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6584	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9812	

T2-T6 Correlation to SCR

	Var1	Var2
Mean	0.7391	0.3140
Variance	0.0581	0.0163
Observations	88	27
Pooled Variance	0.0484	
Hyp. Mean Diff.	0	
df	113	
t Stat	8.7790	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6584	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9812	

T2-T6 Correlation to PCR

(**)

	Var1	Var2
Mean	0.7895	0.6523
Variance	0.0234	0.0046
Observations	88	27
Hyp. Mean Diff.	0	
df	98	
t Stat	6.5599	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6606	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9845	

T2-T7 Correlation to CCR

	Var1	Var2
Mean	0.5320	0.3719
Variance	0.0317	0.0404
Observations	88	27
Pooled Variance	0.0337	
Hyp. Mean Diff.	0	
df	113	
t Stat	3.9645	
P(T<=t) one-tail	0.0001	
t Critical one-tail	1.6584	
P(T<=t) two-tail	0.0001	
t Critical two-tail	1.9812	

T2-T7 Correlation to SCR

	Var1	Var2
Mean	0.7391	0.8844
Variance	0.0581	0.0458
Observations	88	27
Pooled Variance	0.0552	
Hyp. Mean Diff.	0	
df	113	
t Stat	-2.8087	
P(T<=t) one-tail	0.0029	
t Critical one-tail	1.6584	
P(T<=t) two-tail	0.0059	
t Critical two-tail	1.9812	

T2-T7 Correlation to PCR

(**)

	Var1	Var2
Mean	0.7895	0.8133
Variance	0.0234	0.0033
Observations	88	27
Hyp. Mean Diff.	0	
df	108	
t Stat	-1.2107	
P(T<=t) one-tail	0.1143	
t Critical one-tail	1.6591	
P(T<=t) two-tail	0.2287	
t Critical two-tail	1.9822	

T2-T8 Correlation to CCR

	Var1	Var2
Mean	0.5320	0.7263
Variance	0.0317	0.0054
Observations	88	15
Pooled Variance	0.0281	
Hyp. Mean Diff.	0	
df	101	
t Stat	-4.1526	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6601	
P(T<=t) two-tail	0.0001	
t Critical two-tail	1.9837	

T2-T8 Correlation to SCR

	Var1	Var2
Mean	0.7391	0.1057
Variance	0.0581	0.0065
Observations	88	15
Pooled Variance	0.0509	
Hyp. Mean Diff.	0	
df	101	
t Stat	10.0495	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6601	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9837	

T2-T8 Correlation to PCR

	Var1	Var2
Mean	0.7895	0.4323
Variance	0.0234	0.0057
Observations	88	15
Pooled Variance	0.0209	
Hyp. Mean Diff.	0	
df	101	
t Stat	8.8335	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6601	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9837	

T2-T9 Correlation to CCR

	Var1	*Var2*
Mean	0.5320	0.4501
Variance	0.0317	0.0202
Observations	88	15
Pooled Variance	0.0301	
Hyp. Mean Diff.	0	
df	101	
t Stat	1.6889	
P(T<=t) one-tail	0.0472	
t Critical one-tail	1.6601	
P(T<=t) two-tail	0.0943	
t Critical two-tail	1.9837	

T2-T9 Correlation to SCR

	Var1	*Var2*
Mean	0.7391	0.9046
Variance	0.0581	0.0113
Observations	88	15
Pooled Variance	0.0516	
Hyp. Mean Diff.	0	
df	101	
t Stat	-2.6084	
P(T<=t) one-tail	0.0052	
t Critical one-tail	1.6601	
P(T<=t) two-tail	0.0105	
t Critical two-tail	1.9837	

T2-T9 Correlation to PCR

	Var1	*Var2*
Mean	0.7895	0.8156
Variance	0.0234	0.0021
Observations	88	15
Pooled Variance	0.0204	
Hyp. Mean Diff.	0	
df	101	
t Stat	-0.6545	
P(T<=t) one-tail	0.2571	
t Critical one-tail	1.6601	
P(T<=t) two-tail	0.5143	
t Critical two-tail	1.9837	

T3-T4 Correlation to CCR

	Var1	*Var2*
Mean	0.6498	0.4492
Variance	0.0446	0.0275
Observations	46	44
Pooled Variance	0.0362	
Hyp. Mean Diff.	0	
df	88	
t Stat	4.9966	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6624	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9873	

T3-T4 Correlation to SCR

	Var1	*Var2*
Mean	0.6316	0.8222
Variance	0.0798	0.0473
Observations	46	44
Pooled Variance	0.0640	
Hyp. Mean Diff.	0	
df	88	
t Stat	-3.5745	
P(T<=t) one-tail	0.0003	
t Critical one-tail	1.6624	
P(T<=t) two-tail	0.0006	
t Critical two-tail	1.9873	

T3-T4 Correlation to PCR

	Var1	*Var2*
Mean	0.7518	0.8168
Variance	0.0293	0.0102
Observations	46	44
Pooled Variance	0.0200	
Hyp. Mean Diff.	0	
df	88	
t Stat	-2.1832	
P(T<=t) one-tail	0.0158	
t Critical one-tail	1.6624	
P(T<=t) two-tail	0.0317	
t Critical two-tail	1.9873	

T3-T6 Correlation to CCR

	Var1	*Var2*
Mean	0.6498	0.8159
Variance	0.0446	0.0231
Observations	46	27
Pooled Variance	0.0368	
Hyp. Mean Diff.	0	
df	71	
t Stat	-3.5743	
P(T<=t) one-tail	0.0003	
t Critical one-tail	1.6666	
P(T<=t) two-tail	0.0006	
t Critical two-tail	1.9939	

T3-T6 Correlation to SCR

	Var1	*Var2*
Mean	0.6316	0.3140
Variance	0.0798	0.0163
Observations	46	27
Pooled Variance	0.0566	
Hyp. Mean Diff.	0	
df	71	
t Stat	5.5070	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6666	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9939	

T3-T6 Correlation to PCR

	Var1	*Var2*
Mean	0.7518	0.6523
Variance	0.0293	0.0046
Observations	46	27
Pooled Variance	0.0203	
Hyp. Mean Diff.	0	
df	71	
t Stat	2.8826	
P(T<=t) one-tail	0.0026	
t Critical one-tail	1.6666	
P(T<=t) two-tail	0.0052	
t Critical two-tail	1.9939	

T3-T5 Correlation to CCR

(**)

	Var1	*Var2*
Mean	0.6498	0.5382
Variance	0.0446	0.0262
Observations	46	53
Hyp. Mean Diff.	0	
df	84	
t Stat	2.9175	
P(T<=t) one-tail	0.0023	
t Critical one-tail	1.6632	
P(T<=t) two-tail	0.0045	
t Critical two-tail	1.9886	

T3-T5 Correlation to SCR

	Var1	*Var2*
Mean	0.6316	0.7846
Variance	0.0798	0.0343
Observations	46	53
Pooled Variance	0.0554	
Hyp. Mean Diff.	0	
df	97	
t Stat	-3.2244	
P(T<=t) one-tail	0.0009	
t Critical one-tail	1.6607	
P(T<=t) two-tail	0.0017	
t Critical two-tail	1.9847	

T3-T5Correlation to PCR

	Var1	*Var2*
Mean	0.7518	0.8297
Variance	0.0293	0.0165
Observations	46	53
Pooled Variance	0.0225	
Hyp. Mean Diff.	0	
df	97	
t Stat	-2.5790	
P(T<=t) one-tail	0.0057	
t Critical one-tail	1.6607	
P(T<=t) two-tail	0.0114	
t Critical two-tail	1.9847	

T3-T8 Correlation to CCR

(**)

	Var1	*Var2*
Mean	0.6498	0.7263
Variance	0.0446	0.0054
Observations	46	15
Hyp. Mean Diff.	0	
df	59	
t Stat	-2.0955	
P(T<=t) one-tail	0.0202	
t Critical one-tail	1.6711	
P(T<=t) two-tail	0.0404	
t Critical two-tail	2.0010	

T3-T8 Correlation to SCR

	Var1	*Var2*
Mean	0.6316	0.1057
Variance	0.0798	0.0065
Observations	46	15
Pooled Variance	0.0624	
Hyp. Mean Diff.	0	
df	59	
t Stat	7.0774	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6711	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0010	

T3-T8 Correlation to PCR

	Var1	*Var2*
Mean	0.7518	0.4323
Variance	0.0293	0.0057
Observations	46	15
Pooled Variance	0.0237	
Hyp. Mean Diff.	0	
df	59	
t Stat	6.9801	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6711	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0010	

T4-T5 Correlation to CCR

	Var1	*Var2*
Mean	0.4492	0.5382
Variance	0.0275	0.0262
Observations	44	53
Pooled Variance	0.0268	
Hyp. Mean Diff.	0	
df	95	
t Stat	-2.6639	
P(T<=t) one-tail	0.0045	
t Critical one-tail	1.6611	
P(T<=t) two-tail	0.0091	
t Critical two-tail	1.9852	

T4-T5 Correlation to SCR

	Var1	*Var2*
Mean	0.8222	0.7846
Variance	0.0473	0.0343
Observations	44	53
Pooled Variance	0.0402	
Hyp. Mean Diff.	0	
df	95	
t Stat	0.9204	
P(T<=t) one-tail	0.1798	
t Critical one-tail	1.6611	
P(T<=t) two-tail	0.3597	
t Critical two-tail	1.9852	

T4-T5 Correlation to PCR

	Var1	*Var2*
Mean	0.8168	0.8297
Variance	0.0102	0.0165
Observations	44	53
Pooled Variance	0.0137	
Hyp. Mean Diff.	0	
df	95	
t Stat	-0.5395	
P(T<=t) one-tail	0.2954	
t Critical one-tail	1.6611	
P(T<=t) two-tail	0.5908	
t Critical two-tail	1.9852	

T4-T7 Correlation to CCR

(**)

	Var1	Var2
Mean	0.4492	0.3719
Variance	0.0275	0.0404
Observations	44	27
Hyp. Mean Diff.	0	
df	47	
t Stat	1.6795	
P(T<=t) one-tail	0.0498	
t Critical one-tail	1.6779	
P(T<=t) two-tail	0.0997	
t Critical two-tail	2.0117	

T4-T7 Correlation to SCR

	Var1	Var2
Mean	0.8222	0.8844
Variance	0.0473	0.0458
Observations	44	27
Pooled Variance	0.0468	
Hyp. Mean Diff.	0	
df	69	
t Stat	-1.1759	
P(T<=t) one-tail	0.1218	
t Critical one-tail	1.6672	
P(T<=t) two-tail	0.2437	
t Critical two-tail	1.9949	

T4-T7 Correlation to PCR

	Var1	Var2
Mean	0.8168	0.8133
Variance	0.0102	0.0033
Observations	44	27
Pooled Variance	0.0076	
Hyp. Mean Diff.	0	
df	69	
t Stat	0.1636	
P(T<=t) one-tail	0.4353	
t Critical one-tail	1.6672	
P(T<=t) two-tail	0.8705	
t Critical two-tail	1.9949	

T4-T9 Correlation to CCR

	Var1	Var2
Mean	0.4492	0.4501
Variance	0.0275	0.0202
Observations	44	15
Pooled Variance	0.0257	
Hyp. Mean Diff.	0	
df	57	
t Stat	-0.0189	
P(T<=t) one-tail	0.4925	
t Critical one-tail	1.6720	
P(T<=t) two-tail	0.9850	
t Critical two-tail	2.0025	

T4-T9 Correlation to SCR

(**)

	Var1	Var2
Mean	0.8427	0.9046
Variance	0.0355	0.0113
Observations	44	15
Hyp. Mean Diff.	0	
df	44	
t Stat	-1.5686	
P(T<=t) one-tail	0.0620	
t Critical one-tail	1.6802	
P(T<=t) two-tail	0.1239	
t Critical two-tail	2.0154	

T4-T9 Correlation to PCR

	Var1	Var2
Mean	0.8168	0.8156
Variance	0.0102	0.0021
Observations	44	15
Pooled Variance	0.0082	
Hyp. Mean Diff.	0	
df	57	
t Stat	0.0450	
P(T<=t) one-tail	0.4821	
t Critical one-tail	1.6720	
P(T<=t) two-tail	0.9642	
t Critical two-tail	2.0025	

T5-T6 Correlation to CCR

(**)

	Var1	Var2
Mean	0.5382	0.8159
Variance	0.0262	0.0231
Observations	53	27
Pooled Variance	0.0252	
Hyp. Mean Diff.	0	
df	78	
t Stat	-7.4015	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6646	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9908	

T5-T6 Correlation to SCR

(**)

	Var1	Var2
Mean	0.7846	0.3140
Variance	0.0343	0.0163
Observations	53	27
Pooled Variance	0.0283	
Hyp. Mean Diff.	0	
df	78	
t Stat	11.8288	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6646	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9908	

T5-T6 Correlation to PCR

(**)

	Var1	Var2
Mean	0.8297	0.6523
Variance	0.0165	0.0046
Observations	53	27
Pooled Variance	0.0126	
Hyp. Mean Diff.	0	
df	78	
t Stat	6.6904	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6646	
P(T<=t) two-tail	0.0000	
t Critical two-tail	1.9908	

T5-T7 Correlation to CCR

(**)

	Var1	*Var2*
Mean	0.5382	0.3719
Variance	0.0262	0.0404
Observations	53	27
Pooled Variance	0.0309	
Hyp. Mean Diff.	0	
df	78	
t Stat	3.9974	
P(T<=t) one-tail	0.0001	
t Critical one-tail	1.6646	
P(T<=t) two-tail	0.0001	
t Critical two-tail	1.9908	

T5-T7 Correlation to SCR

(**)

	Var1	*Var2*
Mean	0.7846	0.8844
Variance	0.0343	0.0458
Observations	53	27
Pooled Variance	0.0382	
Hyp. Mean Diff.	0	
df	78	
t Stat	-2.1611	
P(T<=t) one-tail	0.0169	
t Critical one-tail	1.6646	
P(T<=t) two-tail	0.0338	
t Critical two-tail	1.9908	

T5-T7 Correlation to PCR

(**)

	Var1	*Var2*
Mean	0.8297	0.8133
Variance	0.0165	0.0033
Observations	53	27
Pooled Variance	0.0121	
Hyp. Mean Diff.	0	
df	78	
t Stat	0.6274	
P(T<=t) one-tail	0.2661	
t Critical one-tail	1.6646	
P(T<=t) two-tail	0.5322	
t Critical two-tail	1.9908	

T6-T8 Correlation to CCR

(**)

	Var1	*Var2*
Mean	0.8159	0.7263
Variance	0.0231	0.0054
Observations	27	15
Hyp. Mean Diff.	0	
df	40	
t Stat	2.5685	
P(T<=t) one-tail	0.0070	
t Critical one-tail	1.6839	
P(T<=t) two-tail	0.0141	
t Critical two-tail	2.0211	

T6-T8 Correlation to SCR

	Var1	*Var2*
Mean	0.3140	0.1057
Variance	0.0163	0.0065
Observations	27	15
Pooled Variance	0.0129	
Hyp. Mean Diff.	0	
df	40	
t Stat	5.7007	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6839	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0211	

T6-T8 Correlation to PCR

	Var1	*Var2*
Mean	0.6523	0.4323
Variance	0.0046	0.0057
Observations	27	15
Pooled Variance	0.0050	
Hyp. Mean Diff.	0	
df	40	
t Stat	9.6759	
P(T<=t) one-tail	0.0000	
t Critical one-tail	1.6839	
P(T<=t) two-tail	0.0000	
t Critical two-tail	2.0211	

T7-T9 Correlation to CCR

(**)

	Var1	*Var2*
Mean	0.3719	0.4501
Variance	0.0404	0.0202
Observations	27	15
Hyp. Mean Diff.	0	
df	37	
t Stat	-1.4671	
P(T<=t) one-tail	0.0754	
t Critical one-tail	1.6871	
P(T<=t) two-tail	0.1508	
t Critical two-tail	2.0262	

T7-T9 Correlation to SCR

(**)

	Var1	*Var2*
Mean	0.8844	0.9046
Variance	0.0458	0.0113
Observations	27	15
Hyp. Mean Diff.	0	
df	40	
t Stat	-0.4087	
P(T<=t) one-tail	0.3425	
t Critical one-tail	1.6839	
P(T<=t) two-tail	0.6849	
t Critical two-tail	2.0211	

T7-T9 Correlation to PCR

	Var1	*Var2*
Mean	0.8133	0.8156
Variance	0.0033	0.0021
Observations	27	15
Pooled Variance	0.0029	
Hyp. Mean Diff.	0	
df	40	
t Stat	-0.1311	
P(T<=t) one-tail	0.4482	
t Critical one-tail	1.6839	
P(T<=t) two-tail	0.8964	
t Critical two-tail	2.0211	

Appendix F: Average Rank Regression Analysis for Stimulus Set A

Averages of the Ranks for Experimental Stimulus A

Stim A	T2	T3	T4	T5	T6	T7	T8
1A	10.68	9.48	10.91	11.90	5.13	10.24	2.60
1C(X)_(1)	1.88	1.87	1.61	1.76	1.75	2.53	1.07
1C(X)_(2)	2.10	2.22	1.83	1.90	2.06	2.29	1.93
1C(X)_(3)	1.98	2.00	1.57	2.24	2.13	2.94	1.73
1D	5.93	5.52	5.78	6.66	5.25	5.00	2.27
1E	7.81	7.26	7.86	7.82	3.44	7.65	2.20
2A	12.32	12.22	11.41	13.38	9.44	11.65	7.07
2C	4.86	7.17	3.35	3.83	9.81	3.88	6.60
2F	14.27	13.57	13.13	14.31	10.00	13.35	5.40
3A	13.83	15.04	13.55	15.83	13.56	10.71	9.40
3D	9.26	11.52	9.04	9.76	13.06	8.18	8.87
3F	15.71	15.00	15.17	16.64	13.13	13.35	9.60
4B	10.12	10.87	9.26	10.93	8.38	7.71	8.60
4C	4.05	6.96	3.91	3.90	6.81	3.76	9.13
4E	10.40	11.35	9.77	10.93	8.06	8.35	8.27
5B	9.14	8.48	9.18	10.21	7.69	7.18	6.33
5D	8.05	7.43	6.59	8.83	7.50	5.76	6.13
5F	13.81	12.83	13.87	14.93	8.88	12.12	6.47
6B	13.79	18.13	12.35	14.76	15.38	8.94	12.27
6C	8.40	11.04	6.30	8.21	14.81	5.35	12.20
6E	14.38	15.65	12.87	16.17	14.94	9.71	12.53

Ranks Assigned (Using SPSS) to Experimental Stimulus A

Stim A	T2	T3	T4	T5	T6	T7	T8
1A	14	10	14	14	5	16	6
1C(X)_(1)	1	1	2	1	1	2	1
1C(X)_(2)	3	3	3	2	2	1	3
1C(X)_(3)	2	2	1	3	3	3	2
1D	6	4	6	6	6	6	5
1E	7	7	9	7	4	10	4
2A	15	15	15	15	13	18	12
2C	5	6	4	4	14	5	11
2F	19	17	18	16	15	20.5	7
3A	18	19	19	19	18	17	17
3D	11	14	10	10	16	12	15
3F	21	18	21	21	17	20.5	18
4B	12	11	12	12.5	11	11	14
4C	4	5	5	5	7	4	16
4E	13	13	13	12.5	10	13	13
5B	10	9	11	11	9	9	9
5D	8	8	8	9	8	8	8
5F	17	16	20	18	12	19	10
6B	16	21	16	17	21	14	20
6C	9	12	7	8	19	7	19
6E	20	20	17	20	20	15	21

Results of the Regression Analysis

T2 vs. CCR

Regression Statistics	
Multiple R	0.5658
R Square	0.3201
Adjusted R Square	0.2843
Standard Error	5.2493
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	246.4583	246.4583	8.9443	0.0075
Residual	19	523.5417	27.5548		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.0417	2.2981	2.1938	0.0409	0.2317	9.8517
X Variable 1	1.8958	0.6339	2.9907	0.0075	0.569	3.2226

T2 vs. SCR

Regression Statistics	
Multiple R	0.8660
R Square	0.7500
Adjusted R Square	0.7369
Standard Error	3.1829
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	577.5146	577.5146	57.0058	0.0000
Residual	19	192.4854	10.1308		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.8792	1.3935	1.3486	0.1933	-1.0374	4.7957
X Variable 1	2.9021	0.3844	7.5502	0	2.0976	3.7066

T3 vs. CCR

Regression Statistics	
Multiple R	0.7398
R Square	0.5473
Adjusted R Square	0.5235
Standard Error	4.2830
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	421.4583	421.4583	22.9749	0.0001
Residual	19	348.5417	18.3443		
Total	20	770			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3.2083	1.8751	1.711	0.1034	-0.7163	7.133
X Variable 1	2.4792	0.5172	4.7932	0.0001	1.3966	3.5617

T3 vs. SCR

Regression Statistics	
Multiple R	0.7137
R Square	0.5094
Adjusted R Square	0.4836
Standard Error	4.4590
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	392.2333	392.2333	19.7276	0.0003
Residual	19	377.7667	19.8825		
Total	20	770			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3.4833	1.9521	1.7844	0.0903	-0.6025	7.5692
X Variable 1	2.3917	0.5385	4.4416	0.0003	1.2646	3.5187

T4 vs. CCR

Regression Statistics	
Multiple R	0.4700
R Square	0.2209
Adjusted R Square	0.1799
Standard Error	5.6190
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	170.1000	170.1000	5.3874	0.0316
Residual	19	599.9000	31.5737		
Total	20	770			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	6.05	2.46	2.4593	0.0237	0.9011	11.1989
X Variable 1	1.575	0.6786	2.3211	0.0316	0.1547	2.9953

T4 vs. SCR

Regression Statistics	
Multiple R	0.9183
R Square	0.8432
Adjusted R Square	0.8349
Standard Error	2.5208
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	649.2646	649.2646	102.1741	0.0000
Residual	19	120.7354	6.3545		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.3292	1.1036	1.2044	0.2432	-0.9807	3.639
X Variable 1	3.0771	0.3044	10.1081	0	2.4399	3.7142

T5 vs. CCR

Regression Statistics	
Multiple R	0.5572
R Square	0.3105
Adjusted R Square	0.2742
Standard Error	5.2844
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	238.9333	238.9333	8.5564	0.0087
Residual	19	530.5667	27.9246		
Total	20	769.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.1333	2.3135	2.2189	0.0389	0.2912	9.9755
X Variable 1	1.8667	0.6381	2.9251	0.0087	0.531	3.2023

T5 vs. SCR

Regression Statistics	
Multiple R	0.8554
R Square	0.7318
Adjusted R Square	0.7177
Standard Error	3.2960
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	563.0953	563.0953	51.8341	0.0000

Residual	19	206.4047	10.8634		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.9938	1.443	1.3817	0.1831	-1.0264	5.0139
X Variable 1	2.8656	0.398	7.1996	0	2.0325	3.6987

T6 vs. CCR

Regression Statistics	
Multiple R	0.8965
R Square	0.8037
Adjusted R Square	0.7934
Standard Error	2.8204
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	618.8583	618.8583	77.7966	0.0000
Residual	19	151.1417	7.9548		
Total	20	770			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.5583	1.2348	1.262	0.2222	-1.0261	4.1428
X Variable 1	3.0042	0.3406	8.8202	0	2.2913	3.717

T6 vs SCR

Regression Statistics	
Multiple R	0.3612
R Square	0.1305
Adjusted R Square	0.0847
Standard Error	5.9362
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	100.4646	100.4646	2.8510	0.1077
Residual	19	669.5354	35.2387		
Total	20	770			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	7.1958	2.5989	2.7688	0.0122	1.7564	12.6353
X Variable 1	1.2104	0.7169	1.6885	0.1077	-0.29	2.7108

T7 vs. CCR

Regression Statistics	
Multiple R	0.3918
R Square	0.1535
Adjusted R Square	0.1090
Standard Error	5.8552
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	118.1250	118.1250	3.4456	0.0790
Residual	19	651.3750	34.2829		
Total	20	769.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	6.875	2.5634	2.682	0.0148	1.5098	12.2402
X Variable 1	1.3125	0.7071	1.8562	0.079	-0.1674	2.7924

T7 vs. SCR

Regression Statistics	
Multiple R	0.9447
R Square	0.8924
Adjusted R Square	0.8868
Standard Error	2.0874
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	686.7146	686.7146	157.6072	0.0000
Residual	19	82.7854	4.3571		
Total	20	769.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.0542	0.9138	1.1535	0.263	-0.8585	2.9669
X Variable 1	3.1646	0.2521	12.5542	0	2.637	3.6922

T8 vs. CCR

Regression Statistics	
Multiple R	0.9618
R Square	0.9250
Adjusted R Square	0.9211
Standard Error	1.7432
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	712.2646	712.2646	234.3973	0.0000
Residual	19	57.7354	3.0387		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.8708	0.7632	1.1411	0.268	-0.7265	2.4682
X Variable 1	3.2229	0.2105	15.31	0	2.7823	3.6635

T8 vs. SCR

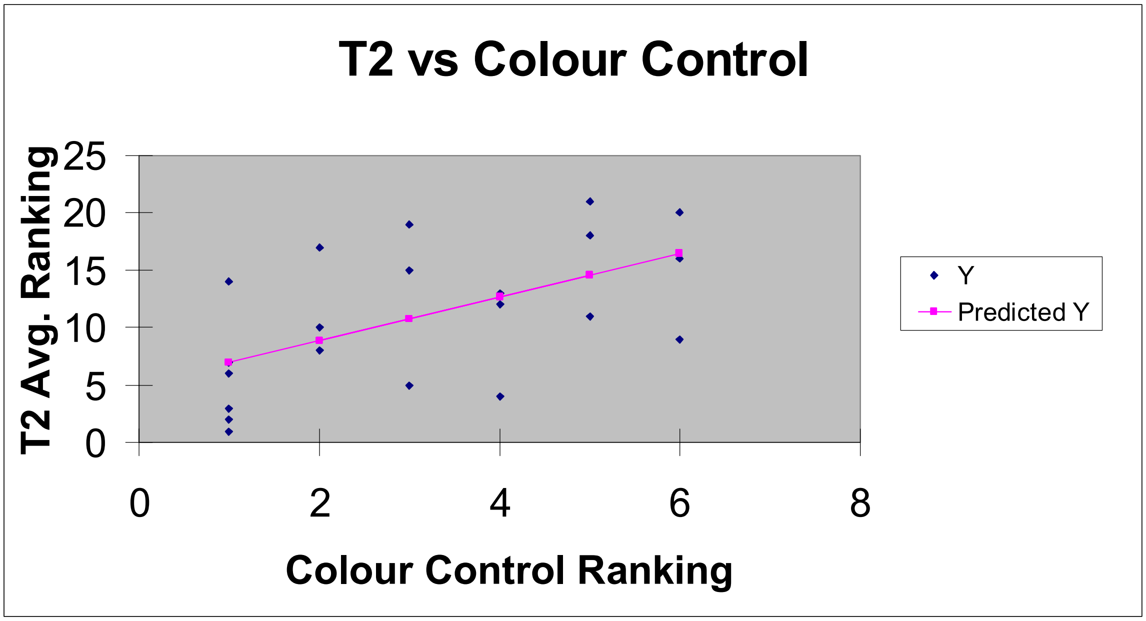
Regression Statistics	
Multiple R	0.2089
R Square	0.0436
Adjusted R Square	-0.0067
Standard Error	6.2256
Observations	21

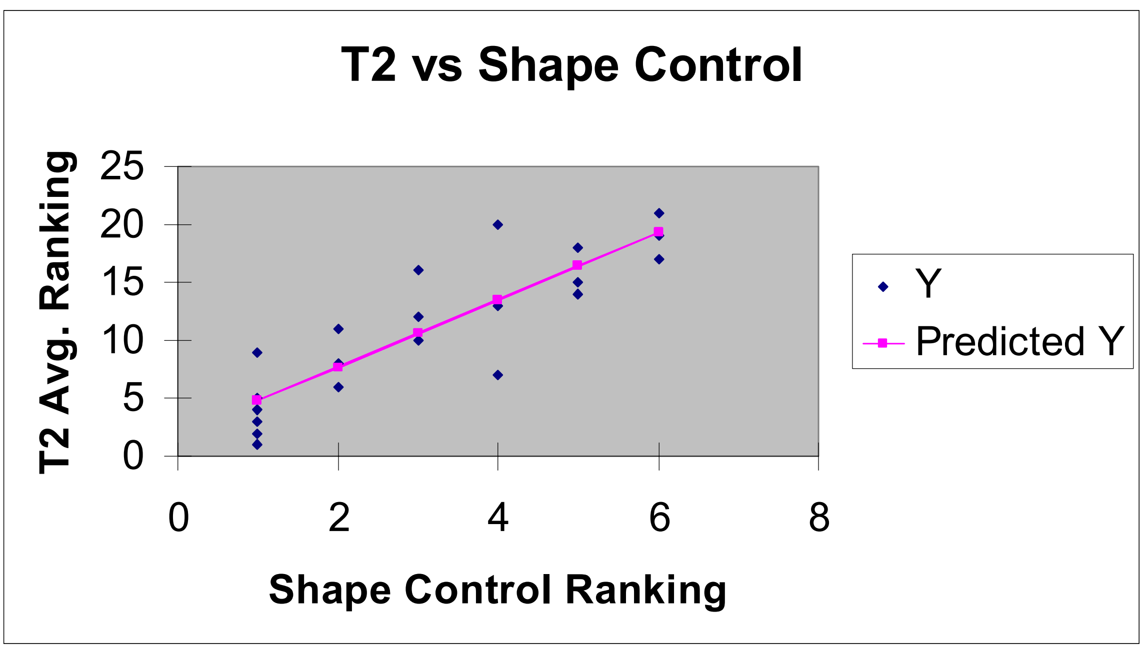
ANOVA

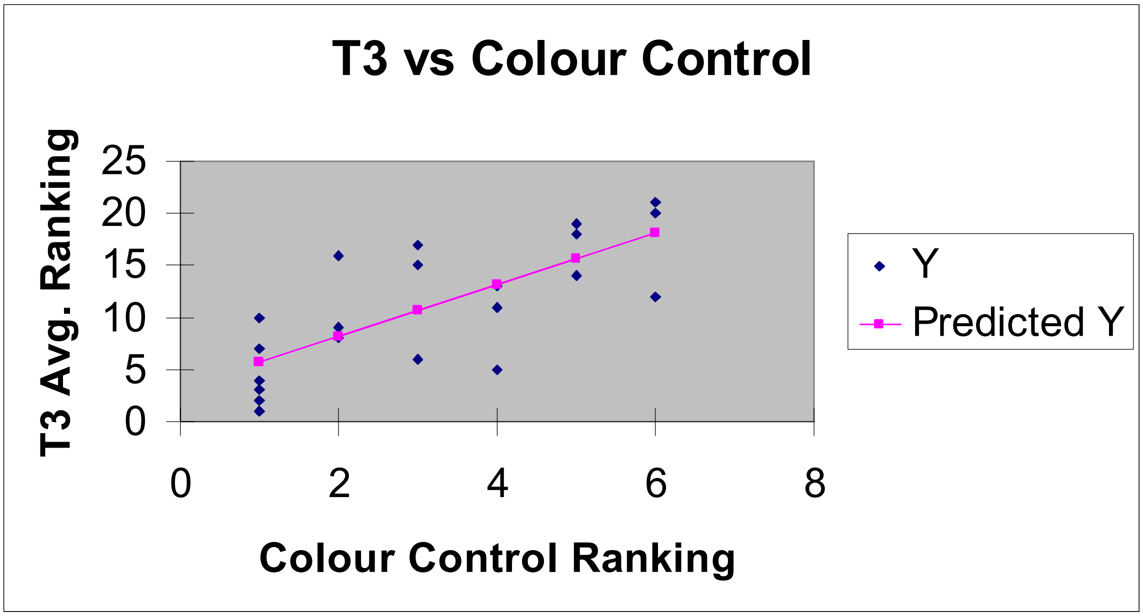
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	33.6000	33.6000	0.8669	0.3635
Residual	19	736.4000	38.7579		
Total	20	770			

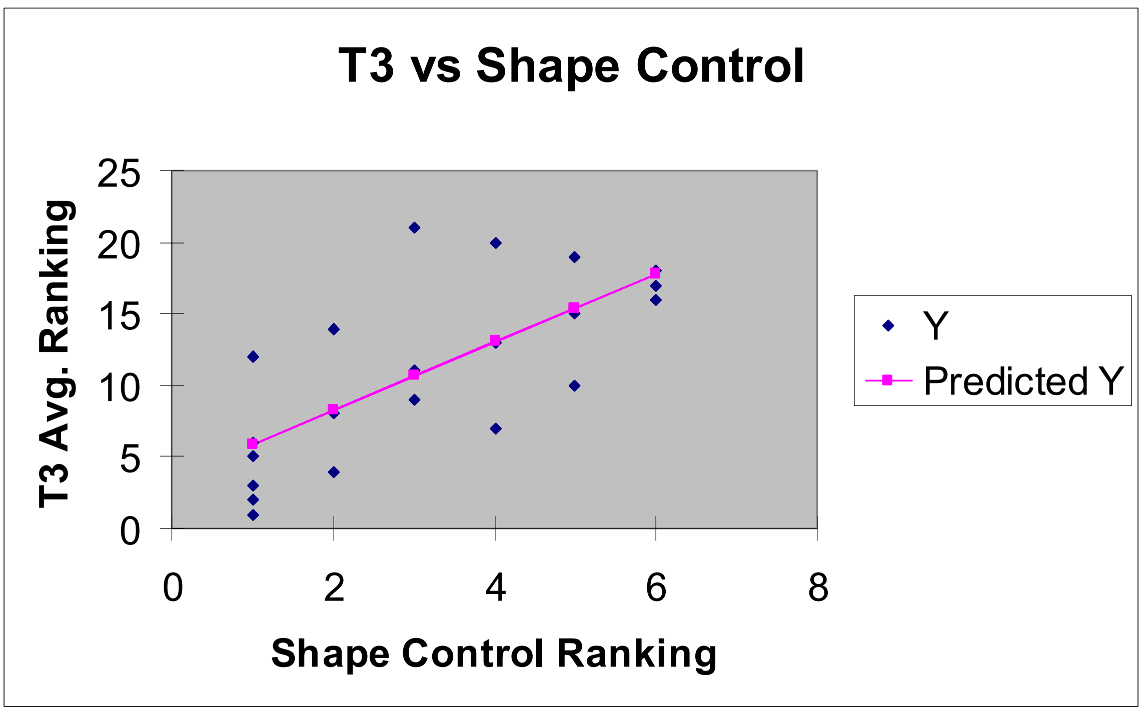
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.8	2.7255	3.2287	0.0044	3.0954	14.5046
X Variable 1	0.7	0.7518	0.9311	0.3635	-0.8736	2.2736

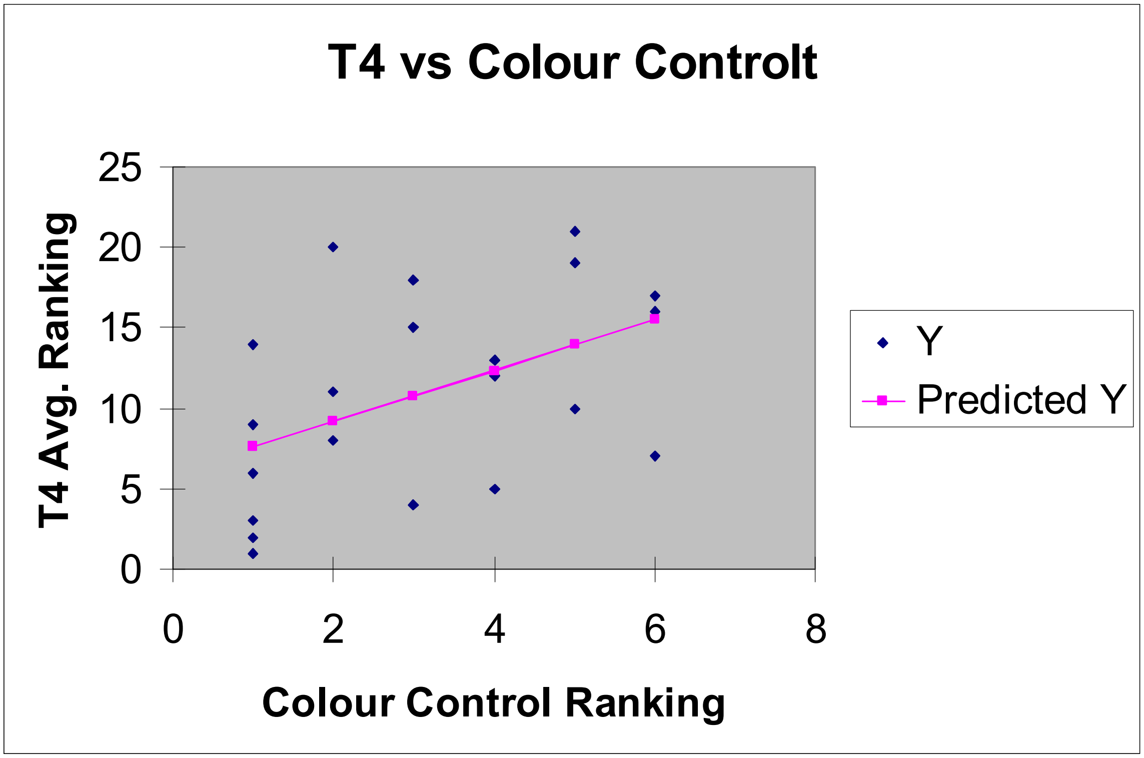
Regression Line Plots



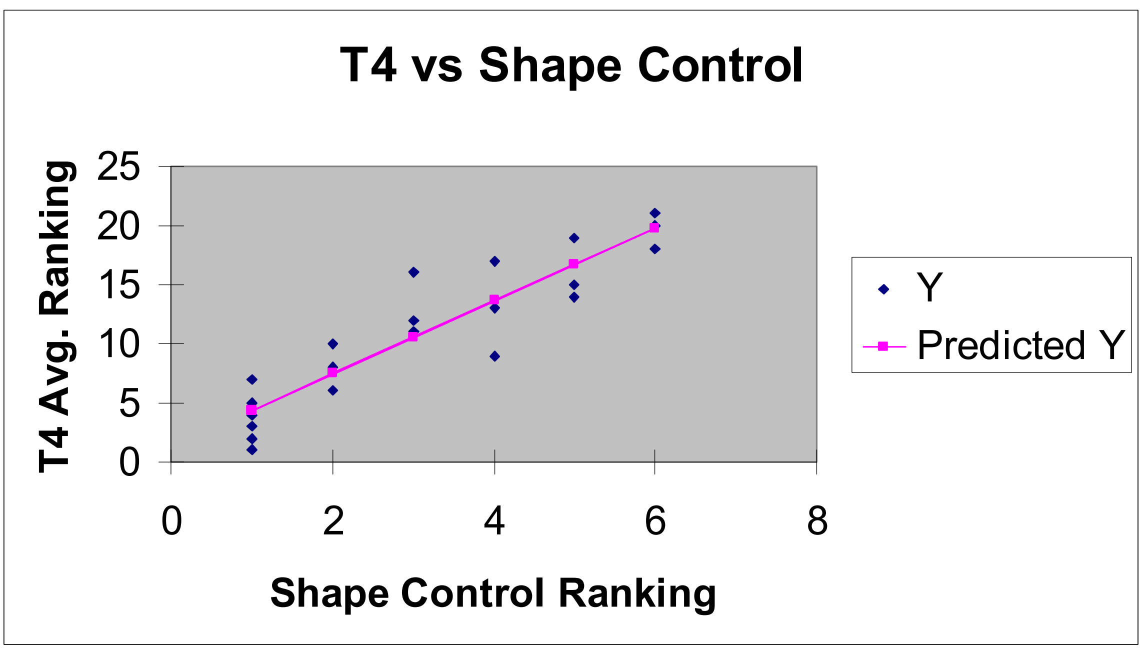



T3 vs Colour Control
T3 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

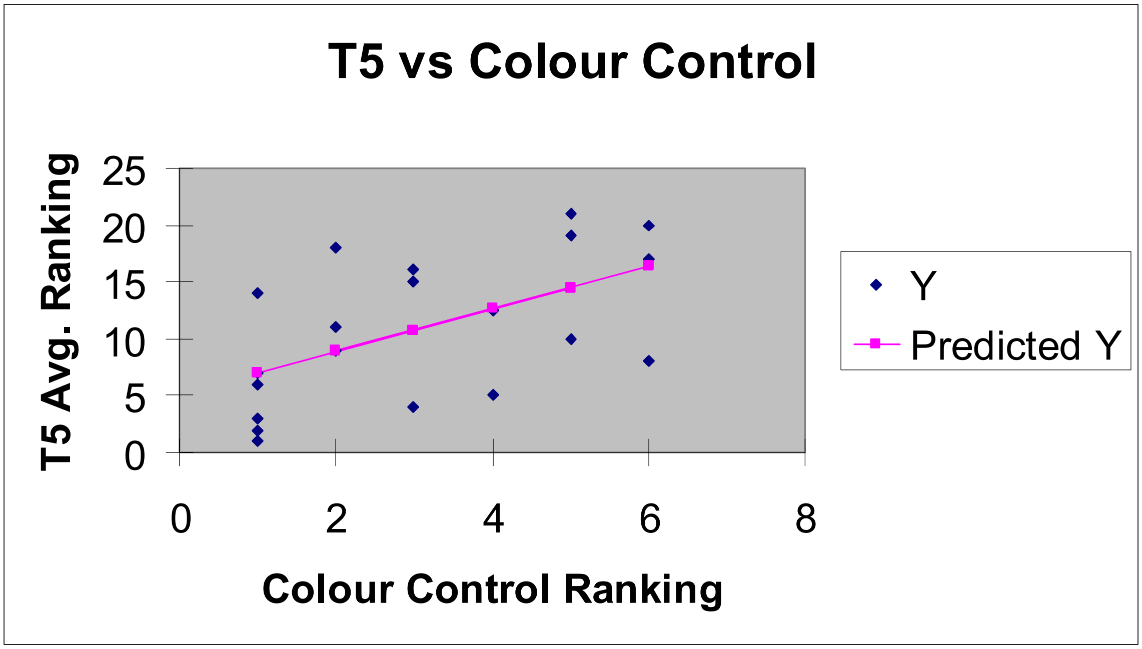

T3 vs Shape Control
T3 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

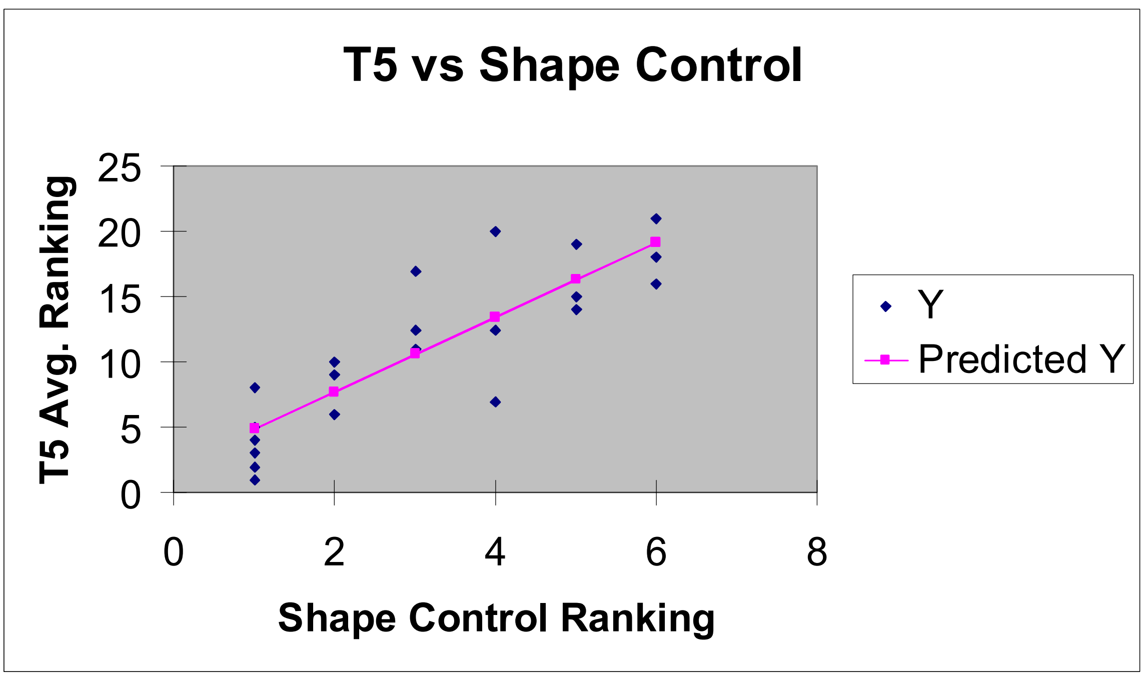


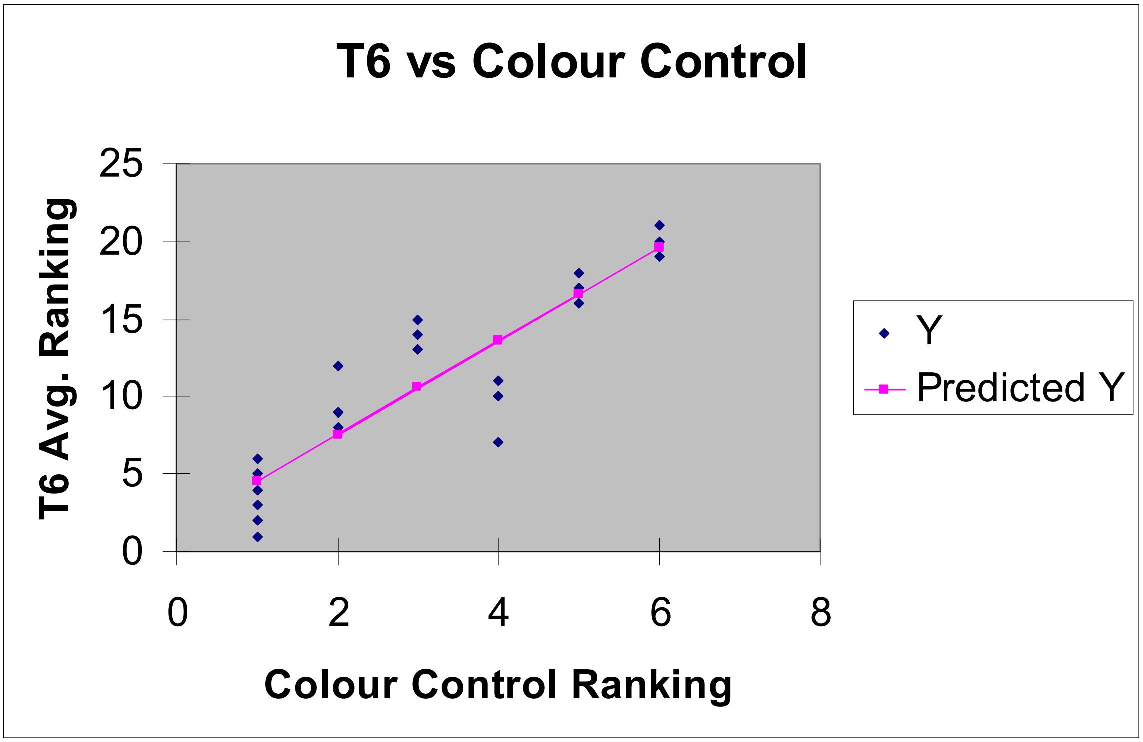
T4 vs Colour Controlt
T4 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

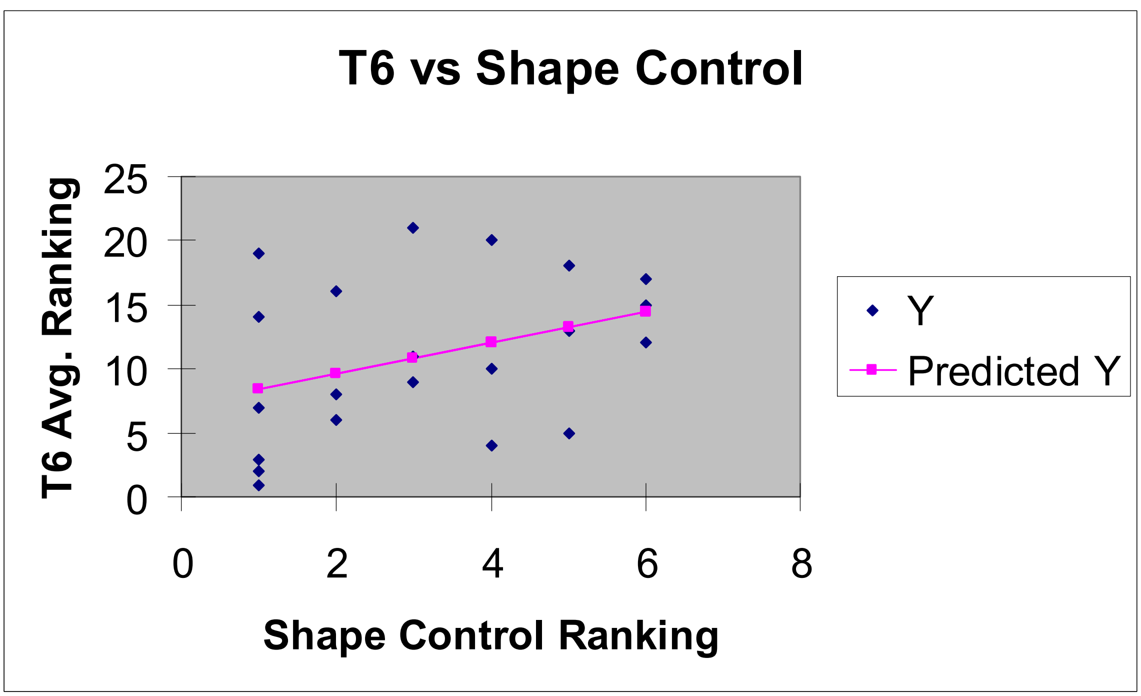


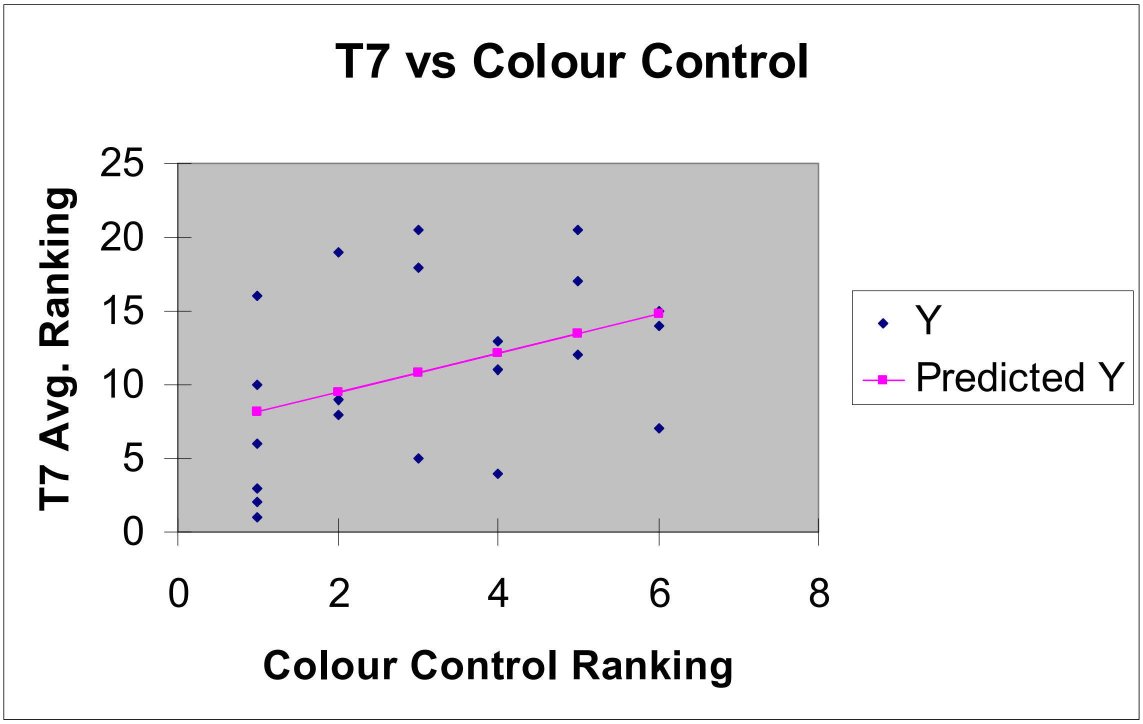
T4 vs Shape Control
T4 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

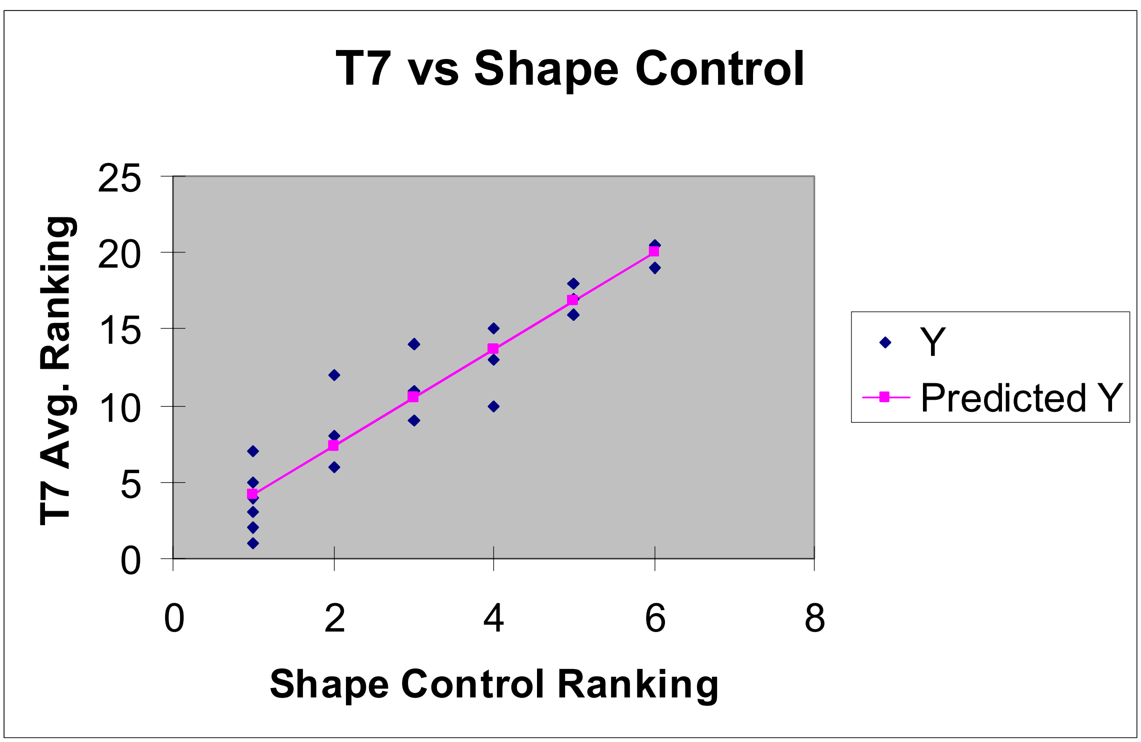

T5 vs Colour Control
T5 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

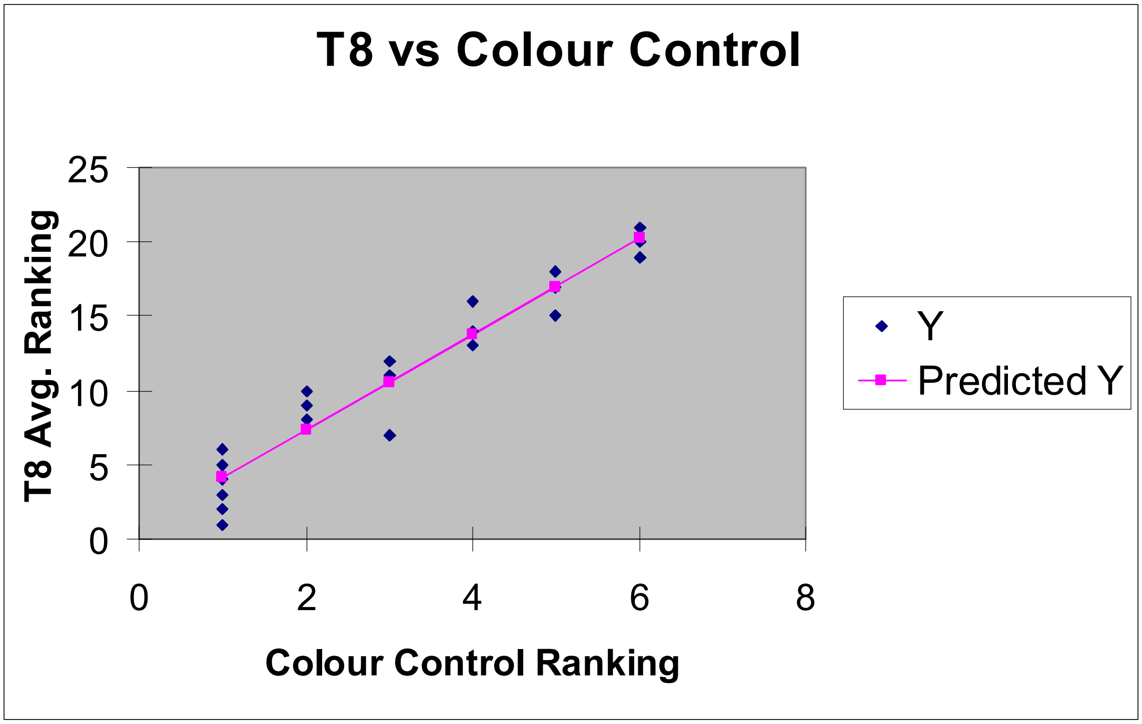

T5 vs Shape Control
T5 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

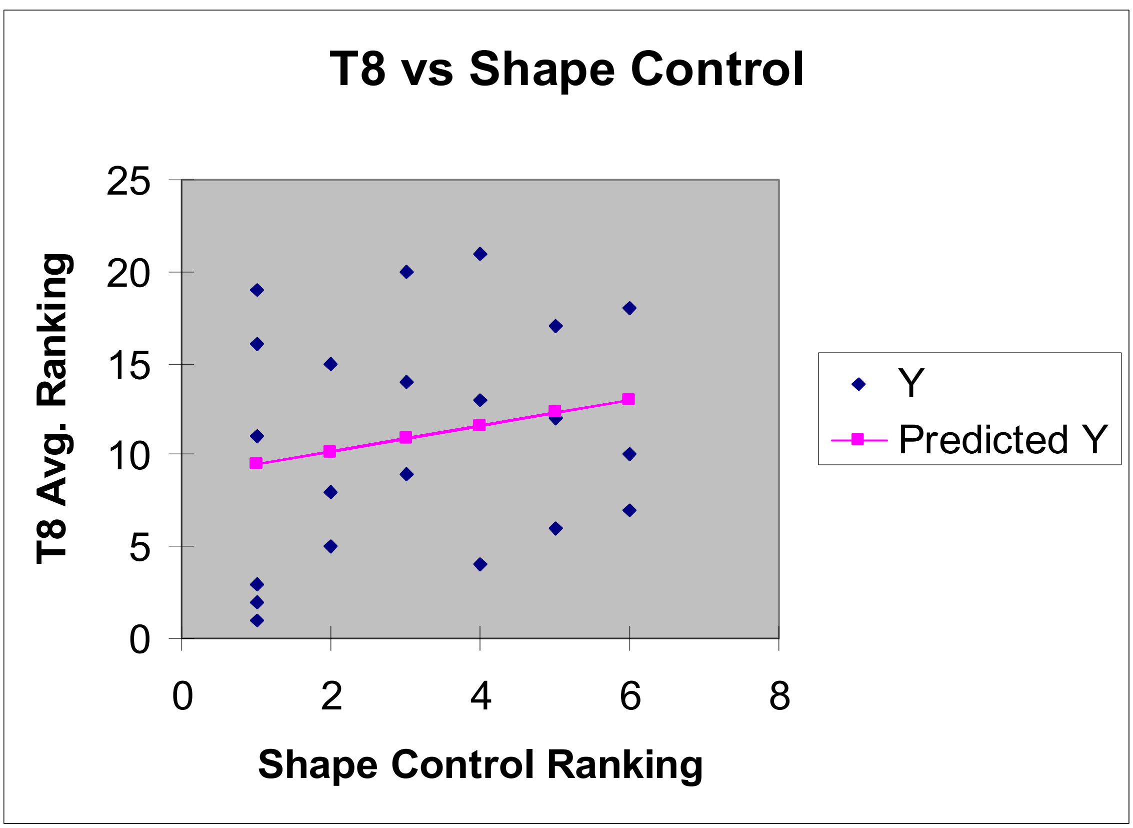

T6 vs Colour Control
T6 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y


T6 vs Shape Control
T6 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y


T7 vs Colour Control
T7 Avg. Ranking
Colour Control Ranking
0
5
10
15
20
25
0
2
4
6
8
Y
Predicted Y


T7 vs Shape Control
T7 Avg. Ranking
Shape Control Ranking
0
5
10
15
20
25
0
2
4
6
8
Y
Predicted Y


T8 vs Colour Control
T8 Avg. Ranking
Colour Control Ranking
Y
Predicted Y
0
5
10
15
20
25
0
2
4
6
8


T8 vs Shape Control
T8 Avg. Ranking
Shape Control Ranking
Y
Predicted Y
0
5
10
15
20
25
0
2
4
6
8

Appendix G: Average Rank Regression Analysis for Stimulus Set B

- Page 135: Averages of the ranks and the assigned rank
- Pages 136 to 141: Regression Results
- Pages 142 to 148: Graphs of the regression Lines

Averages of the Ranks for Experimental Stimulus B

Stim B	T2	T3	T4	T5	T6	T7	T9
1B	8.15	6.92	7.55	7.83	3.45	5.73	7.13
1C(X)_(1)	1.57	1.17	1.36	1.50	1.09	1.91	1.53
1C(X)_(2)	2.24	2.25	1.86	2.00	1.45	2.18	1.33
1C(X)_(3)	1.83	1.42	1.82	1.71	2.00	2.27	2.00
1C(X)_(4)	2.09	2.25	1.73	1.83	2.82	2.00	2.33
1F	12.61	11.92	13.18	13.63	4.18	11.18	12.60
2B	9.43	9.42	7.82	10.04	6.45	7.36	8.13
2D	6.80	7.58	5.68	6.92	7.00	5.09	5.73
2E	10.22	10.75	10.68	10.00	6.36	9.55	8.07
3B	11.15	12.00	8.82	11.96	9.09	7.45	7.87
3C	5.39	6.25	3.95	5.58	9.45	3.36	2.93
3E	12.43	13.58	11.64	13.25	8.36	9.45	8.13
4A	11.59	11.17	11.38	12.00	6.00	8.91	11.20
4D	6.74	6.17	5.95	6.50	5.73	4.91	6.53
4F	13.37	12.50	13.82	13.71	6.55	12.64	13.33
5A	11.52	11.25	11.50	12.29	6.27	9.18	10.73
5C	3.09	4.00	3.05	3.21	7.00	2.45	3.00
5E	9.85	9.92	10.14	9.96	6.10	8.55	7.80
6A	15.93	15.75	13.64	16.21	11.73	10.64	12.53
6D	11.74	11.45	8.64	12.00	10.18	6.36	7.60
6F	16.78	17.08	15.23	17.70	11.55	13.00	14.27

Ranks Assigned (Using SPSS) to Experimental Stimulus B

StimB	T2	T3	T4	T5	T6	T7	T9
1B	9	8	9	9	5	9	9
1C(X)_(1)	1	1	1	1	1	1	2
1C(X)_(2)	4	3.5	4	4	2	3	1
1C(X)_(3)	2	2	3	2	3	4	3
1C(X)_(4)	3	3.5	2	3	4	2	4
1F	18	16	18	18	6	19	19
2B	10	10	10	12	12	11	14.5
2D	8	9	7	8	14.5	8	7
2E	12	12	14	11	11	17	13
3B	13	17	12	13	17	12	12
3C	6	7	6	6	18	6	5
3E	17	19	17	17	16	16	14.5
4A	15	13	15	14.5	8	14	17
4D	7	6	8	7	7	7	8
4F	19	18	20	19	13	20	20
5A	14	14	16	16	10	15	16
5C	5	5	5	5	14.5	5	6
5E	11	11	13	10	9	13	11
6A	20	20	19	20	21	18	18
6D	16	15	11	14.5	19	10	10
6F	21	21	21	21	20	21	21

Results of the Regression Analysis

T2 vs. CCR

Regression Statistics	
Multiple R	0.6397
R Square	0.4093
Adjusted R Square	0.3782
Standard Error	4.8929
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	315.1313	315.1313	13.1631	0.0018
Residual	19	454.8688	23.9405		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.2625	2.1421	1.9899	0.0612	-0.221	8.746
X Variable 1	2.1438	0.5909	3.6281	0.0018	0.907	3.3805

T2 vs. SCR

Regression Statistics	
Multiple R	0.8965
R Square	0.8037
Adjusted R Square	0.7934
Standard Error	2.8204
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	618.8583	618.8583	77.7966	0.0000
Residual	19	151.1417	7.9548		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.5583	1.2348	1.262	0.2222	-1.0261	4.1428
X Variable 1	3.0042	0.3406	8.8202	0	2.2913	3.717

T3 vs. CCR

Regression Statistics	
Multiple R	0.6965
R Square	0.4852
Adjusted R Square	0.4581
Standard Error	4.5663
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	373.3333	373.3333	17.9049	0.0005
Residual	19	396.1667	20.8509		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3.6667	1.9991	1.8342	0.0823	-0.5175	7.8508
X Variable 1	2.3333	0.5514	4.2314	0.0005	1.1792	3.4875

T3 vs. SCR

Regression Statistics	
Multiple R	0.8446
R Square	0.7133
Adjusted R Square	0.6982
Standard Error	3.4077
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	548.8583	548.8583	47.2635	0.0000
Residual	19	220.6417	11.6127		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2.1083	1.4919	1.4132	0.1738	-1.0143	5.2309
X Variable 1	2.8292	0.4115	6.8748	0	1.9678	3.6905

T4 vs. CCR

Regression Statistics	
Multiple R	0.5440
R Square	0.2959
Adjusted R Square	0.2589
Standard Error	5.3417
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	227.8646	227.8646	7.9859	0.0108
Residual	19	542.1354	28.5334		
Total	20	770			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	5.2708	2.3386	2.2539	0.0362	0.3761	10.1655
X Variable 1	1.8229	0.6451	2.8259	0.0108	0.4728	3.1731

T4 vs. SCR

Regression Statistics	
Multiple R	0.9574
R Square	0.9167
Adjusted R Square	0.9123
Standard Error	1.8377
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	705.8333	705.8333	209.0000	0.0000
Residual	19	64.1667	3.3772		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.9167	0.8045	1.1394	0.2687	-0.7673	2.6006
X Variable 1	3.2083	0.2219	14.4568	0	2.7438	3.6728

T5 vs. CCR

Regression Statistics	
Multiple R	0.6138
R Square	0.3768
Adjusted R Square	0.3440
Standard Error	5.0240
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	289.9313	289.9313	11.4868	0.0031
Residual	19	479.5688	25.2405		
Total	20	769.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.5375	2.1995	2.063	0.0531	-0.0661	9.1411
X Variable 1	2.0563	0.6067	3.3892	0.0031	0.7864	3.3261

T5 vs. SCR

Regression Statistics	
Multiple R	0.9077
R Square	0.8239
Adjusted R Square	0.8146
Standard Error	2.6708
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	633.9703	633.9703	88.8767	0.0000
Residual	19	135.5297	7.1331		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.4438	1.1693	1.2347	0.232	-1.0036	3.8911
X Variable 1	3.0406	0.3225	9.4274	0	2.3656	3.7157

T6 vs. CCR

Regression Statistics	
Multiple R	0.8772
R Square	0.7695
Adjusted R Square	0.7573
Standard Error	3.0555
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	592.1161	592.1161	63.4229	0.0000
Residual	19	177.3839	9.3360		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.7646	1.3377	1.3191	0.2028	-1.0352	4.5644
X Variable 1	2.9385	0.369	7.9639	0	2.1662	3.7108

T6 vs. SCR

Regression Statistics	
Multiple R	0.3156
R Square	0.0996
Adjusted R Square	0.0522
Standard Error	6.0387
Observations	21

ANOVA

	df	SS	MS	F	Significance F
Regression	1	76.6536	76.6536	2.1021	0.1634
Residual	19	692.8464	36.4656		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	7.6771	2.6437	2.9039	0.0091	2.1437	13.2104
X Variable 1	1.0573	0.7292	1.4499	0.1634	-0.469	2.5836

T7 vs. CCR

Regression Statistics	
Multiple R	0.4961
R Square	0.2461
Adjusted R Square	0.2065
Standard Error	5.5273
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	189.5250	189.5250	6.2035	0.0222
Residual	19	580.4750	30.5513		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.775	2.4198	2.3865	0.0276	0.7102	10.8398
X Variable 1	1.6625	0.6675	2.4907	0.0222	0.2654	3.0596

T7 vs. SCR

Regression Statistics	
Multiple R	0.9574
R Square	0.9167
Adjusted R Square	0.9123
Standard Error	1.8377
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	705.8333	705.8333	209.0000	0.0000
Residual	19	64.1667	3.3772		
Total	20	770			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.9167	0.8045	1.1394	0.2687	-0.7673	2.6006
X Variable 1	3.2083	0.2219	14.4568	0	2.7438	3.6728

T9 vs. CCR

Regression Statistics	
Multiple R	0.4919
R Square	0.2420
Adjusted R Square	0.2021
Standard Error	5.5407
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	186.2146	186.2146	6.0658	0.0235

Residual	19	583.2854	30.6992		
Total	20	769.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	5.8208	2.4257	2.3997	0.0268	0.7438	10.8979
X Variable 1	1.6479	0.6691	2.4629	0.0235	0.2475	3.0484

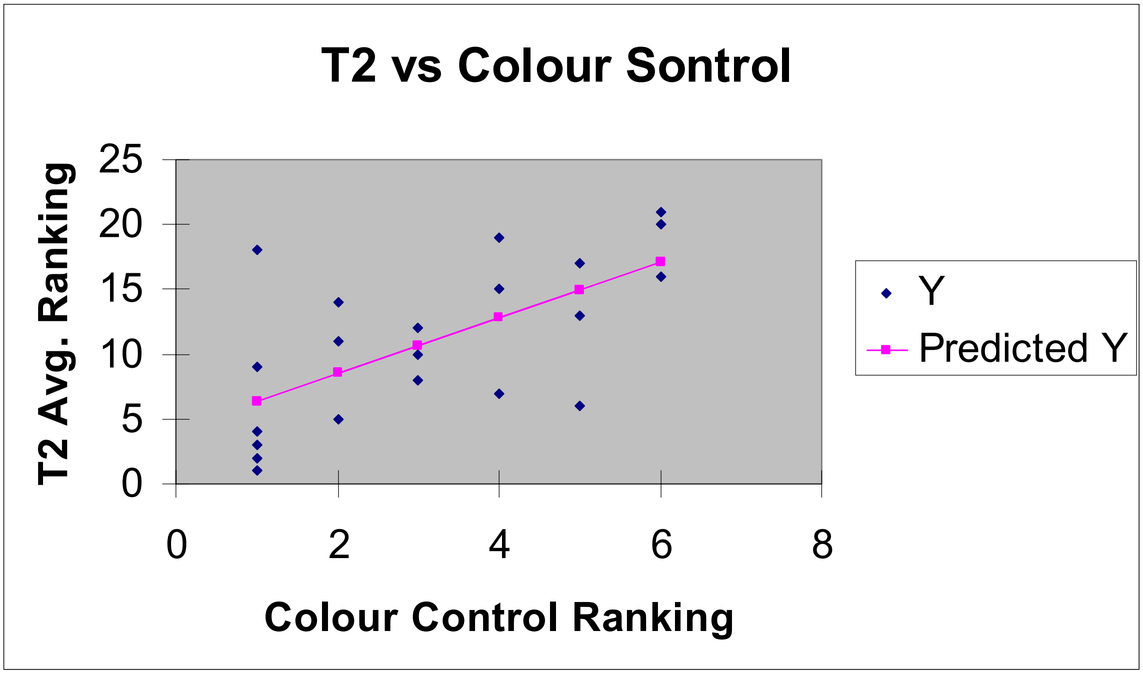
T9 vs. SCR

Regression Statistics	
Multiple R	0.9599
R Square	0.9214
Adjusted R Square	0.9173
Standard Error	1.7838
Observations	21

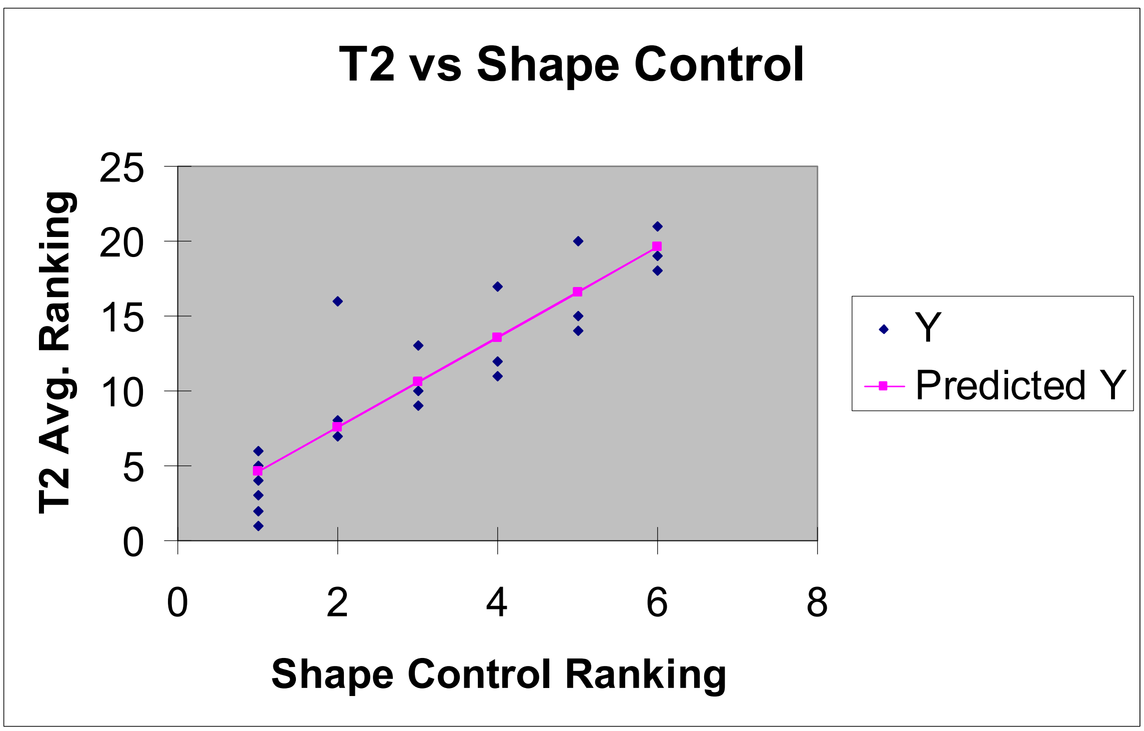
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	709.0453	709.0453	222.8423	0.0000
Residual	19	60.4547	3.1818		
Total	20	769.5			

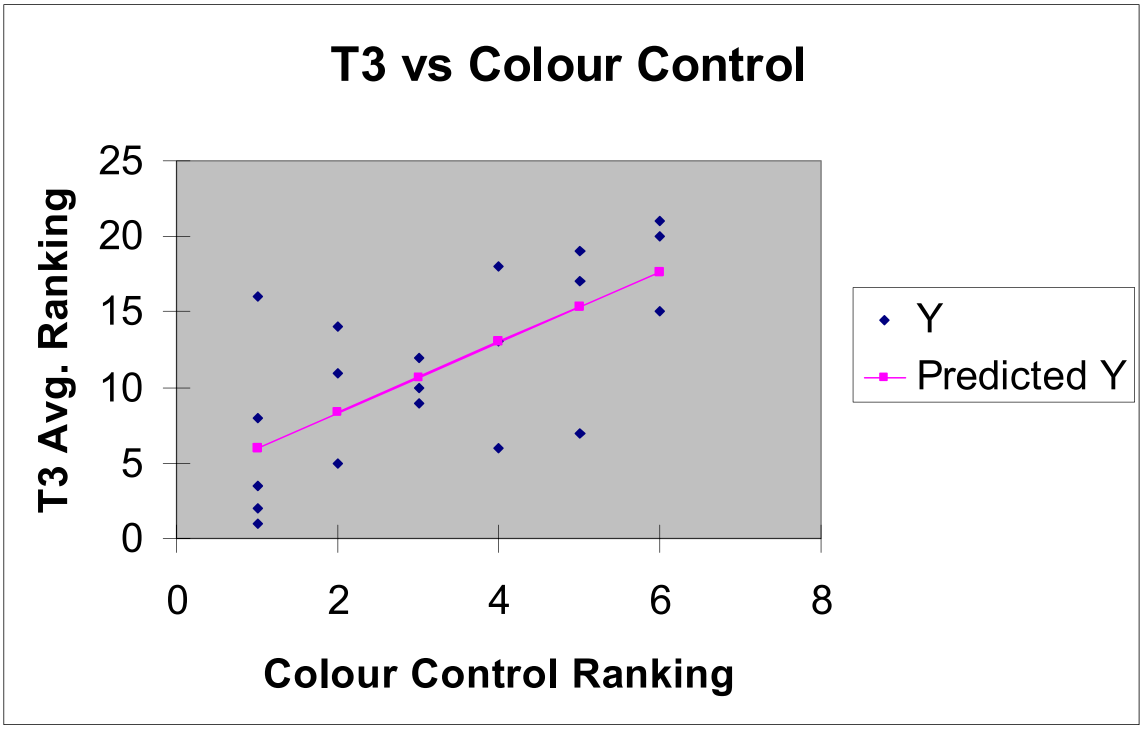
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.8938	0.7809	1.1445	0.2666	-0.7408	2.5283
X Variable 1	3.2156	0.2154	14.9279	0	2.7648	3.6665

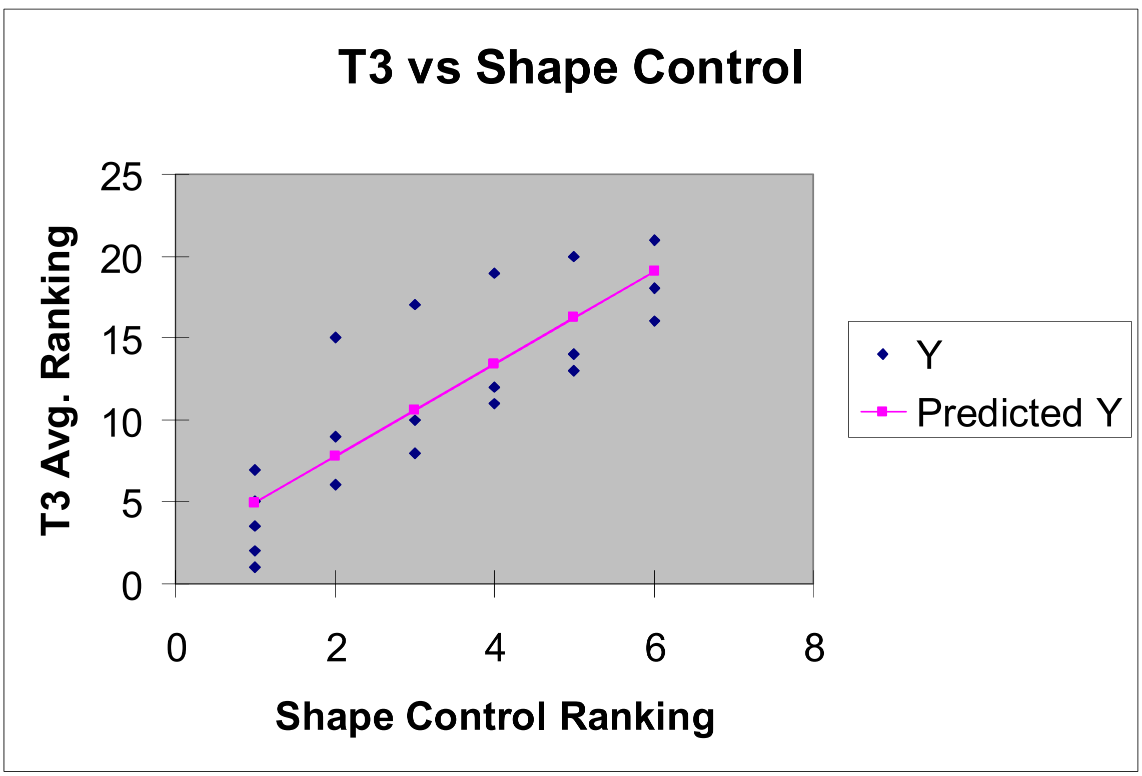


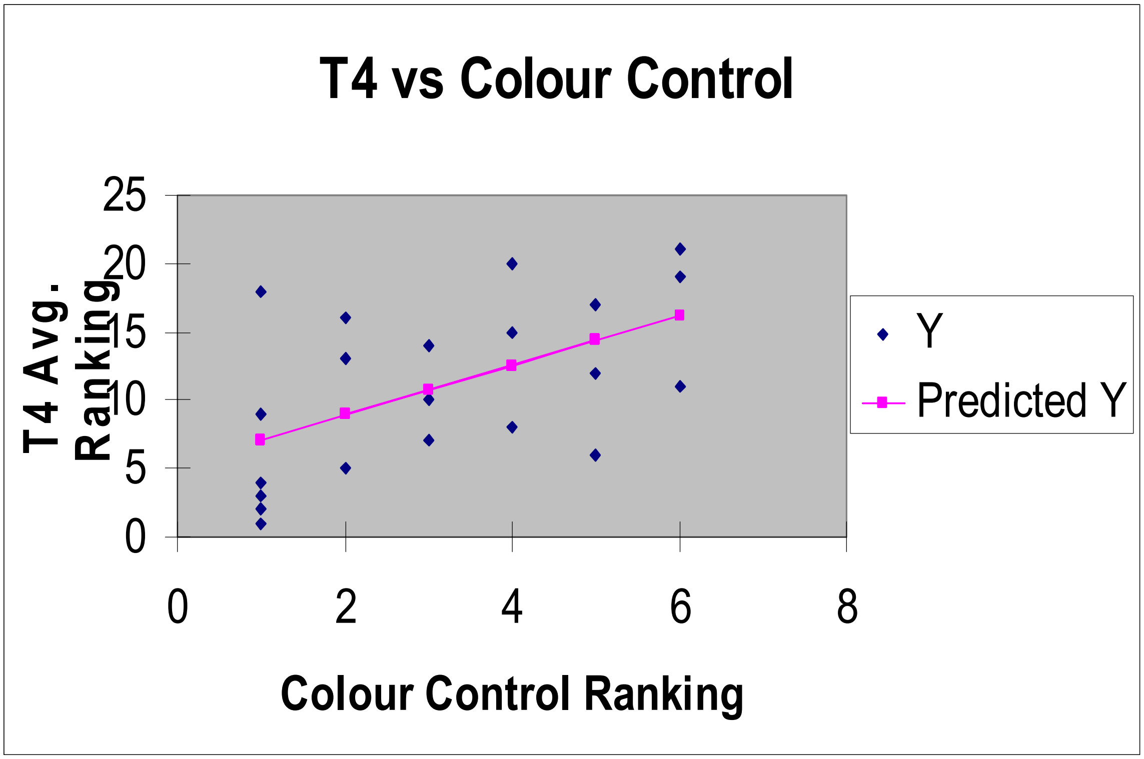
T2 vs Colour Sontrol
T2 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

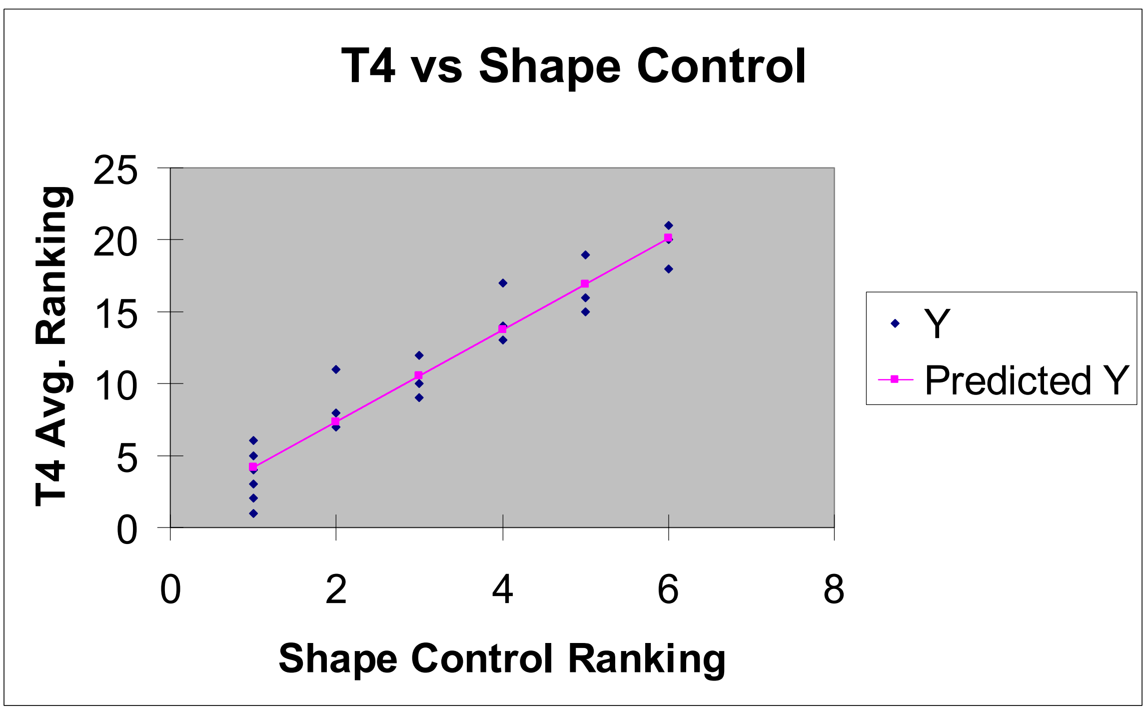


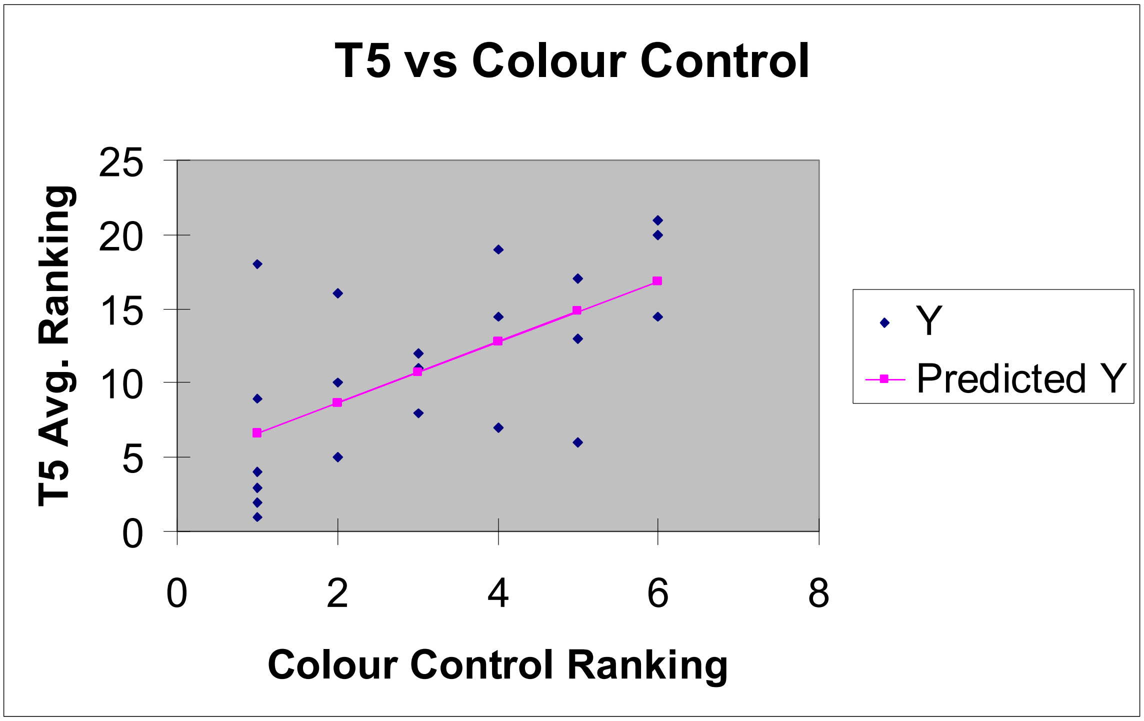
T2 vs Shape Control
T2 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

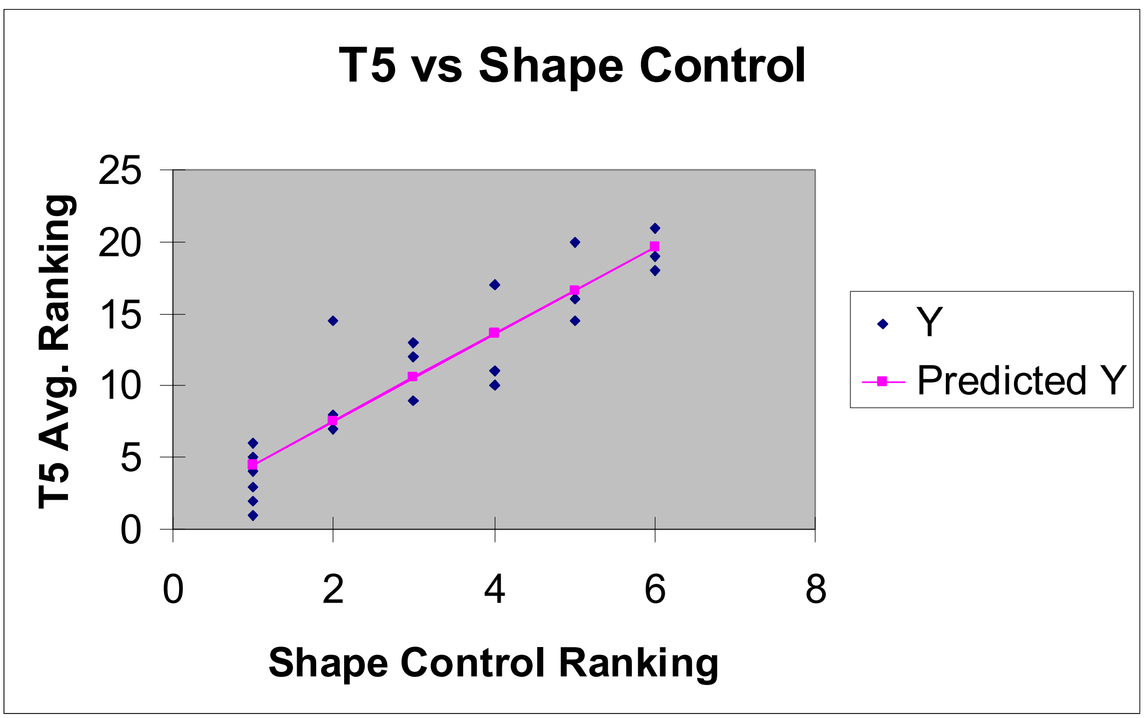

T3 vs Colour Control
T3 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

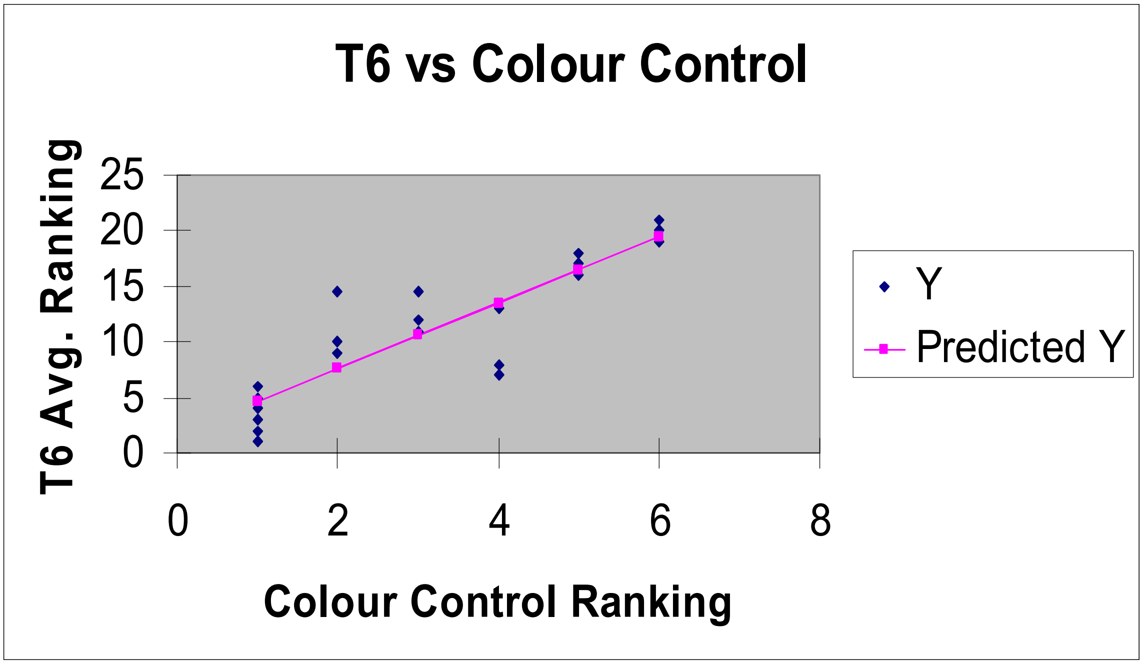

T3 vs Shape Control
T3 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

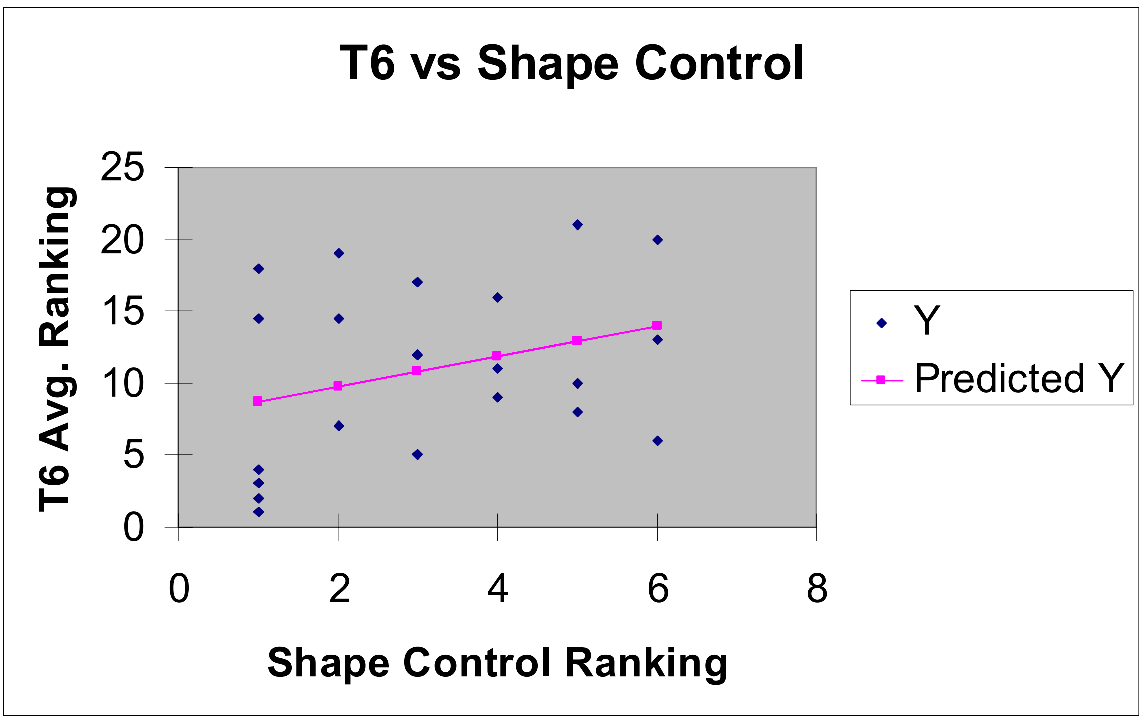

T4 vs Colour Control
T4 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

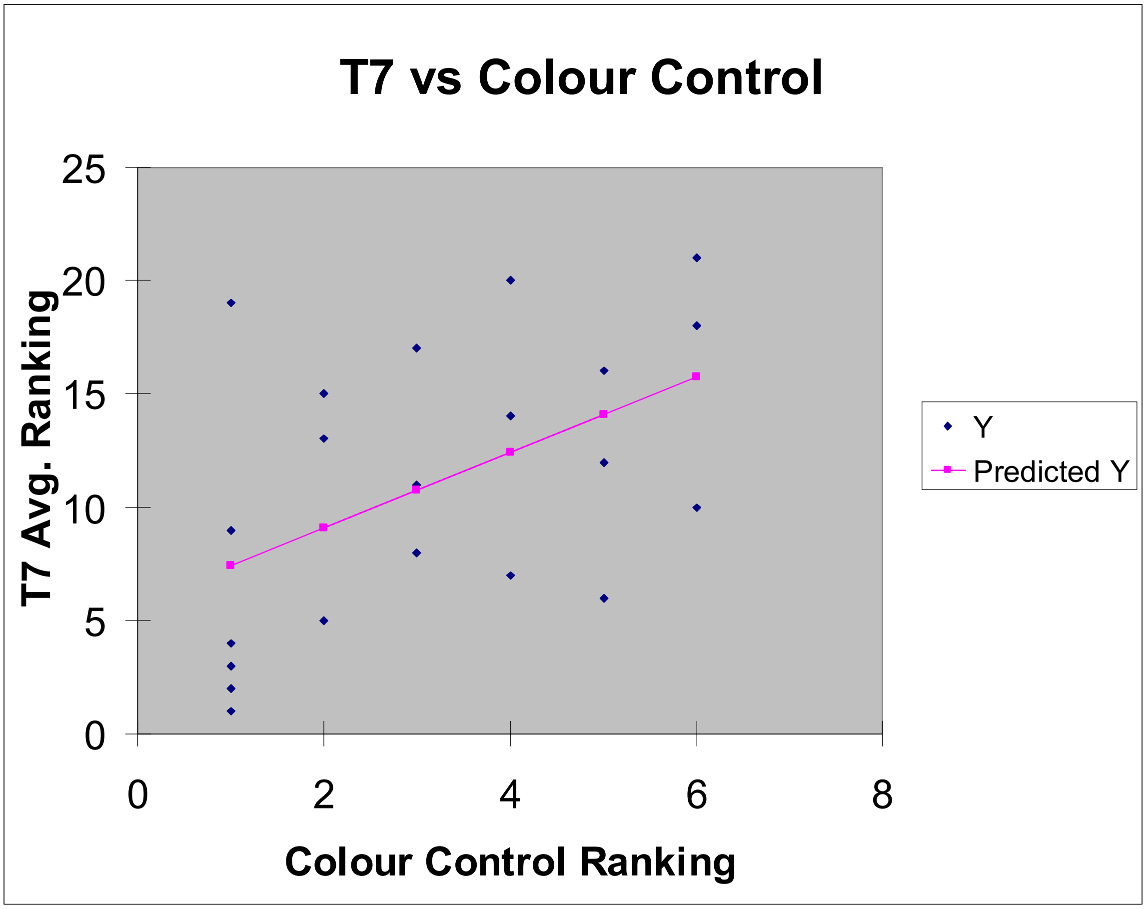

T4 vs Shape Control
T4 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

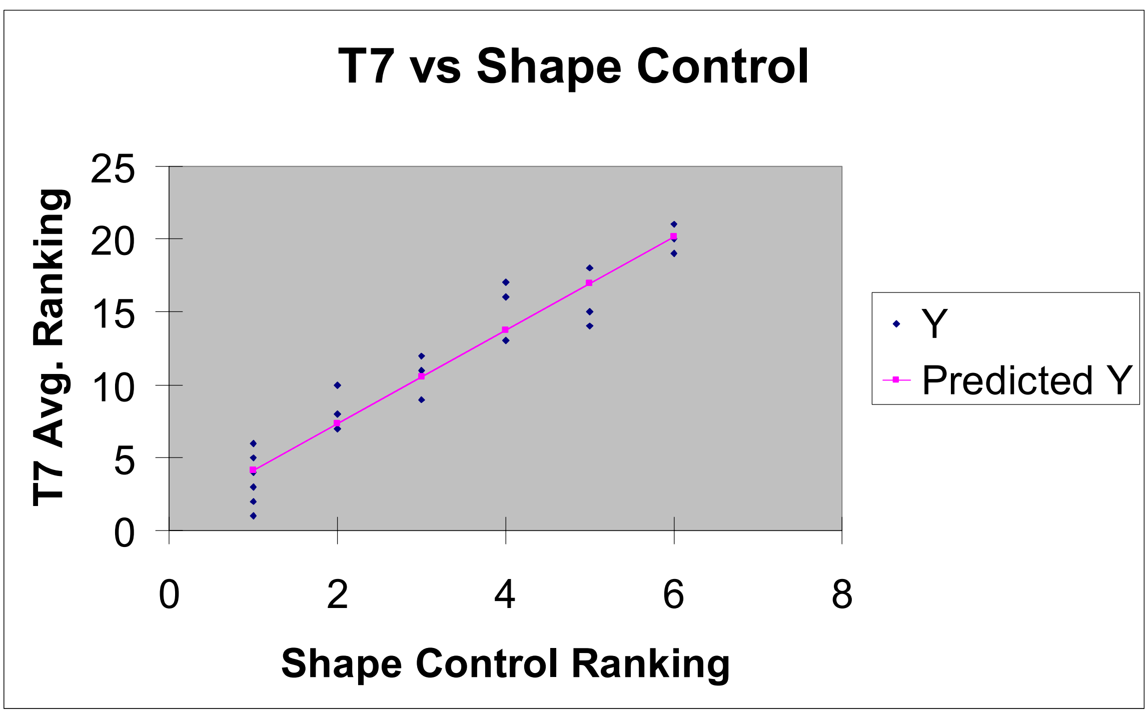

T5 vs Colour Control
T5 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y

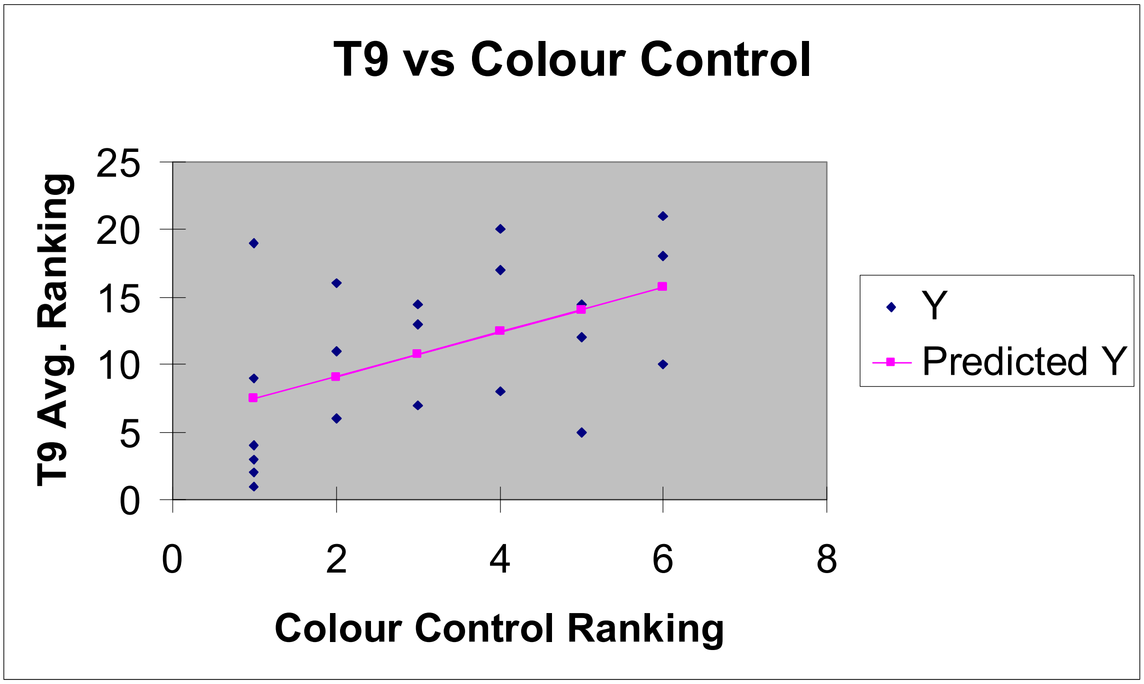

T5 vs Shape Control
T5 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

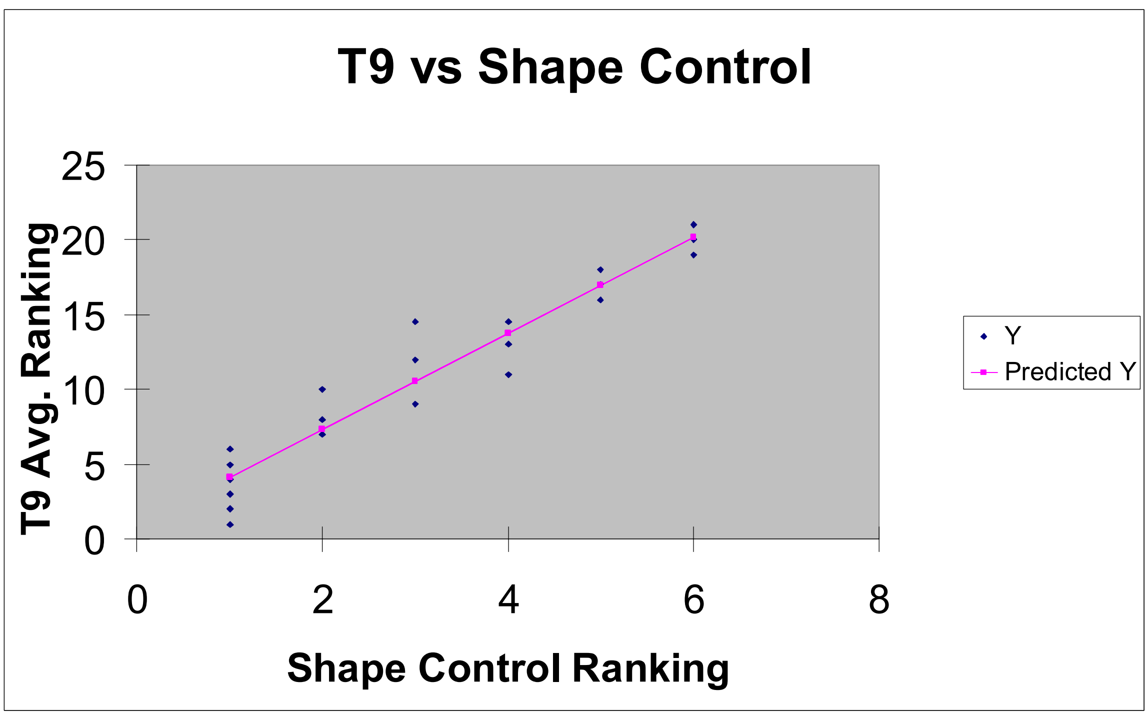

T6 vs Colour Control
T6 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y


T6 vs Shape Control
T6 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y


T7 vs Colour Control
T7 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y


T7 vs Shape Control
T7 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y


T9 vs Colour Control
T9 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Colour Control Ranking
Y
Predicted Y


T9 vs Shape Control
T9 Avg. Ranking
25
20
15
10
5
0
0
2
4
6
8
Shape Control Ranking
Y
Predicted Y

Appendix H: Testing of the Experimental Stimulus Sets for the Objects Location Bias

Testing the two Versions of Stimulus A for T2

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
10.24	14	12.5	14
1.97	2.5	1.5	1
1.88	1	3	3
1.97	2.5	2	2
5.97	6	5.75	6
7.91	7	7.375	7
11.52	15	15.625	19
4.71	5	5.5	5
14.15	20	14.75	17
13.97	19	13.25	15
9.21	10.5	9.5	11
15.71	21	15.75	20
9.76	12	11.625	13
3.85	4	4.875	4
10.15	13	11.5	12
9.21	10.5	8.875	9
8.12	8	7.75	8
13.76	17	14	16
13.47	16	15.125	18
8.18	9	9.375	10
13.81	18	16.625	21

Spearman Correlation: **0.956**
Sig $p < 0.001$

Testing the two Versions of Stimulus B for T2

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
8.2	9	7.83	8
1.525	1	1.83	4
2.35	4	1.50	2
1.925	2	1.17	1
2.15	3	1.67	3
12.6	17	12.67	18
9.4	10	9.67	11
6.425	7	9.33	10
10.25	12	10.00	12.5
11.1	13	11.50	14
5.15	6	7.00	6
12.8	18	10.00	12.5
11.45	15	12.50	17
6.675	8	7.17	7
13.3	19	13.83	19
11.425	14	12.17	16
3.025	5	3.50	5
10.125	11	8.00	9
15.8	20	16.83	21
11.75	16	11.67	15
16.975	21	15.50	20

Spearman Correlation: **0.952**
Sig $p < 0.001$

Testing the two Versions of Stimulus A for T3

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
10.81	10	6.43	7
2.06	1	1.43	3
2.63	3	1.29	2
2.38	2	1.14	1
6.44	4	3.43	4
7.88	8	5.86	5.5
13.13	16	10.14	11
7.75	7	5.86	5.5
14.69	19	11.00	14.5
15.06	20	15.00	18
11.44	13	11.71	16
14.44	18	16.29	19
10.88	11	10.86	12.5
6.56	5	7.86	10
11.56	14	10.86	12.5
8.94	9	7.43	8
7.38	6	7.57	9
12.38	15	13.86	17

Testing the two Versions of Stimulus B for T3

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
6.00	6	7.83	8.5
1.17	1	1.17	1
2.17	3	2.33	4
1.33	2	1.50	2
2.50	4	2.00	3
11.00	15	12.83	17.5
9.33	10	9.50	10
7.33	9	7.83	8.5
9.67	11.5	11.83	13
11.33	16.5	12.67	16
6.50	7	6.00	7
14.33	19	12.83	17.5
10.00	13	12.33	15
7.00	8	5.33	6
11.33	16.5	13.67	19
12.17	18	10.33	12
4.50	5	3.50	5
9.67	11.5	10.17	11

14.25	17	27.00	21
11.06	12	11.00	14.5
15.25	21	16.57	20

Spearman Correlation: **0.903**
Sig $p < 0.001$

15.83	20	15.67	20
10.60	14	12.17	14
16.67	21	17.50	21

Spearman Correlation: **0.954**
Sig $p < 0.001$

Testing the two Versions of Stimulus A for T4

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
12.25	14.5	10.20	14
2.00	2	1.40	1
2.38	3	1.53	2
1.38	1	1.67	3
6.25	6.5	5.53	7
7.88	8	7.86	9
12.25	14.5	10.93	15
4.13	4	2.93	4
14.63	20	12.33	17
13.50	16	13.57	19
9.38	10	8.87	12
15.50	21	15.00	21
9.88	11	8.93	13
4.25	5	3.73	5
11.50	13	8.79	11
10.75	12	8.29	10
6.25	6.5	6.79	8
14.00	18	13.80	20
14.38	19	11.27	16
8.38	9	5.20	6
13.75	17	12.40	18

Spearman Correlation: **0.956**
Sig $p < 0.001$

Testing the two Versions of Stimulus B for T4

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
9.33	11	6.88	9
1.00	1	1.50	2
2.17	2.5	1.75	4
2.17	2.5	1.69	3
2.50	4	1.44	1
15.67	19	12.25	18
8.17	10	7.69	10
6.17	7	5.50	7.5
11.67	14	10.31	14
9.50	12	8.56	11
4.33	6	3.81	6
12.83	16	11.19	17
12.50	15	10.93	15
7.17	8	5.50	7.5
16.00	20	13.00	19
13.00	17	10.94	16
3.33	5	2.94	5
11.00	13	9.81	13
14.17	18	13.44	20
7.83	9	8.94	12
17.33	21	14.44	21

Spearman Correlation: **0.977**
Sig $p < 0.001$

Testing the two Versions of Stimulus B for T5

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
7.56	9	8.38	9
1.38	1	1.75	2
2.31	4	1.38	1
1.63	2.5	1.88	3
1.63	2.5	2.25	4
12.63	17	15.63	18.5
10.13	12	9.88	10
6.69	8	7.38	7
10.00	11	10.00	11
11.81	16	12.25	13
5.75	6.5	5.25	6
13.44	19	12.88	14

Testing the two Versions of Stimulus A for T6

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
4.25	5	6	6
2.375	2	1.125	1
1.625	1	2.5	3
2.5	3	1.75	2
5.375	6	5.125	5
2.875	4	4	4
7.75	9.5	11.125	14
9.25	15	10.375	13
8.75	14	11.25	15
12.5	17.5	14.625	18
12.375	16	13.75	16.5
12.5	17.5	13.75	16.5

11.31	14	13.38	16	7.375	8	9.375	11
5.75	6.5	8.00	8	6.875	7	6.75	7
12.75	18	15.63	18.5	7.875	11.5	8.25	10
11.19	13	14.50	17	7.75	9.5	7.625	9
2.81	5	4.00	5	7.875	11.5	7.125	8
9.50	10	10.88	12	8.125	13	9.625	12
15.44	20	17.75	20	15.375	21	15.375	21
11.38	15	13.25	15	14.75	19.5	14.875	19
16.81	21	19.71	21	14.75	19.5	15.125	20

Spearman Correlation: **0.948**
Sig $p < 0.001$

Spearman Correlation: **0.961**
Sig $p < 0.001$

Testing the two Versions of Stimulus A for T7

Avg Rank V1	Rank V1	Avg Rank V2	Rank V2
12.56	16	7.63	14
1.67	2	3.50	2
1.33	1	3.38	1
1.89	3	4.13	3
5.56	7	4.38	4
9.78	13	5.25	8
13.44	18	9.63	19
3.11	5	4.75	6
15.22	20	11.25	21
13.22	17	7.88	15
7.89	9	8.50	17
16.11	21	10.25	20
8.00	10	7.38	13
3.00	4	4.63	5
10.11	14	6.38	10
8.11	11	6.13	9
6.56	8	4.88	7
14.78	19	9.13	18
9.67	12	8.13	16
4.33	6	6.50	11
11.89	15	7.25	12

Spearman Correlation: **0.875**
Sig $p < 0.001$