An Empirical Study of Software Project Managers Using a Case-Based Reasoner

Carolyn Mair
Southampton Solent University
carolyn.mair@solent.ac.uk

Miriam Martincova
Southampton Solent University

Martin Shepperd
Brunel University
martin.shepperd@brunel.ac.uk

Abstract

BACKGROUND – whilst substantial effort has been invested in developing and evaluating knowledge-based techniques for project prediction, little is known about the interaction between them and expert users.

OBJECTIVE – the aim is to explore the interaction of cognitive processes and personality of software project managers undertaking tool-supported estimation tasks such as effort and cost prediction.

METHOD – we conducted personality profiling and observational studies using think-aloud protocols with five senior project managers using a case-based reasoning (CBR) tool to predict effort for real projects.

RESULTS – we found pronounced differences between the participants in terms of individual differences, cognitive behaviour and estimation outcomes, although there was a general tendency for over-optimism and over-confidence.

CONCLUSIONS – in order to improve task effectiveness in the workplace we need to understand the cognitive behaviour of software professionals in addition to conducting machine learning research.

1. Introduction

This paper aims to better understand the interaction between cognitive processes involved in problem solving and personality on software project professionals using case-based reasoning (CBR) tools to perform project effort estimation tasks. Project effort dominates project costs and calculating this can be characterised as infrequent, but high-value problem-solving. Therefore, professional estimators must assume significant responsibility for any estimate. Despite this, the role of the human in software project cost estimation is little understood. The dual-process theory of cognition [1] leads to a tendency to trust analytic justifications (explanations) over intuitive ones, yet to prefer intuitive judgements over analytic ones. One implication of this is that formal prediction systems can turn into "expert judgement in disguise" [2].

The rationale for trying to understand how to predict more accurately is illustrated by Flyvbjerg et al. [3] who describe project cost underestimation and overrun as a global problem which has not diminished in the last 50 years and argue that "no learning seems to take place" (p.16). In addition, to project cost underestimation, the benefits are overestimated. These phenomena are widespread and not restricted to IT projects. In order to reduce these occurrences we seek to understand the cognitive processes involved in such problem solving situations and also the interaction of these processes with the personality of the problem solver.

Problems may be divided into two general classes: well-defined and ill-defined. A well-defined problem can be described by finding a sequence of actions that alter the problem state and ultimately lead to a desirable goal. Typically, problems in software cost estimation might be viewed as ill-defined due to the inherent ambiguity and uncertainty.

In order to focus our enquiry, we restrict ourselves to estimation supported by CBR tools [4]. This is loosely based on human cognitive processes [5]. However, humans demonstrate a wide range of individual differences which may have a bearing upon performance. In particular we are interested in the interaction between personality and the cognitive processes involved in ill-defined problem solving.

This small-scale study was conducted at a large software development organisation, referred to as Company ABC, who allowed us access to expert participants. Thus we were able to work closely with five experienced project managers as they performed realistic estimation tasks using their company data and our CBR tool archANGEL [6]. First we conducted individual semi-structured interviews in order to obtain background information, then we trained the participants in the think-aloud protocol [7]. Once confident that they could solve a problem whilst thinking aloud, we asked the participants to solve effort estimation problems using the think-aloud protocol. Using this data plus the output log from archANGEL
we constructed models of their problem solving processes. We also assessed their personality using the Big Five personality trait instrument [8] to try to explain, at least in part, the substantial differences in approach and outcome.

The remainder of the paper is organised as follows. First we review related work that has adopted a human-centric view of software project prediction. This is followed by some contextual information on Company ABC, the participants and the task (and the data behind it). Then we describe our data collection and analysis methods. We then present our findings and a discussion of their significance. We conclude with a summary, review of the study limitations and recommendations for follow up work.

2. Related Work

Myrtveit and Stensrud [9] conducted one of the first experiments that investigated the relationship between formal or model-based approaches and experts. They found that a combination of expert and CBR tool outperformed either technique alone. In a different application domain (engineering projects), Rush and Roy [10] identify the need to study the intersection of expert and computerised cost models as “very little research tackles the issues of capturing and integrating [expert judgement] and rationale into the cost estimating process” (p271). Unfortunately, the situation has not changed substantially since they published their views.

Whilst the main thrust of research has been into the development of formal models or prediction systems there has also been some investigation of what is collectively known as ‘expert judgement’. Jørgensen has produced a useful systematic literature review [11] in which he points out, “there is no substantial evidence in favour of use of estimation models, and that there are situations where we can expect expert estimates to be more accurate than formal estimation models”. Of the 15 studies Jørgensen identifies, 5 find in favour of formal models, 5 find in favour of expert judgement, and 5 are unable to differentiate. This is one motivation for our work since the notion of replacing humans with algorithms or models seems somewhat optimistic and moreover, may not be beneficial. In addition, we see relatively low levels of formal prediction systems being used in practice. One possible explanation is that the human and cognitive aspects of using formal techniques and tools have received little attention [12].

There has also been a growing interest in the personality characteristics of software engineers. Researchers have used psychometric instruments such as questionnaires to measure personality in order to better understand the role personality plays in project outcome. Gorla and Lam [13] report that selecting the ‘right’ people for a project team can improve the chances of success. Capretz [14], using the Myers-Briggs Type Inventory (MBTI), found that there are more ISTJ (introverted, sensing, thinking, judging) type of software engineers than any other type. However, he adds the caveat that "no personality instruments will ever accurately predict success in this field" (p214).

Beyond computer science, there has been much interest in the cognitive aspects of problem solving per se. A systematic literature review by Mair et al. [15] found that analogical reasoning can assume a significant role in non-trivial problem solving, but that CBR tools do not model this in a biologically plausible way. For example, the ability to induce structure and therefore find deeper analogies is widely seen as a characteristic of expert performance. Yet, CBR tools fail to provide support for this type of reasoning; focusing instead on surface characteristics of problems, typically modelled as simple vectors of features.

The study described in this paper builds upon earlier work [16] which analysed problem solving strategies for two other participants (P1 and P2). In addition we provide a description of the results for task performance for all participants. Note, we also conducted preliminary semi-structured interviews but these are excluded from this paper for reasons of space.

3. Study Context

3.1. The Company

The Company ABC is a major, international software house that provides "business solutions" to its clients. It has been involved in information and communications technology (ICT) outsourcing for almost half a century. Its clients include government, local government, banks and the financial sector, the manufacturing and energy sectors worldwide.

Subsequent to conducting this research, ABC was taken over by a world leading, computer hardware and software company. ABC has had an extensive software measurement programme in place since the early 1990s and has 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), lines of code (LOC)) and total effort. Unfortunately there are extensive problems of missing values compounded by issues of data trustworthiness. However, guided by a Metrics Specialist from ABC, we identified a subset of 18 comparable, recent UK enhancement projects that
were relevant to the experience of the participants. These were then used as the basis of a relevant realistic prediction task for our expert participants.

Five highly experienced project managers (4 male and 1 female, age range 40-50) with typically 20+ years experience from ABC in the UK volunteered to participate. Since they were volunteers we have a convenience as opposed to a representative sample.

3.2. The CBR Tool

Over some years we have been interested in the application of CBR technology to support project managers' prediction and decision-making processes [17]. The basic philosophy is that history repeats itself but not exactly. Therefore the challenge is to find similar, but not necessarily identical cases so that a past solution can be used, perhaps with some adaptation, to solve the new problem. For effort prediction, a project is represented as a case which is made up of a vector of features that describe the project and its solution: the actual effort employed. This common choice does not represent a particularly rich problem structure. However, it has the benefits of simplicity. For more details see [18].

One of the tools we have developed is archANGEL which is a successor to the original ANGEL tool reported in [6]. It is a sophisticated shell (i.e. it is not specific to any particular problem domain or case-base) that provides support for both continuous and categorical case features. It has a simple spreadsheet style interface but is able to retrieve similar cases (analogies) in high-dimensional feature space using standardised Euclidean distance. It uses various search and meta-heuristic search algorithms to find better feature subsets based upon a wrapper [19]. It is highly configurable so that the user may choose which analogies to use and which features with which to select them.

archANGEL provides support for analogy based prediction and has been subject to considerable investigation by ourselves and other research teams. However, as our systematic review [20] revealed, of 20 empirical validation studies, the results are quite mixed (9 supporting, 4 neutral and 7 against) so we have moved to a new position of seeing CBR technology as an adjunct or support for the professional (rather than as a replacement).

3.3. Data

As indicated above, ABC has collected a considerable amount of project data. However for the purposes of this investigation we focused on a small subset. We chose projects completed between 2003 and 2008 that were comparable in terms of technology, client and methodology and were UK-based and relevant to our participants. In addition, we needed to exclude projects with incomplete data or situations where the quality of the data was questionable. Applying these filters resulted in 18 projects being selected.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>UFP</td>
<td>Unadjusted FPs</td>
<td>90</td>
<td>658</td>
<td>1719</td>
</tr>
<tr>
<td>AFP</td>
<td>Adjusted FPs</td>
<td>90</td>
<td>701</td>
<td>1822</td>
</tr>
<tr>
<td>LOC</td>
<td>Lines of code</td>
<td>2676</td>
<td>29940</td>
<td>64031</td>
</tr>
<tr>
<td>Duration</td>
<td>In days</td>
<td>192</td>
<td>393</td>
<td>544</td>
</tr>
<tr>
<td>Effort</td>
<td>In person hours</td>
<td>6174</td>
<td>18200</td>
<td>50886</td>
</tr>
<tr>
<td>MaxFTE</td>
<td>Max full-time equivalent staff</td>
<td>5.8</td>
<td>15.7</td>
<td>30</td>
</tr>
<tr>
<td>PeakStaff</td>
<td>Max number of personnel</td>
<td>8</td>
<td>25.2</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 1: Summary data of project size and effort information

The projects ranged in Effort from 6174 to 50886 person-hours, approximately 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. 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.

Although not contrived, it is important to note the absence of simple linear relationships between the size type measures such as function points (FPs) and project effort. In Figs. 1 and 2 we use scatterplots to show the poor fit of a simple linear relationship between project size as measured by either FPs or LOC and development effort. (The highlighted data points indicate projects with extreme values.)

More generally, the ill-defined relationship means that simple ratio-based approaches to estimation are unlikely to be effective given the nature of the scatter about the regression line. Table 2 indicates the diversity of productivity rates either in terms of LOC or FPs. We see in this sample of 18, supposedly homogeneous projects, variation in productivity rates of twenty and thirty-fold respectively. This underscores the challenges of software project effort and cost estimation. Nor is there a stable ratio between specification size (reported in FPs) and implementation size (reported in LOC) leading to a Pearson correlation coefficient of r = 0.63. These characteristics of the data caused the participants some difficulties for the prediction task, something we will shortly return to.
<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours per FP</td>
<td>7.1</td>
<td>41.5</td>
<td>153.5</td>
</tr>
<tr>
<td>Hours per LOC</td>
<td>0.2</td>
<td>1.3</td>
<td>6.6</td>
</tr>
<tr>
<td>LOC per FP</td>
<td>7.9</td>
<td>52.8</td>
<td>137.7</td>
</tr>
</tbody>
</table>

Table 2: Summary data of project productivity ratios

Finally, we note that even a naïve approach to this data set using our CBR tool archANGEL with a leave-one-out cross-validation could yield average prediction errors of less than 20%. Although the purpose of this research is not to demonstrate the superiority of one particular method, it is a helpful finding in that participants could potentially obtain good results from use of this CBR method. Of course the extent to which they accomplish this is part of our investigation.

4. Procedure

4.1. Personality Measures

We are interested in measuring the underlying personality traits of our participants, that is the “enduring patterns of perceiving, relating to, and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts” ([21] p.686). Trait measures assessing personality have several advantages over personality type measures such as the popular Myers-Briggs Type Inventory [22]. Trait measures tend to describe personality much better because personality traits are seen as part of a continuous dimension with many people falling in the middle [23]; whereas personality types are seen as qualitatively distinct categories. Also, traits are genetically or biologically based [24] and reflect durable personality patterns whereas types tend to be a product of a particular place, time, and culture [23]. Finally trait scales provide numerical scores for each characteristic that makes more detailed statistical analysis possible. However, please note, we do not derive any statistical inference from the small sample in this study.

To assess personality we used the BFI [8] which is based on five major traits: namely Openness to Experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism. This measure consists of 44 items and it uses a five point Likert scale (1=disagree strongly, 2=disagree a little, 3=neither agree nor disagree, 4=agree a little, 5=agree strongly). Each trait in the BFI is now summarised in more detail.

Openness to Experience: this trait measures proactive seeking and appreciation of experience for its own sake, tolerance of, and exploration of the unfamiliar. It suggests curiosity, imagination, creativity and many interests. Low scores indicate a preference for a balance between old and new values.

Conscientiousness: this trait measures the degree to which a person is persevering, responsible, and organized as opposed to lazy, irresponsible, and impulsive.

Extraversion: a high score for Extraversion characterises sociability, 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, 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. The Extraversion trait in the BFI measures the degree to which a person is sociable, leader-like, and assertive as opposed to withdrawn, quiet, and reserved.

The Agreeableness trait in the BFI measures the degree to which a person is warm and cooperative as opposed to unpleasant and disagreeable.

Neuroticism: scoring high on this trait (indicating emotionality or anxiety) characterises high levels of depression and anxiety, low self-esteem and feelings of guilt.

4.2. The Estimation task

The estimation task was devised by taking one of the 18 completed projects previously described in Section 3.3 and reducing the actual effort threefold (i.e. 15258 to 5086 person-hours). Each participant was then given this scenario: ‘This [reduced] estimate has been provided by another, unknown, manager. Your task is to perform a sanity check using the data set of projects and the CBR tool archANGEL. Please express whether you agree or disagree with the estimate. If you disagree, please provide your preferred estimate as both a point value and a range’.

Since the participants were unfamiliar with the CBR tool, we provided a 15-20 minute demonstration with a ‘toy’ example. However, since we were not investigating specific tool learnability or usability per se, one of the investigators ‘drove’ the tool, while following instructions given by the participant. Thus this was a form of action research [25]. As the participant carried out the estimation task, he or she used a think-aloud protocol which was audio-recorded and subsequently transcribed. In addition, archANGEL also generated a log file of usage which provided a record for subsequent analysis.

Other than requiring the occasional prompt of "what are you thinking now" the participants appeared comfortable with thinking aloud.

5. Results

In this section we review the detailed results of this study. Recall that the BFI was administered and that it measures the five dimensions of personality (Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism). Most participants showed very similar scores, however where there is a deviation this is highlighted in bold, Table 3 summarises the individual participant scores and shows participant score/max possible score. Please note this is a convenience, not a random, sample so we make no generalisations or statement about ‘typical’ project manager personality.

There was a good deal of similarity between the participants. All five were highly conscientious. That is, well-organized, reliable, hard-working, ambitious, persevering individuals who have a strong focus on achieving goals. The score for Neuroticism was medium for all participants which suggests that overall they were emotionally stable. All participants scored high on Agreeableness which indicated that participants could easily identify with, and understand, another’s situation, feelings and motives which may be reflected in their moods.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Neuroticism</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Med</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Med</td>
</tr>
<tr>
<td>P2</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Med</td>
</tr>
<tr>
<td>P3</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Med</td>
</tr>
<tr>
<td>P4</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Med</td>
</tr>
<tr>
<td>P5</td>
<td>Med</td>
<td>High</td>
<td>Med</td>
<td>High</td>
<td>Med</td>
</tr>
</tbody>
</table>

Table 3: BFI scores for the study participants

The areas of difference between our participants related to Openness to some extent, and more notably for Extraversion.

Openness: The mean score for Openness was slightly above the average. This indicates that participants tend to take risks and welcome new, exciting experiences and sensations, but that they are not generally impulsive and think carefully before making a decision. Two participants scored medium and three participants scored high on this trait. This indicates that the former prefer a balance between old and new values and are fairly open to new experiences whereas latter are more curious, imaginative, creative, and have many interests.

Extraversion: Three out of five managers' score for Extraversion was above average showing that they are extraverted individuals who are enthusiastic, active, outgoing and sociable, preferring to be around people much of the time. However, P1 scored low, and P5 scored medium.

Table 4 provides quantitative data regarding the task performance of the five participants. The participants were not time constrained and we tried not to convey any expectations. However, the time to complete task ranged between 19 and 42 minutes and
there does not appear to be any relationship between time taken and accuracy of prediction. For example, P1 took the longest time (42 minutes) yet in the end was unwilling to provide a specific effort prediction. The other participants provided a point value first and were then prompted for a range.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Time (mins)</th>
<th>Predicted</th>
<th>Range</th>
<th>Accuracy (Residual)</th>
<th>Accuracy (Abs Rel Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>42</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P2</td>
<td>32</td>
<td>20000</td>
<td>16000-20000</td>
<td>4742</td>
<td>31%</td>
</tr>
<tr>
<td>P3</td>
<td>24</td>
<td>15000</td>
<td>11000-19458</td>
<td>-258</td>
<td>2%</td>
</tr>
<tr>
<td>P4</td>
<td>19</td>
<td>12000</td>
<td>11000-13000</td>
<td>-3258</td>
<td>21%</td>
</tr>
<tr>
<td>P5</td>
<td>20</td>
<td>10000</td>
<td>9500-10500</td>
<td>-5258</td>
<td>35%</td>
</tr>
<tr>
<td>Mean</td>
<td>27.4</td>
<td>14250</td>
<td>9500-20000</td>
<td>-1008</td>
<td>22%</td>
</tr>
</tbody>
</table>

**Table 4: Participant overall performance**

The true effort value for the project was 15258 person-hours. Only one range (that provided by P3) encompassed the true effort value. Three out of four participants underestimated the effort although this may have been influenced by the low anchor of the prediction provided (5086 person-hours) [26], [27]. Finally, we note that P5 was the most confident (with the narrowest bounds) but at the same time the least accurate. By contrast, P2 was the most accurate yet least confident (with the widest bounds). Although P2, P3 and P4 did not differ much in their BFI profiles, P2 did have the highest score (44/50) of the group for Openness. This may also reflect differing levels of self-reflection and criticism demonstrated in the two participants reported in this paper. Reflection is a component of metacognition (thinking about thinking) which is generally seen as positively correlated with task performance. For example, in a study investigating the effects of metacognitive training on mathematical reasoning and metacognitive skills, Krumsaks and Mevarech [28] found that exposure to metacognitive instruction led to improved performance in a number of areas including the use of logical arguments to support math reasoning.

In the present study, each manager adopted a different approach to constructing an interval about the predicted point (see Table 5) and only two participants (P4 and P5) assumed strictly symmetric distributions. There was also some diversity in terms of spread ranging from 10% to 57%. NB all percentages are calculated relative to the predicted value.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Strategy description</th>
<th>(Lower, estimate, upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Insufficient info to predict</td>
<td>n.a.</td>
</tr>
<tr>
<td>P2</td>
<td>Predicted = upper bound</td>
<td>(-20%, 0%, 0%)</td>
</tr>
<tr>
<td>P3</td>
<td>Predicted is approx mid point; upper bound based on archANGEL</td>
<td>(-27%, 0%, 30%)</td>
</tr>
<tr>
<td>P4</td>
<td>Predicted is midpoint</td>
<td>(-8%, 0%, 8%)</td>
</tr>
<tr>
<td>P5</td>
<td>Predicted is midpoint</td>
<td>(-5%, 0%, 5%)</td>
</tr>
</tbody>
</table>

**Table 5: Participant estimation range strategies**

Table 6 shows the counts of unique references to cases (i.e. other projects that are potential analogies) and features (i.e. project attributes such as date and LOC). Note that whilst we selected 18 projects or cases, one serves as the target and another was used for a separate task not reported here, hence the maximum number of analogies that might be used is 16. The counts were obtained through an analysis of the think-aloud transcripts and include any reference, not just via archANGEL. Where there are multiple references to the same case or feature this is only counted once. Notable is that P3 uses the fewest features and notably fewer cases in the prediction process. By contrast P1 appears extremely unselective and considers all the cases in the case-base.

Using the transcripts from the think-aloud protocols recorded during the prediction task we applied a simple coding scheme detailed in Table 7. The transcripts were analysed independently by the two authors. Discussions resolved inconsistencies and led to the derivation of five categories that described the cognitive processes of P1 and P4 (Table 7).

<table>
<thead>
<tr>
<th>Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarifying</td>
<td>Trying to gain better understanding of the task or problem environment for instance by talking to others e.g. &quot;it’s important that I understand the scope of the project&quot;</td>
</tr>
</tbody>
</table>
Reflecting | Awareness or assessment of one's problem solving process e.g. "so I can better see the wood for the trees" or the quality of the solution e.g. "whereas if you were given a number of days to do [the task], you’d do a better job"

Confidence (negative) | Expectation of a negative outcome e.g. "I treat information as unproven until I’m happy with it" or one’s abilities e.g. "I think it’s a shortfall in my competency … that I wouldn’t be able to use function point information"

Confidence (positive) | Expectation of a positive outcome e.g. "So I would be confident from the point of view that the various parameters …".

Organising | Articulating procedural aspects of the problem solving process e.g. "The first thing I’d want to do would be to filter …".

**Table 7: Code definitions**

Next we examine each participant’s problem solving process in more detail. For reasons of space we focus upon P1 and P4.

<table>
<thead>
<tr>
<th>Code</th>
<th>P1</th>
<th>P4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarifying</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Reflecting</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>Negative Confidence</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Positive Confidence</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Organising</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Total code count</td>
<td>53</td>
<td>23</td>
</tr>
<tr>
<td>Task time (minutes)</td>
<td>42</td>
<td>19</td>
</tr>
</tbody>
</table>

**Table 8: Code frequencies by participant**

**Figure 3: Code proportions by participant**

Table 8 and Figure 3 show the absolute counts and relative proportions of the code types for participants P1 and P4 respectively. The proportions are a better indicator of differences because the task durations and therefore the transcript lengths differed. Both were reflective in their problem solving behaviour, e.g. P4 commented "I would ask serious questions about that and I would be...". Here they are articulating whether they need to probe more deeply. Another example comes from P1 who observes "so I can better see the wood for the trees". In this case P1 seems to be reminding themselves not to lose the "big picture" through excessive detail. Such reflection is to be expected from mature and effective professionals.

Nevertheless it is clear that there are substantial differences between P1 and P4. Most notably P1 makes far more negative confidence utterances than P4, e.g. "my expertise in function points is not good" revealing a lack of confidence in their own abilities and "I treat information as unproven until I’m happy with it" revealing a cautious, even pessimistic, view of the problem domain. Interestingly P4 seemed to be aware of their rather more positive view of the problem domain and ended the task by commenting "and it would be optimistic estimate as well but I always go for optimistic".

Recall that P1 was in the end unwilling to provide a solution for the prediction task. Much of the problem-solving was characterised by a need to "have a handle on the data". However, P1’s endeavours were confounded by the inability to find linear relationships and this "set off alarm bells". As a result P1 considered that the "underlying database may be unreliable" and so did not wish to proceed. The idea of seeking complex, non-linear relationships in high-dimensional feature-space threatened P1’s need for understanding and so P1 chose not to engage with a CBR tool that did not support P1’s preferred problem solving style.

Some of the tension that P1 seemed to feel between both finding the data useful and unhelpful might be explained by the dual-process theory of cognition, that is preferring intuitive answers but more analytic justifications. For example, P1 states "I’m aware that what you think numbers are telling you may turn out to be totally different and relationships that you think are there aren’t really, but nonetheless I’d still draw some comfort by [them]". The difficulty was the data were not saying what P1 expected from their intuition.

By contrast P4 started to engage with the CBR tool quickly and then advanced to searching for analogies using more than one feature. Another characteristic of P4’s approach was hypothesis testing in order to assess the adequacy of a proposed solution. This may have been easier for P4 due to a more pronounced Openness trait than P1 (see Table 3). Although P4 was willing to provide a solution (12000 hours and quite tight bounds 11000-13000) they did comment that they would like
"more historical data" indicating an awareness of the limitations of their solution.

6. Discussion

The five participants differed in personality in terms of two traits: Extraversion and Openness. We also observed non-trivial differences in prediction processes.

P1 took considerably longer than any other participant yet was reluctant to commit to a specific answer without further information. P1 also made little use of archANGEL whilst the others used it extensively. P2-P5 were happy to provide single point values and then when prompted a confidence interval in the form of a range.

As the participants did not vary on other traits, the question arises to what extent can the different approaches be explained in terms of Extraversion and Openness? Other studies have found major 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 Openness, or risk taking, may be constrained by the artificial nature of the task since little rested upon the outcome. We also note that, results from the Extraversion and Openness traits seem to be contradictory in that they are negatively correlated. Further investigation is required along with the use of other personality measures such as the MBTI. 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 Huit [29] 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 internal consistency whilst in contrast P2, P3 and P4 operated more in an outside world and attended to the “external reality”. P5 fell somewhere between the two groups in terms of Extroversion, but of the participants who provided an estimate he or she was the most optimistic and most over-confident.

Consequently P1, found the 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 Openness trait. P1, perhaps surprisingly, given the below average score for Extraversion exhibited an above average score for Openness. 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, e.g., a multi-million pound or euro project.

7. Conclusions

To recap, in this paper we have described an empirical investigation of the relationship among personality, cognitive and prediction behaviour of professionals when using a CBR tool. We worked with five experienced software project managers and observed an estimation task using a think-aloud protocol. The task was based upon actual project data from the collaborating organisation. This was complemented by a personality assessment using the BFI personality questionnaire. This led to four main findings.

First, it is clear that project effort prediction is not easy. The data were drawn from a single organisation and were enhancement projects for a single client using similar technology. Yet there were no simple underlying relationships in the data.

Second, although the participants were all from the same company there were wide differences in performance ranging from a 2% to a -35% error plus one participant did not even wish to propose a solution. The estimates tended to be optimistic (3 out of 4) and over-confident (3 out of 4 ranges not encompassing the true value). This is line with other studies such as [30].

Third, the ability to utilize a CBR tool is in part determined by personality which in turn impacts on problem solving behavior.

Fourth, software tool developers and those suggesting formal estimation methods should not assume a single style of problem solving. We observed in one out of five participants such a contrast between the participant’s preferred approach and that expected by the CBR tool that the manager was not able to make meaningful use even when supported by an expert user. Other studies (see the systematic review [15]) have also commented upon how the ‘mismatch’ between experts’ cognitive processes and CBR tools contributes to the erratic performance of CBR-based prediction.

As with any empirical study there are threats to validity. In this case the main issues revolve around
generalisation. However, in one sense observation studies by their very nature cannot generalise. We do not have a random sample, and the small sample of five participants was drawn from a single company. However, this work generalises into theory, such as those relating task performance to personality trait, e.g. [29] and contributing to a body of evidence.

Clearly there is scope for much follow up work. We used the BFI because we sought to understand the impact of personality traits on cognitive performance, however, because traits are unchanging, they are not context specific. The MBTI, extensively used in Occupational psychology, is context specific and therefore, may be more useful in determining the personality we demonstrate in the workplace.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council under grants EP/G007683 and EP/G008388. We are also indebted to the participants in this study who freely gave of their time.

References


