Personality and analogy-based project estimation

Martin Shepperd, Carolyn Mair, Miriam Martincova and Mark Stephens

Abstract

The aim of this research is to investigate the relationship between personality and expert prediction behaviour when estimating software project effort using analogical reasoning. For some years we have been developing tools and techniques for estimation by analogy (EBA). However the variability of results from using these tools and techniques can be difficult to interpret. We have conducted a pilot study to integrate knowledge from cognitive psychology and computer science to investigate how to improve estimation when using analogy-based tools. We interviewed and assessed the personality of two experienced project managers to gain an understanding of their background and the problem solving strategies they currently employ. Following these interviews, the project managers were given a typical project effort estimation task. The project managers were asked to complete the task using our analogical reasoning tool and articulate their processes by means of a 'think aloud' protocol. We found significant differences in prediction approach that may be in part be explained by personality differences. One aspect, i.e. the strong need to acquire personal understanding may present obstacles to the successful use of some prediction tools.

Keywords*: personality, analogy, case-based reasoning, cognitive psychology, software projects, effort prediction.*

1. Introduction

Our research aims to better understand the cognitive processes and the impact of personality on estimation by analogy for software professionals. We are investigating how these psychological aspects impinge on project cost estimation specifically when the estimator is using software tools to facilitate his or her prediction. Estimation by analogy (EBA), as encapsulated by case-based reasoning (CBR) tools, is loosely based on human cognitive processes. However, human demonstrate a wide range of individual differences and cognitive styles which may have a bearing upon performance. In particular we are interested in personality.

Despite the economic importance of accurate project cost estimation there remain many difficulties associated with reliably estimating, particularly at an early stage of the project. Many techniques have been proposed and many validation studies conducted. Unfortunately no clear picture emerges. Although Jørgensen has championed research into the use of what is often called expert judgment (i.e. humans making predictions unaided by formal systems), the interplay between expert and EBA has not been the target of empirical study. Hence we believe this research to be novel. For a comprehensive review of cost prediction research see (Jørgensen and Shepperd, 2007).

The remainder of the paper is organized as follows. In the next section we review the state of play for EBA project prediction research. This is followed by a brief description of the cognitive psychology viewpoint on personality (or individual differences) and then a discussion of how this has been applied to software projects in general. In the third section we describe our empirical investigation of project managers from our partner organization. This involved semi-structured interviews, personality tests and a think-aloud protocol whilst solving a real prediction task using our CBR tool. Preliminary results are presented in Section Four, and the paper concludes with a discussion of our findings.

2. Related work

Over the years many different approaches have been proposed for the task of predicting software project costs (usually effort) at an early stage in the process. One approach that has received considerable attention and some success is EBA. The idea here is that costs are best estimated by reference to similar past projects for which total cost or effort is already known. Thus there are essentially three steps. First, to code the new or target problem in terms of known characteristics, for example size in function points. Second, to retrieve completed projects that are similar to the target project. Third, to use the retrieved costs, possibly with adaptation, to predict the new project cost.

Whilst EBA may be conceived as an informal problem solving strategy there has been a good deal of research formalising this style of prediction into an artificial intelligence technique known as case-based reasoning (CBR). This constitutes the underpinnings of our EBA tool, called ANGEL (Shepperd and Schofield, 1997). Subsequently there have been a number of independent studies that have sought to compare the accuracy of EBA predictions with those obtained by other means, most commonly the benchmark technique of linear regression analysis. In 2005 we identified 20 such studies. We found approximately equal evidence for and against analogy-based methods (Mair and Shepperd, 2005). This naturally poses the question, why should EBA prediction accuracy be so variable?

One avenue we are therefore exploring is the interplay between the expert and estimation technique as encapsulated by ANGEL. Surprisingly, this is a rather neglected topic despite studies such as (Myrtveit and Stensrud, 1999) who reported that the *combination* of expert and CBR tool outperformed either individually. In particular there is a rich vein of psychology work that considers the cognitive processes of expert problem-solving (of which prediction is an example of an ill-defined problem). For a more extensive review of this literature and how it might apply to project prediction see (Mair et al. 2009). However in this paper we focus upon the role of personality in particular.

Project managers play a vital role in the success of software project cost estimation. It is therefore important that psychometric data also be collected. However, research in software engineering has mainly focused on models, algorithms and improvement of tools while overlooking the importance of human factors such as the personality of software professionals. Personality is made up of the characteristic patterns of thoughts, feelings, and behaviours that make a person unique. It is typically measured in terms of type or trait.

Type theories of personality propose that types are qualitatively distinct categories. That is that people are either introverts or extraverts. However, types do not reflect durable personality patterns; they tend to be a product of a particular place, time, and culture (Carver & Scheier, 2004). On the other hand, personality traits are persistent and exhibited in a wide range of social and personal contexts. Furthermore, trait theorists view personality as the result of internal biologically based characteristics that influence our behaviour. In addition, according to trait theorists, introversion and extraversion are part of a continuous dimension, with many people falling in the middle (Carver & Scheier, 2004).

Myers-Briggs Type Indicator (MBTI) is a popular example of a personality measure. Empirical studies using the MBTI have tended to find that certain personalities are disproportionately represented in software engineering personnel. For example, Capretz et al. (2003) found that most software engineers are introverts. The most common type found was ISTJ (introverted, sensing, thinking, judging). In a systematic literature review (Beecham et al. 2008) engineers were found to be sociable yet introverted, needing stability yet desiring a variety of new tasks and challenges. Gorla and Lam (2004) identified the best personality attributes for individual roles within a software development team. According to their survey results, the optimal personality for a team leader was intuitive and feeling; the optimal personality for a programmer was extrovert, sensing and judging; whereas the most suited type for system analyst was thinking and sensing. Of course, project teams comprise more than one individual and personality heterogeneity within the team can lead to success of the project (Howard 2001). It is therefore important to build up a team where all personality categories are represented. The majority of studies in software engineering have used the MBTI (Myers & McCaulley, 1985) as a measure of personality which demands the individual adopt the personality type he or she would use in a specific situation (e.g. at work). However, we are more concerned with biological bases of personality which are durable and manifest in a range of situations. For this reason, we elected to use the Eysenck Personality Questionnaire EPQ-R Short Scale (Eysenck and Eysenck, 1991), and the Impulsiveness (IVE) questionnaire (Eysenck and Eysenck, 1975).

3. Our study

We believe there is a need for cognitive, qualitative empirical research of professionals making predictions. This is to complement extensive current research into algorithmic and statistical aspects of prediction. We focus on making predictions using analogy. In a systematic review of the relevant empirical studies we found an absence of published work on the interplay of personal differences (personality) and experts carrying out prediction tasks (Mair et al., 2009). So this is the motivation for the empirical study reported in the remainder of this paper.

Working with our collaborator we conducted interviews, personality profiles and then a think-aloud protocol with a prediction task supported by our CBR tool. The participants were highly experienced project managers employed by the collaborator for many years. To assess personality we used the Eysenck model of personality which is based upon three major traits: Extraversion, Neuroticism (emotionality) and Psychoticism (tough-mindedness). The Eysenck Personality Questionnaire (EPQ) is a development of various personality questionnaires over 40 years and is regarded as a highly reliable and valid measure of personality. Internal consistencies and test-retest reliabilities of all three factors are above 0.7, many above 0.8 (on a 0 to 1 scale) and thus are highly satisfactory. The validity of these scales is supported by much experimental evidence. In fact, it is the best supported of any generally available personality measure.

The organization we are collaborating with is a major, international software developer with clients around the globe. They have had an extensive software measurement programme in place since the early 1990s and have amassed a database of over 10,000 projects. This database includes information about duration, team size, methods and language, client details, project size (typically measured in function points (FPs) and lines of code (LOC)) and total effort. Unfortunately there are extensive problems of missing values and these problems are compounded by issues of trustworthiness. However, guided by a Metrics Specialist from our collaborator, we identified a small subset of 18 comparable, recent UK enhancement projects. These were then used as the basis of a relevant realistic prediction task for our expert participants.

Table 1: Summary Data of Project Size and Effort Information

Variable	Min	Median	Mean	Max
Unadjusted FP	90	459	658	1719
Adjusted FP Count	90	525	701	1822
Total Logical LOC	2676	25959	29940	64031
Duration	192	380	393	544
Effort	6174	14760	18200	50886

The projects range in Effort from 6174 to 50886 person hours, essentially an order of magnitude. There is a slight tendency for the mean to exceed the median implying a positive skew to the data, in other words a few atypically large values (Table 1). A total of 16 variables were selected including methodology, client, full project name, language, maximum staffing, start and end dates and the information from Table 1.

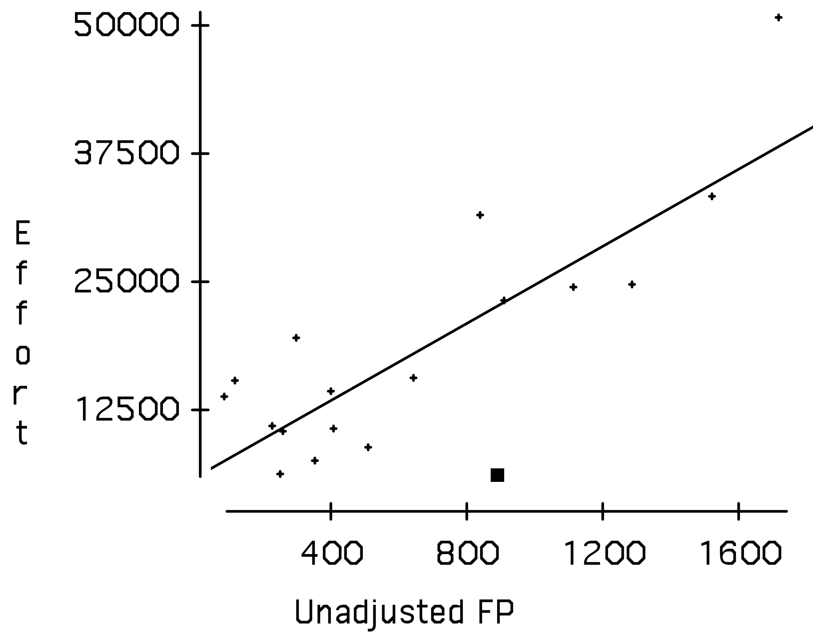
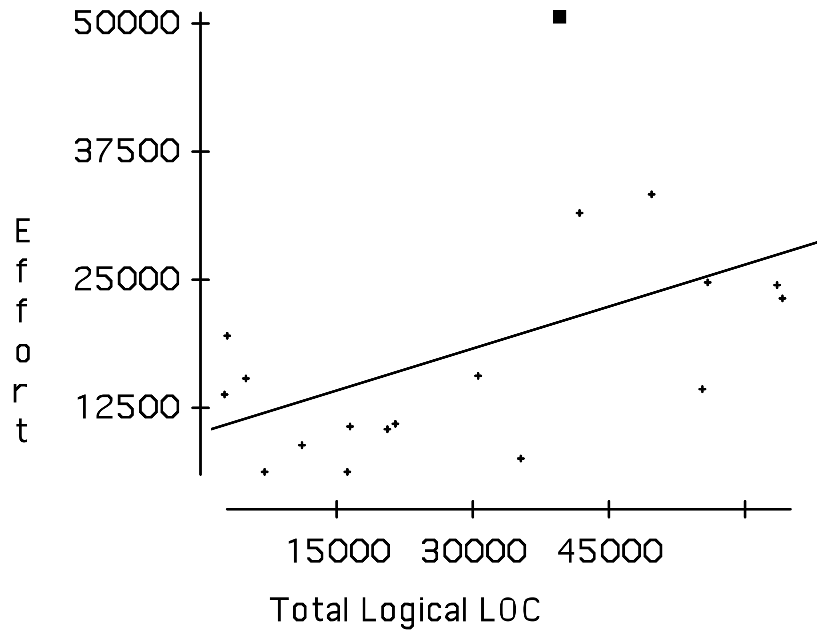


Figure 1a: Relationship between project size in FPs and Effort

Figure 1b: Relationship between project size in Logical LOC and Effort



In Figs 1a and 1b we use scatterplots to show the poor (certainly non-linear) relationship between size as measured by either FPs or LOC. The highlighted projects indicate extreme anomalous values. The regression lines are added for purposes of clarity only. This ill-defined relationship between obvious size measures and effort is a naturally occurring example of why cost prediction is difficult. There was also no clear relationship (not illustrated) between LOC and FPs even though all projects were developed in the same language. This indicates that the backfiring method of converting LOC to FPs is unlikely to be useful in this particular environment. More generally, it means that simple ratio-based approaches to estimation are unlikely to be effective (or indeed any form of linear regression modelling). This characteristic of the data caused the participants some difficulties for the prediction task, something we will shortly return to. Finally, we note in passing that even a naïve approach to this data set using our EBA tool ANGEL with a jackknife cross-validation could yield average prediction errors of less than 20% MMRE.

The estimation task was devised by taking one of the 18 completed projects and reducing the actual effort threefold (i.e. 15258 to 5086 person-hours). Each participant was then given the scenario that this estimate had been provided by another, unknown manager and their task was to perform a 'sanity check' using the data set of projects and the EBA tool ANGEL. Since they were unfamiliar with this tool, other than a prior 15-20 minute demonstration with a 'toy' example, one of the investigators would 'drive' the tool on their behalf. As they carried out the estimation task they were encouraged to use a think-aloud protocol and this was recorded and subsequently transcribed. In addition, the analogy tool ANGEL also generated a log file of usage.

4. Results

From the think-aloud transcription from the two project managers we produced process maps of how the two participants approached the estimation task (Figs. 1 and 2). These are intended to indicate the major steps in the prediction task which took about (60 minutes and 20 minutes respectively) for the two participants. The diagram is divided into three 'swimming lanes' which represent:

- references to the target problem (i.e. the estimate for which the sanity check is required)
- the prediction activities (e.g. computing a ratio)
- explicit requests for external information (e.g. speak to a colleague), which of course could not be satisfied due to the constrained nature of the investigation

The thought clouds denote self-reflective processes such as familiarity with a technique or degree of confidence in a result. In addition, time is conveyed by moving top-down through the diagram and lastly the small number boxes represent explicit references to other projects, in other words making use of analogies.

Even a cursory comparison of the two process maps reveals some clear differences. Most significantly, P1 was reluctant to make a decision concerning the prediction task without additional information. By contrast, P2 not only decided that the provided estimate was too low (5086 hours) but suggested an alternative value (12000 hours). Recall, the true value was 15286 hours. Interestingly P1 took approximately three times longer than P2. Given our interest is in analogical reasoning we also note that P2 made 16 explicit analogical references compared with 5 for P1. In general P1 did not wish to use the ANGEL tool to retrieve similar projects but preferred to establish and "validate" theories concerning relationships among the data. In particular, P1 sought to find linear relationships and stable ratios, for example in terms of delivery rates. As we discussed in the previous section such relationships are not to be found within this project data set and this caused P1 some difficulty. Interestingly towards the end of the task P1 speculated that the problem might lie within the quality of the data provided for the task. At one stage P2 also looked for some linear relationship between FPs and effort but when this was not supported by the data, P2 retreated from this viewpoint and investigated a new approach based on finding analogies using two (not one) dimension. This proved to be successful.

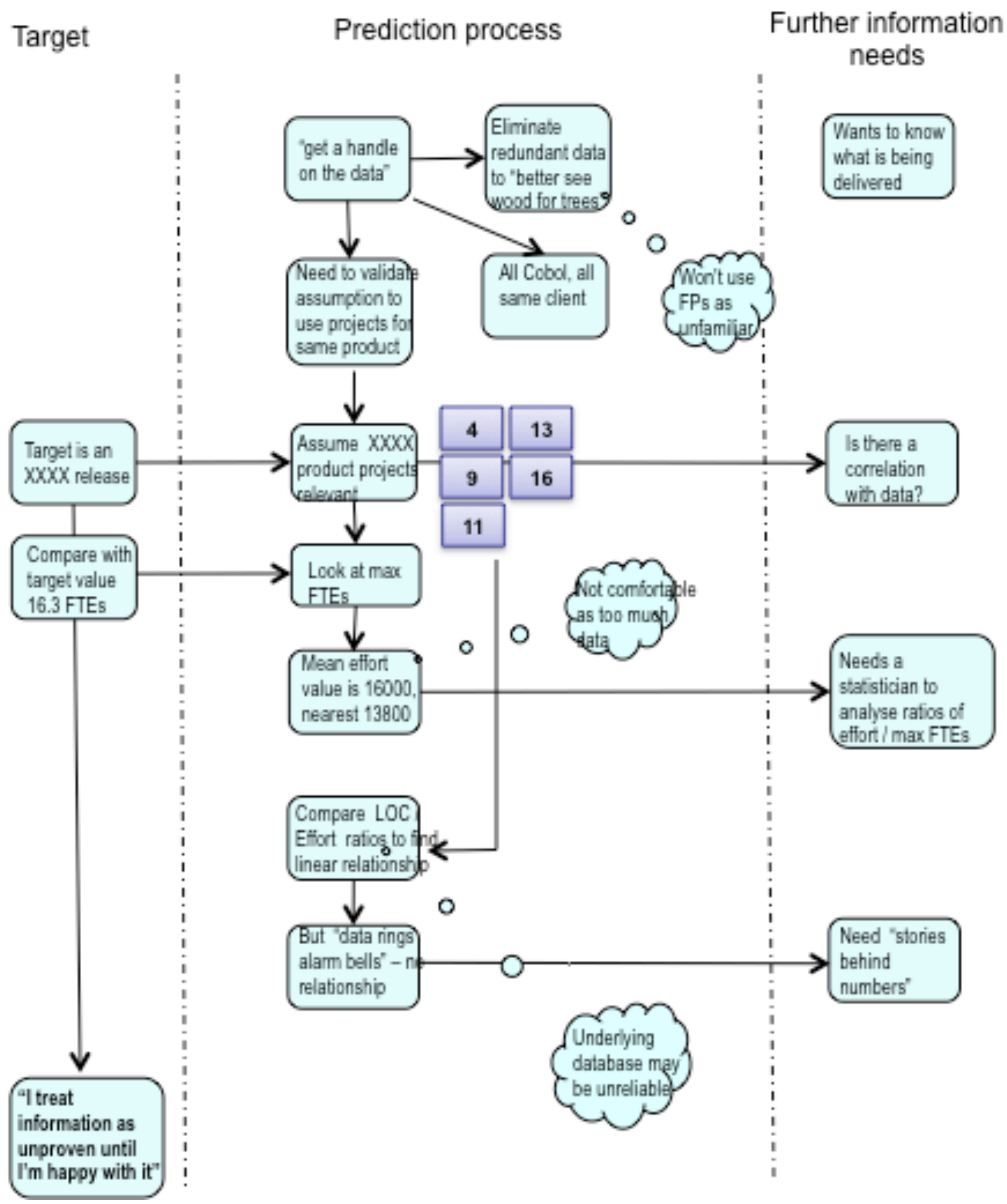


Figure 2: Process map of P1 carrying out the prediction task

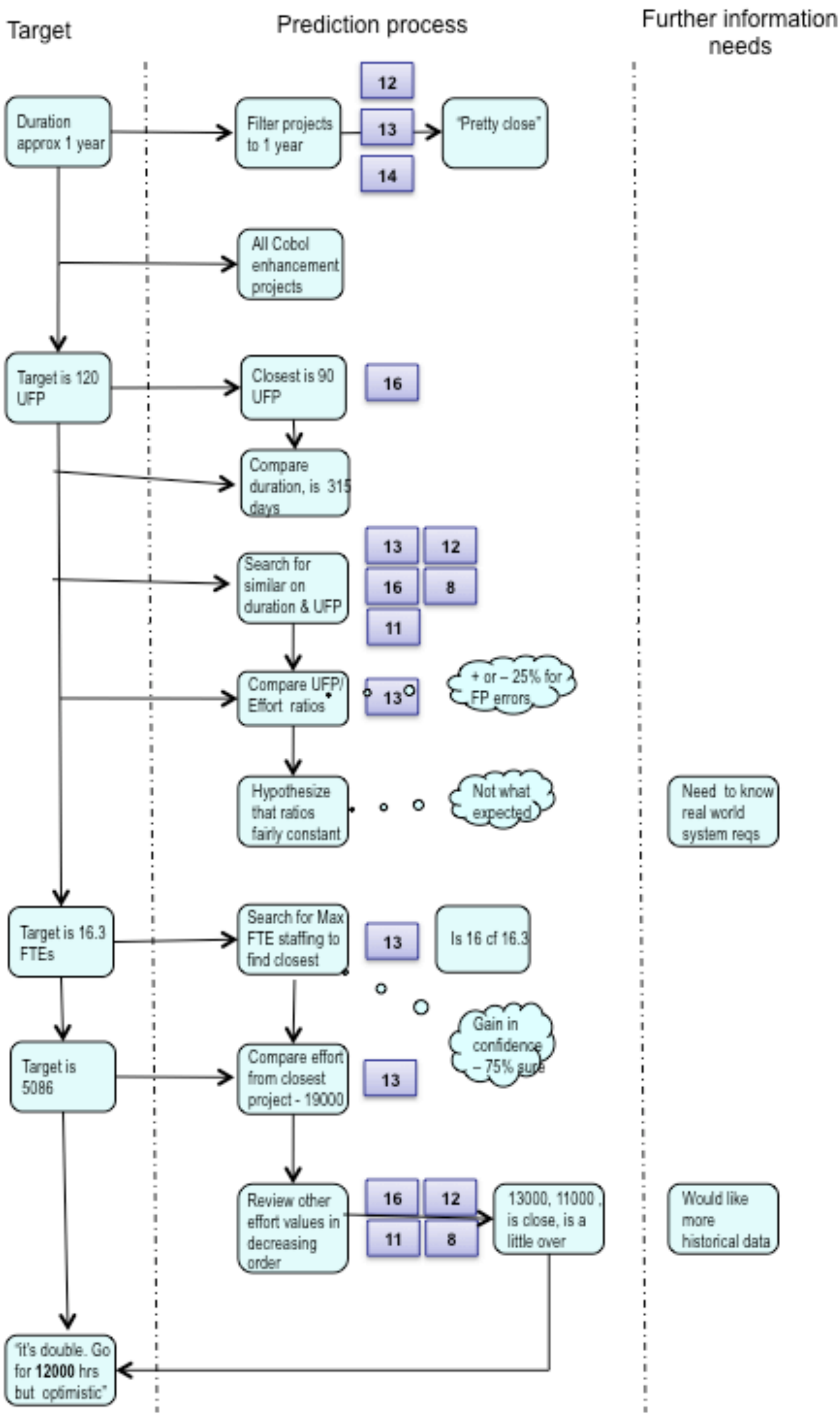


Figure 3: Process map of P2 carrying out the prediction task

Next we used the Adult EPQ-Revised Short Scale, comprising of 48 items, and Eysenck's Impulsiveness (IVE) questionnaire, which consists of 54 items, and is designed to measure three personality traits: Impulsiveness, Venturesomeness, and Empathy. The IVE supplements and complements the EPQ-R Short Scale measure. A dichotomous response format is used in both questionnaires with respondents ticking "Yes" or "No". Hence our participants' profiles are comprised of 6 dimensions: Extraversion (E, sociability), Psychoticism (P, tough-mindedness), Neuroticism (N, anxiety), Impulsiveness (I), Venturesomeness (V, tendency to be adventurous), and Empathy (Emp).

Table 2 Personality Profiles of the Study Participants

Participant	**P Psychoticism**	**E Extraversion**	**N Neuroticism**	**I Impulsive-ness**	**V Venturesome-ness**	**E Empathy**
P1	Below average	Below average	Below average	Below average	Above average	Average
P2	Below average	Above average	Below average	Below average	Average	Average

Across the two questionnaires, the participants scored below average on the Psychoticism, Neuroticism, and Impulsiveness traits. However, P1 scored below and P2 above average on Extraversion. On Venturesomeness, P1 scored above and P2 below average. P1 scored below average on empathy and P2's score was average (Table 2).

A high score on *Psychoticism*, or tough-mindedness, is associated with aggressiveness, hostility, anger, non-conformity and inconsideration. Our participants scored below average on this trait, demonstrating tender-mindedness, empathy, unselfishness, altruism, warmth, and placidness. Responses indicated that the participants enjoyed cooperating with others, were conforming, took notice of what other people thought, and cared about good manners.

A high score on *Extraversion* characterises sociability, outgoingness, and a need for external stimulation and action. In general, an extravert dislikes solitary pursuits, preferring excitement often achieved through taking chances and acting on impulse. In contrast the introvert tends to be quiet, retiring and studious. He or she can be reserved and distant except in intimate friendships, tends to plan ahead, and usually is not impulsive. Introverts prefer order and keep their feelings controlled. Hence, introverts are generally reliable, somewhat pessimistic, even tempered and tend to place great value on ethical standards. Our participants' scores differed on this trait. P1 was below average and P2 above. The effects of this difference on task performance are discussed below.

Scoring high on *Neuroticism* (emotionality or anxiety) characterizes high levels of depression and anxiety, low self-esteem, and feelings of guilt. Both participants scored below average on this trait which suggests emotional stability.

High scores on *Impulsiveness* are associated with an inclination to act on impulse. Again both participants scored lower than average, suggesting that they are not impulsive and think carefully before making a decision.

Venturesomeness characterizes being adventurous and exhibiting risk-taking behaviour. On this trait, P1 scored above average and P2 below average. We interpret this as P1 enjoying taking risks, welcoming new, exciting experiences and sensations. However, these scores contradict the participants' scores on the Extraversion trait. This might be explained by the fact that each trait comprises a range of dimensions. We intend to investigate this apparent incongruence later. The effects of this difference on task performance are further discussed below.

Empathy is associated with the capacity to share and understand another's state of mind or emotion. The participants' scores were average. This suggests they can easily identify with and understand another's situation, feelings, and motives. They can become engrossed in their friends' problems which may be reflected in their moods.

Thus, to summarise, P1 and P2 differ in terms of Extraversion and Venturesomeness out of the six traits (Table 2). Now we turn to the question of how these differences might explain at least in part differences in the participants' prediction behaviour. Other researchers such as (Huitt, 1992) have found that when solving problems, introverted individuals tend to take time to think and clarify their ideas before engaging, whilst extraverts tend want to talk through their ideas in order to clarify them. In addition, introverts are often concerned with their own understanding of important concepts and ideas, whilst extroverts seek feedback from others about the value of their ideas. In essence, P1 as an introvert operated in an "inner world of ideas" and attended to his / her internal consistency whilst P2 operated more in an outside world and attended to the "external reality".

Consequently P1 found the whole basis of the CBR tool, which is an example of a lazy learner, incongruent with the introvert's preferred style of problem solving. Finding analogies in high-order feature space is unintuitive and fits ill with simple linear modelling. Much of the raison d'être of estimation by analogy is that data are irregular and that by using past history one can avoid the need to build explanatory models. There is no support for inductive reasoning, i.e. to induce general theories or principles from example project data.

It is less easy to explain the impact of differences in the Venturesomeness trait. P1, perhaps surprisingly, given the below average score for Extraversion exhibited an above average score for Venturesomeness. However, it may be that the risk seeking behaviour of the participants did not play much of a role since the task was artificial without little likelihood of harm, particularly compared with, say, a multi-million pound or euro project.

5. Discussion

To recap, in this paper we have described a pilot study that we have conducted to empirically investigate the relationship between personality or individual differences and prediction behaviour of experts when using an estimation by analogy method. To do this we conducted semi-structured interviews to establish a context. For this pilot we worked with two experienced participants. This was followed by an estimation task using a think-aloud protocol. The task was based upon actual project data from the collaborating organization. This was complemented by a personality assessment using the Adult EPQ-Revised Short Scale and Eysenck Impulsiveness (IVE) questionnaires.

The two participants differed in personality in terms of two traits: Extraversion and Venturesomeness. We also observed significant differences in prediction processes. P1 took considerably longer than P2 and was reluctant to commit to an answer without further information whilst P2 was happy to provide a single point value. P1 made little use of the EBA tool ANGEL whilst P2 used it extensively and effectively.

The question therefore arises to what extent can we explain this in terms of Extraversion and Venturesomeness. Other studies have found significant differences in problem-solving approach between introverts and extroverts and we believe this may be salient to our study. However, the impact *in our study* of Venturesomeness or risk taking may be constrained by the artificial nature of the task since little rested upon the outcome.

So, what practical significance does this study have? One clear lesson is that designers of both estimation methods and tools need to take care to avoid a "one size fits all" mentality. It may well be that a partial explanation for the mixed accuracy results from EBA stems from individual differences of the estimators. Results such as from this study certainly should be fed back into the design of the next generation of EBA tools.

Finally there are some limitations to the results. In this pilot we only analyse results from two participants. We intend to extend this analysis to a larger number of project managers. In addition we note that, the Extraversion and Venturesomeness trait results seem to be contradictory in that they are negatively correlated. Further investigation is required along with the use of other personality measures.

Ackowledgements:
This work is funded by EPSRC grants EP/G007683/1 (Southampton Solent University) and EP/G008388/1 (Brunel University). We are also grateful to the collaborating organisation for supporting this research and allowing us access to their project managers and data.

References:

Beecham, S., Baddoo, N., Hall, T., Robinson, H. & Sharp, H. (2008). "Motivation in software engineering: A systematic literature review", *Information and Software Technology, 50 (9-10),* 860-878.

Capretz, L.F. (2003). "Personality types in software engineering". *International Journal of Human-Computer Studies, 58 (2),* pp207-214.

Carver, C. S., & Scheier, M. F. (2004). *Perspectives on personality* (5th ed.) Boston: Allyn and Bacon.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. Hodder & Stoughton, London.

Eysenck, H. J. & Eysenck, S. B. G. (1991) *Manual of the Eysenck Personality Scales (EPS Adult):* Hodder & Stoughton, London.

Eysenck, S.B.G., Pearson, P.R., Easting, G. & Allsopp, J.P. (1985). "Age norms for Impulsiveness, Venturesomeness and Empathy in Adults". *Personality & Individual Differences, 6(5),* pp613-619.

Gorla, N. & Lam, Y.W. (2004). "Who should work with whom? Building effective software project teams". *Communications of ACM, 47(6),* pp79-82.

Howard, A. (2001). "Software engineering project management". *Communications of the ACM, 44(5),* pp23-24.

Huitt, W. (1992), "Problem solving and decision making: Consideration of individual differences using the Myers-Briggs Type Indicator". *Journal of Psychological Type*, 24, pp33-44.

Jørgensen, M. and M. Shepperd (2007). "A Systematic Review of Software Development Cost Estimation Studies." *IEEE Transactions on Software Engineering* 33(1): 33-53.

Mair, C. and M. Shepperd (2005). The Consistency of Empirical Comparisons of Regression and Analogy-based Software Project Cost Prediction. 4th Intl. Symp. on Empirical Softw. Eng. (ISESE), Noosa Heads, Australia, IEEE Computer Society.

Mair, C., M. Martincova, and Shepperd, M. (2009). A Literature Review of Expert Problem Solving using Analogy. *13th International Conference on Evaluation & Assessment in Software Engineering (EASE 2009)*, Durham, UK, BCS.

Myers, I. & McCaulley, M. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.

Myrtveit, I. and E. Stensrud (1999). "A controlled experiment to assess the benefits of estimating with analogy and regression models." *IEEE Transactions on Software Engineering* 25(4) pp510-525.

Shepperd, M. J. and C. Schofield (1997). "Estimating software project effort using analogies." *IEEE Transactions on Software Engineering* 23(11), pp736-743.