The Effects of User Characteristics on Query Performance in the Presence of Information Request Ambiguity

ABSTRACT

This paper investigates the effects of personality characteristics on individuals’ abilities to compose queries from information requests containing various types of ambiguity. In particular, this research examines the effects of user personality characteristics on query performance in the presence of information requests that contained no, extraneous, syntactic, or both extraneous and syntactic ambiguities. The results indicate that personality characteristics significantly affect users’ abilities to compose accurate queries. Neuroticism, agreeableness, openness to experience, and conscientiousness significantly affected the number of errors made in the query formulations. Conscientiousness affected the length of time taken to compose the queries and neuroticism affected the confidence users had in the accuracy of their queries. Although several personality dimensions affected query performance, no significant interactions between personality dimensions and ambiguity were detected. Furthermore, both query complexity and information request ambiguity exhibited greater impacts on query performance than personality characteristics. Hence, organizations should attempt to train users to deal with query complexity and information request ambiguity before modifying their training programs for personality characteristics.

(Keywords: Information retrieval, personality, NEO PI-R, ambiguity)
1. INTRODUCTION

In today's highly competitive business environments (Goldstein and Storey, 1994; Cascia and Sanseverino, 1997), organizations are encouraging managers and other end users to query information repositories themselves (Delligatta and Umbaugh, 1993; Owei, 2003). Frequently, the queries these users compose are to satisfy information requests they receive from other stakeholders, e.g., managers, trading partners, and regulatory officials. These information requests are often posed in natural language and typically contain ambiguities such as imprecise adjectives and excessive scope (Whitten et al., 2001). To successfully retrieve the desired information, users require an appropriate skill set (Lerouge et al., 2005). This skill set includes the ability to recognise and ultimately resolve the imprecision contained in an information request. Indeed, good communication skills generally distinguish effective managers from poor managers (Stephens, 1982).

Within a business environment, a substantial portion of managerial communication consists of written or verbal requests for information from associates who must transform the request into queries in order to extract data stored in electronic form (Tubre and Collins, 2000). Being in natural language, these information requests often contain ambiguities. Hence, the focus of this research is on the communication ambiguities that may arise between two persons, i.e., the originator of the information (hereafter referred to as the information request provider) and the query developer.

Ambiguities in communication, unless resolved, can affect performance (e.g., Cowie and Lehnert, 1996). Furthermore, personality traits can affect the ability of persons to resolve ambiguities (Mumford et al., 1993). Together, these ideas suggest that people with different

---

1 Within this research, only verbal ambiguity arising between a information request provided and a query developer is investigated as opposed to situational ambiguity and role ambiguity. Both situational and role ambiguities may also be present within managerial decision making contexts.
personality types may cope better with ambiguities and thus achieve better results from database queries than people of other personality types. Accordingly, this paper investigates the effects of personality characteristics on individuals’ abilities to resolve ambiguities in an information retrieval task. In particular, this research examines the effects on query performance of the interaction of personality characteristics in the presence of information requests that contained no, extraneous, syntactic, or both extraneous and syntactic ambiguities. The personality dimensions examined are neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. These dimensions were measured using the Neuroticism, Extraversion, Openness, Personality Inventory-Revised Model (NEO PI-R).

Providing insights into the effects of personality characteristics and communication ambiguities on information retrieval is important for a number of reasons. First, the two types of ambiguity investigated here, syntactic and extraneous, occur frequently in organizational communications. These ambiguities exist because of the inability to communicate clearly in conjunction with the tendency not to question our communication skills. Few organizations and individuals adequately recognise that many communications, both internal and external, contain ambiguities. Second, these ambiguities can potentially have disastrous effects on organizational decision making, e.g., ambiguous information requests by stakeholders can lead to incorrect interpretations by the recipients, queries that produce reports that differ from the desired information, and erroneous decisions. By examining whether or not personality can affect an individuals ability to resolve communication ambiguities, organizations can tailor educational programs to reduce the negative effects of these ambiguities. Third, from an academic perspective, little or no research has investigated the potential affects of personality on the ability to resolve ambiguities.
The next section of this paper reviews research on personality, ambiguity, and information retrieval, and combines the two areas of research to develop a number of hypotheses investigating the relationships between verbal ambiguity, personality, and information retrieval. Section three explains the research method and section four presents the results. The implications of the research are presented in section five before the conclusions, limitations, and directions for future research are discussed in section six.

2. THEORETICAL FOUNDATIONS AND HYPOTHESIS DEVELOPMENT

This section describes the relevance of personality to information systems research and briefly examines one model of personality. The section then discusses the two types of ambiguity investigated within this research and how these ambiguities affect information retrieval. Finally, the section develops five hypotheses that examine the effects of the five personality factors in the presence of syntactic and extraneous ambiguities.

2.1 Personality

Personality refers to the cognitive and affective structures maintained by individuals to facilitate their adjustments to the events, people, and situations they encounter (Gough, 1976). Personality variables such as locus of control, ambiguity tolerance, and attitude affect individuals’ ability to articulate and evaluate information related tasks. These variables impact perceived usefulness, actual system use, and ultimately MIS success (Zmud, 1979; Taylor, 2004). The role of personality on employee behavior has been the focus of organizational psychologists for many years (Johnson, 2003). When carefully applied within the context of specific occupations and organizations, personality variables are significant predictors of job performance (Day and Silverman, 1989; George, 1992; van der Berg and Feij, 2003).
Personality variables can potentially increase our understanding of the links between information systems and human cognition (Wheeler et al., 2004). During the 70s and 80s, individual traits, psychological types, and cognitive style were heavily researched within the area of human-computer interaction (Banker and Kauffman, 2004). With the increasing interest in human-computer interaction, Banker and Kauffman (2004) predict a resurgence of this research stream. Recent research into individual traits and computer anxiety (Thatcher and Perrewe, 2002) and personal attributes and their impact on adaptation to technological change (Gallivan, 2004) support their prediction. When faced with unclear information, personality characteristics impact peoples’ ability to provide solutions to problems (Back and Seaker, 2004). Furthermore, personality characteristics have been shown to affect the query errors made during information retrieval (Bowen et al., 2003).

A number of models of personality exist. Jung (1962) developed one of the best known theoretical bases used to examine personality and personality types. Jung’s theory asserts that individuals have a certain social orientation (either Introverted or Extroverted), prefer one way of perceiving (either Sensing or Intuition), and have one primary way of judging (either Thinking or Feeling) (Johnson et al., 2001). Myers and Briggs added another dimension, during the development of the MBTI, to describe how an individual primarily deals with the outer world (either Judging or Perceiving) (Nordvik, 1996). The resulting Myers Briggs Type Indicator (MBTI) is one widely accepted and reliable measures of personality type (Furnham et al., 2003).

Recently another model is receiving considerable attention. The Five Factor Model has been derived from a “more theoretically neutral position” (Widiger and Trull, 1997, pg. 229). While most other models of personality have evolved from one theoretical perspective (as is the case in MBTI), the Five Factor Model has not been derived from one single theory. Rather, the
Pre-print version of article

Five Factor Model examined the personality traits that people consider most important in describing themselves. The predominant instrument for operationalizing the Five Factor Model is the NEO PI-R (Clark, 2003). This instrument has “demonstrated exceptional psychometric properties” (Zhang, 2002; pg. 19).

The Five Factor model categorizes personality traits into five major dimensions: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). Each dimension is comprised of six facets. The facets contained within each dimension are described in sections 2.1.1 through 2.1.5 and summarized in Table 1. The creation and validation of NEO has had a significant impact on research into the effects of personality (Byrne et al., 2005). The Five-Factor Model of personality has become the dominant basis for investigating the effects of personality traits (Goldgerg, 1993), i.e., the five factor NEO-PI is the most extensively used measure in recent academic research (Furnham, 1996; Lampe, 2004). The model has been used in studies of job performance (Barrick and Mount, 1991; Barrick and Mount, 1993), career success (Judge et al., 1999), and job satisfaction and work adjustment (Tenopyr, 1993).

Table 1: NEO PI-R dimensions and facets

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Neuroticism</th>
<th>Extraversion</th>
<th>Openness to Experience</th>
<th>Agreeableness</th>
<th>Conscientiousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facets</td>
<td>Anxiety</td>
<td>Warmth</td>
<td>Fantasy</td>
<td>Trust</td>
<td>Competence</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
<td>Gregariousness</td>
<td>Aesthetics</td>
<td>Modesty</td>
<td>Self-Discipline</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>Assertiveness</td>
<td>Feelings</td>
<td>Compliance</td>
<td>Achievement-Striving</td>
</tr>
<tr>
<td></td>
<td>Self-consciousness</td>
<td>Activity</td>
<td>Actions</td>
<td>Altruism</td>
<td>Dutifulness</td>
</tr>
<tr>
<td></td>
<td>Impulsive</td>
<td>Excitement-seeking</td>
<td>Ideas</td>
<td>Straightforwardness</td>
<td>Order</td>
</tr>
<tr>
<td></td>
<td>Vulnerability</td>
<td>Positive emotions</td>
<td>Values</td>
<td>Tender-Mindedness</td>
<td>Deliberation</td>
</tr>
</tbody>
</table>

2.1.1 Neuroticism

Neuroticism is the tendency to experience negative emotions such as anxiety, hostility, and depression. From Table 1, the neuroticism dimension is comprised of the following facets:
anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability. Traits in the neuroticism dimension reflect different ways of reacting to distress circumstances (Costa and McCrae, 1995). Low neuroticism individuals are more able to cope with stressful conditions without becoming overly upset or anxious. Individuals with high levels of neuroticism often have stronger emotional reactions to stressful conditions. Neuroticism has been found to be negatively associated with career success (Judge et al., 1999). The Neuroticism dimension is not included in the MBTI.

2.1.2 Extraversion

Extraversion within the Five Factor Model refers to “high activity, sociability, and a tendency to experience positive emotions” (Furnham et al., 2003, pg. 577). From Table 1, the extraversion dimension is comprised of the following facets: warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotions. Extraverts are social, active, talkative, upbeat, energetic, and optimistic (Costa and McCrae, 1992). Individuals deficient in these characteristics are termed Introverted. The conceptualization of extraversion embodied in NEO is not the same as the concept of extraversion in Jung’s theory. For example, within NEO, introspection is not related to extraversion but is a characteristic of openness to experience. Extraversion on the five factor model is, however, positively associated with the Extraversion trait and negatively associated with the Introversion trait on the Extraversion-Introversion dimension of MBTI (Furnham et al., 2003). Extraversion, as operationalized by the NEO-PI R, has been shown to be negatively associated with academic performance. Extraversion has also been shown to be related to job performance among certain occupational groups, e.g., salesmen and managers (Hurtz and Donovan, 2000). The performance of other groups, e.g., professionals
and skilled workers, does not appear to be strongly associated with the dimension (van der Berg and Feij, 2003).

2.1.3 Openness to Experience

Within the Five Factor Model, openness to experience refers to the “tendency to involve in intellectual activities and new experiences” (Furnham et al., 2003, 578). From Table 1, the openness to experience dimension is comprised of the following facets: fantasy, aesthetics, feelings, actions, ideas, and values. The factor relates to intellect, acceptance of novelty, cultural interests, educational aptitude, and creativity, as well as an interest in varied sensory and cognitive experiences (Howard and Howard, 1995). Traits in the openness to experience dimension attempt to capture the process of using cognition, intelligence, and contemplativeness together with unconventionality (Judge et al., 1999). Individuals who score low on this dimension tend to be conventional in behavior and conservative in their outlook, and to prefer familiar and recognisable situations. Low openness to experience individuals also tend to be less creative. Individuals who score high on this dimension are more imaginative and open-minded. They are willing to entertain novel ideas, to think more divergently, and to have positive attitudes towards learning (Costa and McCrae, 1992; Barrick and Mount, 1991). The openness to experience dimension has been shown to be related to training proficiency (Barrick and Mount, 1991). Openness to experience on the five factor model is positively associated with the Intuition trait and negatively associated with the Sensing trait on the Sensing-Intuition dimension of MBTI (Furnham et al., 2003).

2.1.4 Agreeableness
Agreeableness within the Five Factor Model refers to “friendly, considerate, and modest behaviour” (Furnham et al., 2003, pg. 578). From Table 1, the agreeableness dimension is comprised of the following facets: trust, modesty, compliance, altruism, straightforwardness, and tender-mindedness. Traits in the agreeableness dimension reflect styles of interpersonal behaviour and interaction (Costa et al., 1991). Individuals with high levels of agreeableness are fundamentally unselfish. They are compassionate and cooperative, and tend to believe others behave in a similar manner (Judge et al., 1999). Individuals with low levels of agreeableness tend to be non-compliant, critical, and sceptical. Agreeableness on the five factor model is positively associated with the Feeling trait and negatively associated with the Thinking trait on the Thinking-Feeling dimension of MBTI (Furnham et al., 2003).

2.1.5 Conscientiousness

Conscientiousness within the Five Factor Model refers to “persistence, self-discipline, and the need for achievement” (Furnham et al., 2003, pg. 578). From Table 1, the conscientiousness dimension is comprised of the following facets: competence, self-discipline, achievement-striving, dutifulness, order, and deliberation. Conscientiousness on the five factor model is positively associated with the Judging trait and negatively associated with the Perceiving trait on the Judging-Perceiving dimension of MBTI (Furnham et al., 2003). As such, conscientiousness is a measure of the way persons process information, i.e., whether their Thinking trait or Feeling trait is dominant. Traits in the conscientiousness dimension describe differences in an individual’s motivation and persistence. Individuals with a high level of conscientiousness are competent, productive, and well-organized. As such, they approach problem solving in a highly organised and structured manner that tends to lead to sensible decisions. Individuals with low levels of conscientiousness tend to be prone to procrastination and prefer a “lackadaisical”
approach towards achieving their goals. The conscientiousness dimension has been the most consistent predictor of job performance (van der Berg and Feij, 2003).

2.2 Ambiguity and Information Retrieval

Research has been conducted into the modelling of data imprecision and data uncertainty within the design of data models and databases (Ma, 2005). Incongruities between information requests and data representations adversely affect end-user accuracy, time taken, and confidence (Borthick et al., 2001). Especially for stakeholders with little prior association with the relevant data repositories, metadata about the entities, relationships, and attributes in these data repositories are often ambiguous. Unfortunately, overcoming the ambiguity/uncertainty associated with the metadata will not necessarily lead to improved decision making, i.e., resolving metadata ambiguities is a necessary but not sufficient condition for effective information retrieval. Understanding the data model and the associated metadata removes one type of ambiguity, i.e., the query developer must completely understand the data structure with which they are working. The query developer must, however, also understand the information request they are given.

Within the realm of information retrieval, a person receives an information request, interprets that request, and formulates a query to retrieve the required information from a data repository, i.e., a database, data mart, or data warehouse. Knowledge workers can access these data repositories via a wide variety of end-user analytical tools including graphical query interfaces, report writers, OLAP cube builders, and data mining tools as well as the more traditional database query languages (Speier and Morris, 2003). The presence of ambiguity in an information request is likely to lead to multiple valid interpretations of the desired information request. Because of the multiplicity of valid interpretations, the information retrieved may not be
the information desired by the person making the request. Use of the potentially inappropriate information can have significant negative ramifications on business decision-making processes. Walton (1996) identified six ambiguity types: lexical, syntactical, inflective, pragmatic, emphatic, and suggestive. Axelsen et al. (2001) expanded Walton’s taxonomy to include a seventh type of ambiguity, extraneous ambiguity. Axelsen et al. (2001) examined, within one experiment, all seven different types of ambiguity. They found two types of ambiguity, syntactic and extraneous, to significantly affect a person’s performance during the query composition process. Their results indicate that syntactic and extraneous ambiguity strongly affect people’s ability to correctly translate information requests into queries that extract the information desired by the requestor. Interestingly, these two types of ambiguities are closely related to Bonner’s (1994) classification of task characteristics that are elements of task complexity. Bonner classified task characteristics that are elements of task complexity as relating to either the amount of information (extraneous ambiguity) or clarity of information (syntactic ambiguity). Bonner found both appeared to be negatively related to performance, however, only the effect for clarity of information was significant.

Van Gompel et al. (2005) found that globally syntactically ambiguous sentences are as easy to read as sentences containing no ambiguities. They were, however, not investigating the accuracy with which the ambiguities were resolved. Furthermore, they suspected that the person reading the globally ambiguous sentence may be failing to notice or resolve the ambiguity.

2.2.1 Syntactic Ambiguity

---

2 These two types of ambiguity are also considered important because they are two of the most common forms of ambiguity in everyday business communications.
Syntactic ambiguity, i.e., structural or grammatical ambiguity, often results in recipients being unclear or mistaken as to the subject or the object of a sentence. An example of syntactical ambiguity occurs in the information request:

“Provide a report of current wine inventory and suppliers that determines their sales for the last month.”

The request is syntactically ambiguous, as “their” can refer to either the wine inventory or the suppliers. As illustrated by this example, one of the most common forms of syntactic ambiguity is the use of indefinite pronouns where the pronoun’s antecedent is not clear.

2.2.2 Extraneous Ambiguity

Extraneous ambiguity arises when information is included that is not required to complete the current task. Some extraneous communications are clearly not relevant to the task at hand, e.g., small talk, and may even be misleading, e.g., discussions of other projects when one or more participants mistakenly think the discussion does, indeed, affect the current task/project. Axelsen et al. (2001) found that excess information impairs people’s ability to recognise critical elements of an information request. An example of extraneous ambiguity occurs in the information request:

"After the recent spate of increases in premiums, the new insurer wants to know about the major risks of theft or damage. Generate a report containing item number, item name, item maker, item year, and total value, i.e., quantity on hand times average unit cost, where total value is greater $5,000 or item year is less than 1994."

The first sentence represents extraneous ambiguity. The sentence is unnecessary but could potentially be beneficial to the recipients by explaining the motivation for the information request. The extraneous information could, however, confuse the recipients and cause them to
misinterpret the information request, e.g., by expanding the scope of the query. This research extends the work undertaken by Axelsen et al. (2001) by examining whether some personality types can resolve syntactic and extraneous ambiguities better than other personality types.

2.3 Ambiguity, Personality, and Information Retrieval

This section develops a set of testable hypotheses. Each hypothesis examines the effects of the five personality factors on query performance in the presence of syntactic and extraneous ambiguity. The relationships discussed above between ambiguity, personality, and information retrieval performance are depicted in Figure 1.

---

**Figure 1 – Model Tested**

2.3.1 Neuroticism, Ambiguity, and Information Retrieval

Once an information request has been received, to formulate a query users interpret the components of the request relative to the tables and attributes in the data structure. When individuals undertake more demanding attentional tasks, higher levels of neuroticism are associated with worse task performance (Szymura and Wodniecka, 2003; Wallace and Newman,
Pre-print version of article

1998). Similar findings from Hurtz and Donovan (2003) revealed that lower levels of neuroticism (emotional stability) led to better performance.

Introducing ambiguity (either syntactic or extraneous) into an information request increases task complexity (Bonner, 1994) and thus makes the task of formulating the query more demanding and potentially more stressful. Persons with lower levels of neuroticism are better able to remain calm and less anxious when placed in stressful situations (Costa and McCrae, 1995). This increased stress associated with ambiguity invokes a negative emotional response in persons with higher levels of neuroticism and negatively affects their performance (Furnham et al., 2002). High neuroticism users faced with both types of ambiguity simultaneously are likely to experience even more difficulty and even greater negative effects on their performance.

Within the area of information retrieval three different measures of performance are of interest. First, when placed in more stressful situations, persons with higher levels of neuroticism are less likely to remain calm and thus are more likely to make errors. Second, persons who are able to stay calm in a stressful situation should be able to complete their tasks quicker. Third, persons exhibiting higher levels of anxiety are likely to be less confident in the output of their query. This analysis leads to the following three hypotheses:

H1(a): When faced with ambiguity in information requests, users with higher levels of neuroticism will make more semantic errors formulating queries than users with lower levels of neuroticism.

H1(b): When faced with ambiguity in information requests, users with higher levels of neuroticism will take longer formulating queries than users with lower levels of neuroticism.
H1(c): When faced with ambiguity in information requests, users with higher levels of neuroticism will be less confident with the outcome of their queries than users with lower levels of neuroticism.

2.3.2 Extraversion, Ambiguity, and Information Retrieval

Persons with high extraversion scores tend to be more outgoing, high spirited, active, excitement seeking, and cheerful. The relationship between performance and extraversion alters in both significance and direction depending on task and situational variables (Furnham et al., 2002; Hogan and Holland, 2003). The task of composing queries for information requests requires little use of the exuberant traits associated with high levels of extraversion. To perform the task well and to resolve the ambiguity relies on people’s ability to focus on concepts and ideas. Research has indicated that introverts have an advantage in written assessments whereas extraverts typically benefit by oral assessment (Furnham et al., 2002). For this research, individuals are required to compose queries for written tasks, thus tending to favor the introverts. Furthermore, the presence of ambiguity within the task is likely to make an extroverts inability to discuss the ambiguity more frustrating. Thus, because they are required to suppress their enthusiasm to engage the external and sensory aspects of a task and to focus intently on the internal and cognitive aspects of the task, individuals exhibiting high levels of extraversion are likely to find the task of composing a query more difficult and stressful. This increased difficulty is likely to lead to them making more errors and taking more time. Confidence is more problematic. While logic suggests that if a person is experiencing more difficulty, then that person would likely be less confident in the output they produced. Research has shown, however, that higher levels of extraversion lead to overconfidence. Thus, while a direction can
be predicted for accuracy and time, a direction for the confidence hypothesis is not possible and thus H2(c) is stated in the null. This analysis leads to the following three hypotheses:

H2(a): When faced with ambiguity in information requests, users with higher levels of extraversion will make more semantic errors formulating queries than users with lower levels of extraversion.

H2(b): When faced with ambiguity in information requests, users with higher levels of extraversion will take longer formulating queries than users with lower levels of extraversion.

H2(c): When faced with ambiguity in information requests, users with higher levels of extraversion will not differ in their confidence in the output of their queries relative to users with lower levels of extraversion.

2.3.3 Openness, Ambiguity, and Information Retrieval

Recall that traits in the openness to experience dimension reflect the process of using cognition, intelligence, and contemplativeness together with unconventionality (Judge et al., 1999). Individuals with low levels of openness to experience are more conventional and prefer familiar and recognizable situations. Conversely, individuals with higher levels of openness to experience are likely to have greater ability to achieve innovation, to have more positive attitudes towards learning, and to exhibit higher motivation (Barrick and Mount, 1991). These individuals are also more willing to embrace novel ideas “as well as experience emotions more keenly” (Howard and Howard, 1995, pg. 15). Prior research into cognitive style factors that affect database query performance find that persons who rely on intuition as opposed to sensing
composed more accurate queries (Bowen et al., 2003). As noted previously, a positive relationship exists between the openness dimension in the Five Factor Model and the Intuition trait within the MBTI (Furnham et al., 2002).

Determining and extracting the information required from an information system requires creative mappings of real world ideas and concepts to a database structure (Wand and Weber, 1990). Based on prior research, individuals with higher levels of openness are more likely to compose more accurate queries (Furnham et al., 2002; Bowen et al., 2003). Individuals with higher levels of openness to experience exhibit higher levels of flexibility and creativity and, hence, should find the task less daunting. When higher openness individuals are also confronted with excess information and structurally unsound grammar (extraneous and syntactic ambiguity), their greater flexibility and creativity increase the likelihood that such individuals can resolve the ambiguity and perform better than individuals who are less open. That is, individuals who score higher on the openness dimension are likely to be able to produce more accurate queries in less time and be more confident in their output. This analysis leads to the following three hypotheses:

H3(a): When faced with ambiguity in information requests, users with higher levels of openness to experience will make fewer semantic errors formulating queries than users with lower levels of openness to experience.

---

3 Bowen et al., (2003) used the MBTI instrument to measure the dimensions of personality.
4 Recent research into the area of personality research has identified that a possible reason for poor relationship between openness to experience and job performance may be due to the presence of two subfactors affecting performance differently (Griffin and Hesketh, 2004). Openness-to-experience may be comprised of one factor that relates to openness to internal experience and one factor that relates to openness to external experience. Thus, the openness to external factors is likely to be related to the adaptability required for the task of querying. As such, the current research has couched the hypotheses in terms of the longstanding five factor model, but, at the same time, when dealing with the interpretation of the results the possibility of the sixth factor will be taken into consideration.
H3(b): When faced with ambiguity in information requests, users with higher levels of openness to experience will take less time formulating queries than users with lower levels of openness to experience.

H3(c): When faced with ambiguity in information requests, users with higher levels of openness to experience will be more confident with their query output than users with lower levels of openness to experience.

2.3.4 Agreeableness, Ambiguity, and Information Retrieval

Recall that individuals with high levels of agreeableness are compassionate and cooperative whereas individuals with low levels of agreeableness tend to be more non-compliant, critical, sceptical, and competitive (Costa et al., 1991; Judge et al., 1999). The process of query composition requires that essential information in the information request is recognised and that individuals step through the components of each query logically and cautiously. Especially due to their propensity to be critical, sceptical, and competitive, individuals exhibiting lower levels of agreeableness are likely to be better able to recognise, articulate, and evaluate the information necessary to make accurate analyses. Conversely, agreeable individuals with higher levels of straightforwardness, ingenuousness, and modesty are more likely to misinterpret and overlook relevant information (Costa et al., 1991; Judge et al., 1999).

Following the execution of each query, users must evaluate, on an objective and logical basis, the accuracy and relevance of the results generated. The difficulty of these tasks increases with excess information or structural ambiguities. Individuals who are less agreeable, exhibit greater scepticism, and exercise critical thinking skills are more likely to recognise the presence of ambiguities and to be better equipped to resolve them. If the participant’s personality type is better suited to the task, they are likely to exhibit better performance by way of accuracy, take
Pre-print version of article

less time, and be more confidence in their output. This analysis leads to the following three hypotheses:

H4(a): When faced with ambiguity in information requests, users with higher levels of agreeableness make more semantic errors formulating queries than users with lower levels of agreeableness.

H4(b): When faced with ambiguity in information requests, users with higher levels of agreeableness take more time formulating queries than users with lower levels of agreeableness.

H4(c): When faced with ambiguity in information requests, users with higher levels of agreeableness are less confident in their query output than users with lower levels of agreeableness.

2.3.5 Conscientiousness, Ambiguity, and Information Retrieval

The process of composing queries from information requests is iterative. Individuals with higher levels of conscientiousness possess “persistent and achievement oriented” traits (Bryne et al., 2005). When presented with excess information or syntactical ambiguity in an information request, users with higher levels of conscientiousness are more likely to carefully, logically, and persistently work through the request. Furthermore, because they are achievement-oriented and thus more diligent, they are likely to produce more accurate queries and to be more confident in their query results. Their greater diligence is, however, likely to increase the amount of time they spend formulating their queries. This discussion leads to the following hypotheses:

5 Recent research into the area of personality research has identified the possibility of a sixth factor referred to as the honesty-humility factor (Ashton and Lee, 2005). This factor is derived from two of the facets within the agreeableness dimension. Due to the nature of the experimental task forming part of a participant’ assessment the affects of this possible dimension should be minimized. As such, the current research has couched the hypotheses in terms of the long standing five factor model, but at the same time when dealing with the interpretation of the results the possibility of the sixth factor will be taken into consideration.
H5(a): When faced with ambiguity in information requests, users with higher levels of conscientiousness make fewer semantic errors formulating database queries than users with lower levels of conscientiousness.

H5(b): When faced with ambiguity in information requests, users with higher levels of conscientiousness take more time formulating database queries than users with lower levels of conscientiousness.

H5(c): When faced with ambiguity in information requests, users with higher levels of conscientiousness are more confident in their query output than users with lower levels of conscientiousness.

3. METHOD

3.1 Research Design, Participants, and Data Collection

In a laboratory experiment, participants composed and executed queries in SQL for an Oracle database. Seventy-five undergraduate and masters level commerce students participated in the experiment. All participants were familiar with general computing concepts and activities and, prior to the experiment, had received training in developing SQL queries. All participants received a set of instructions containing the scenario, the details of tasks to be performed, the data dictionary, and the entity-relationship diagram (Appendix A). To eliminate potentially different interpretations of non-verbal cues that accompany face-to-face verbal communication (Manusov et al., 1997), ambiguity was manipulated via written information requests. The objective of this research is to study the impacts of personality variables on resolving syntactic, extraneous, and the combination of syntactic and extraneous ambiguity on information retrieval.

---

6 While many users within today’s environment use applications that incorporate a querying by example (QBE) tool the authors have followed the advice by Hayes and Hunton, (2001) who state that “although QBE tools are visual and relatively easy to use, they’re somewhat limited. To create complex queries users must turn to a language call SQL.” They go on to say that “in today’s world it is important to understand the fundamentals of SQL, for it is the basis of all database queries.”
performance. Each information request was designed with four formulations, one clear (no ambiguity) and the remaining three formulations corresponding to each type of ambiguity.

Figure 2 provides an example of an information request, in each of the four formulations, and the corresponding model query. Participants received information requests in each of four possible states: clear, extraneous, syntactic, and both extraneous and clear, i.e., each participant experienced each type of ambiguity.

Four equivalent groups were established using the following technique. The participants were ranked in descending order according to their GPA, i.e., the person with the highest GPA was ranked 1 and the next ranked 2, etc. Participants were assigned to four groups according to their rank, i.e., the highest ranked person to group 1, the second highest to 2, third to group 3, fourth to group 4, fifth to group 4, sixth to group 3, etc. This method of randomization was intended to make the overall ability of the groups as equivalent as possible. The groups were then randomly assigned to a different starting treatment. Thus, participants in each of the four groups experienced each type of ambiguity in the same order (i.e., order was deliberately fixed) but with different starting points. Table 2 shows the order in which the ambiguities were presented to the participants in each group.

Table 2: Information request matrix illustrating type of ambiguity present within the information request

<table>
<thead>
<tr>
<th>Information Request</th>
<th>Group A N=18</th>
<th>Group B N=20</th>
<th>Group C N=21</th>
<th>Group D N=16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Clear</td>
<td>Extraneous</td>
<td>Syntactic</td>
<td>Both</td>
</tr>
<tr>
<td>2</td>
<td>Both</td>
<td>Clear</td>
<td>Extraneous</td>
<td>Syntactic</td>
</tr>
<tr>
<td>3</td>
<td>Syntactic</td>
<td>Both</td>
<td>Clear</td>
<td>Extraneous</td>
</tr>
<tr>
<td>4</td>
<td>Extraneous</td>
<td>Syntactic</td>
<td>Both</td>
<td>Clear</td>
</tr>
<tr>
<td>5</td>
<td>Clear</td>
<td>Extraneous</td>
<td>Syntactic</td>
<td>Both</td>
</tr>
<tr>
<td>6</td>
<td>Both</td>
<td>Clear</td>
<td>Extraneous</td>
<td>Syntactic</td>
</tr>
</tbody>
</table>

1 Recall information retrieval performance is defined within this study as (1) the accuracy (effectiveness) of query formulation, (2) the time required to formulate queries, and (3) users’ confidence in their query formulations.

8 The GPA for each participant was obtained from the university, i.e., not self reported. Due to the possibility that some students were in their first semester of study at the current university, the preferred measure of grade point average for IS/IT courses was not obtainable for all students. Overall GPA was used as the best available alternative.

9 Group was not significantly associated with the number of errors made or the time taken.
Information Request 2

<table>
<thead>
<tr>
<th>Formulation</th>
<th>Information Request</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear</td>
<td>List item number, item name, quantity on hand, and quantity on order for those items where the quantity on hand is greater than 2 times the quantity ordered.</td>
</tr>
<tr>
<td>Syntactic</td>
<td>Management wants a list of inventory items, names, and quantities where the stock levels and the typical amounts ordered are double. (The ambiguity is caused by the use of “and” in the phrase “where the stock levels and the typical amounts ordered are double”. It is unclear as to which amount is to be doubled – the amounts ordered or the stock levels.)</td>
</tr>
<tr>
<td>Extraneous</td>
<td>A recent stocktake of a random sample of inventory items revealed some significant shortages. Provide management with a list containing item number, item name, quantity on hand, and quantity on order for those items where the quantity on hand is greater than 2 times the quantity ordered.</td>
</tr>
<tr>
<td>Syntactic and Extraneous</td>
<td>A recent stocktake of a random sample of inventory items revealed some significant discrepancies. Provide management with a report of inventory items, names, and quantities where the stock levels and the typical amounts ordered are double.</td>
</tr>
</tbody>
</table>

Model Answer

```sql
SELECT item_no, item_name, qty_hand, qty_ordered
FROM inventory
WHERE qty_hand > 2 * qty_ordered;
```

Figure 2: Example of an information request and the four different formulations

The participants had two hours to construct, as accurately as possible, appropriate queries for as many of the twelve information requests as they could (Appendix B). Participants received 5% course credit for participating. Participants were informed that they would be marked on the accuracy of the queries they entered and not merely the number of queries they completed. Because the correct query formulations were generally increasing in complexity, participants were encouraged to do their best on each query before moving to the next information request.

Participants used a UNIX shell script that recorded their entire session. After submitting each query, the grading criteria for the students’ results, not the coding for the statistical analysis, were as follows. The students received a base of 50% of the available 100 points if they produced at least four syntactically correct queries that reasonably addressed the corresponding information requests. Essentially all students received this 50 points. Each completed query was graded on a 0 to 5 scale based on its accuracy. Because of the increasing complexity of the queries, obtaining the same score on each successive query became increasingly challenging.
query attempt, the system displayed the SQL result, i.e., either the rows returned by the query or a syntax error message. Participants could revise their queries as many times as they wished. When they indicated that they were satisfied with the result they obtained for a particular request, participants were prompted to specify their confidence that the query results were correct. After indicating their confidence level, participants proceeded to the next information request.

Note that within this study, the ambiguities contained within the information request could have been resolved by the query developer through the use of the additional information provided e.g., the data dictionary. As such, even though there were ambiguities, e.g., semantic ambiguity, within an information request there is only one possible semantically correct interpretation in the context of this experiment.

3.2 Operationalizing the Variables

The dependent variables were, for each query developed by each participant: the number of semantic errors in the query, the time taken to compose the query, and the participant’s confidence in their query. The number of semantic errors was determined by counting the number of semantic errors in each participant’s last query attempt for each information request. Information requests that were not attempted by a particular participant were not included in the scoring. Furthermore, the final question being attempted at the end of the two hour period may not have been included when it was obvious that the SQL query was not complete. After two individuals independently counted query errors, they cross-checked their error coding sheets for correctness and consistency. When the two coders compared their solutions, the possible outcomes were initial total agreement, one coder being deemed correct, or both coders changing their solution. Given the criteria of making the minimum number changes to reach a semantically correct solution, after re-examining each query and each coder's solution, the coders
were always able to reach agreement on the number of errors, if any, in each participant's query\textsuperscript{11}. The time taken to compose the query for the information request was determined by examining the log files. The dependent variable for the participant’s self-assessed confidence level was entered on the scale: 86-100%, 71-85%, 56-70%, 41-55%, 26-40%, 11-25%, and 0-10%. Participants were asked to determine which of the seven categories they considered best represented their level of confidence. The values were transformed to a seven point scale as follows: ratings of 86-100% transformed to 7, ratings of 71-85% to 6, ratings of 56-70% to 5, ratings of 41-55% to 4, ratings of 26-40% to 3, ratings of 11-25% to 2, and ratings of 0-10% to 1.

### 3.2.1 Independent Variables

The six primary independent variables were ambiguity type and each participant's scores on each of the five dimensions of the NEO PI-R. Ambiguity type was coded as a categorical variable taking on one of four values: clear, extraneous, syntactic, or both extraneous and syntactic. Prior to the experiment all participants completed a Neuroticism, Extraversion, Openness, Personality, Inventory-Revised Model (NEO PI-R) survey to determine their personality types. Each participant completed a self-reported item booklet (Form S) using the hand scoring answer sheet. Due to copyright restrictions, copies of the Form S (completed by participants), the item booklet, and summary sheets are not included in the Appendices.\textsuperscript{12} Each participant's score on...

\textsuperscript{11} To give an indication of the process, the two coders recorded the outcomes for 15% of the queries. The coders were in initial total agreement on 85% of the queries. On the remaining 15%, the more experienced coder was deemed correct on 50%, the less experienced coder was deemed correct on 33.3% of the queries, and both coders changed their error counts on 16.7% of the queries.

\textsuperscript{12} The documents used were obtained via the Australian Council for Educational Resources see http://shop.acer.edu.au/acer-shop/locate?group=RQ
Pre-print version of article

each of the five dimensions of the NEO PI-R was calculated by hand using the procedures outlined in the manual (Costa and McCrae, 1995).

To control for task complexity and each user's query ability, two more variables, query number and grade point average, were used as covariates in the statistical analyzes. Because the model queries that satisfied the 12 information requests became increasingly more challenging, query number was selected as the proxy for query complexity. Alternate complexity measures such as length, difficulty, and effort metrics (Halstead, 1977) were not chosen because they focus almost exclusively on size. The order of the information requests took into account the “challenges” encountered by participants when composing a query. A query containing a sub query or outer join, for example, is likely to be shorter that a query joining multiple tables. Participants, however, often find the shorter query more difficult and thus more challenging.

Because of its greater availability, consistency, comparability, and verifiability, grade point average (GPA), instead of number of IS/IT subjects taken by each participant, was chosen as the proxy for query ability. GPA was obtained from university records.

4. RESULTS

4.1 Summary Results

Table 3, Panel A summarizes the participants’ characteristics. Table 3, Panel B summarizes the performance for each group. Table 3, Panel C summarizes, by ambiguity type, the performance of the participants. These results indicate that query developers’ effectiveness and confidence were affected primarily by syntactic ambiguity but not by extraneous ambiguity. The means for the number of semantic errors and confidence for the clear and extraneous categories are not statistically different (semantic errors p = 0.9992; confidence p = 0.9753). Similarly, the mean number of semantic errors and confidence for the syntactic and both categories are not
statistically different (semantic errors p = 0.2555; confidence p = 0.6486). Thus, the variation in the number of semantic errors and confidence appears to stem primarily from syntactic ambiguity. The mean number of semantic errors and confidence for the extraneous and syntactic categories are statistically different (semantic errors p = 0.0018; confidence p = 0.0540).

The results indicate that syntactic ambiguity alone had only a marginal affect on time. The means for time taken for the clear and extraneous categories are not statistically different (p = 0.3293). The means for time taken for the extraneous and syntactic categories are also not statistically different (p = 0.2797). When, however, information requests contained both syntactic and extraneous ambiguity, query developers took significantly longer than when faced with only one type of ambiguity. The means for time taken for the syntactic and both categories are statistically different (p = 0.0096). One approach people often take to reduce ambiguity in an information request is by including additional explanatory information. This research shows that if the initial request also contained syntactic ambiguity, the inclusion of extra information was not helpful relative to resolving the syntactic ambiguity, i.e., query developers did not make significantly fewer errors.

Table 3 – Summary Results

<table>
<thead>
<tr>
<th>Panel A - Participant Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
</tr>
</tbody>
</table>

25
Grade Point Average (7-point scale, 7 highest)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.94</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>5.08</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>4.90</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>4.86</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Number of males</th>
<th>Number of females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>4</td>
</tr>
</tbody>
</table>

Neuroticism

<table>
<thead>
<tr>
<th>Neuroticism</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>52.22</td>
<td>9.53</td>
</tr>
<tr>
<td></td>
<td>54.70</td>
<td>12.52</td>
</tr>
<tr>
<td></td>
<td>51.19</td>
<td>9.41</td>
</tr>
<tr>
<td></td>
<td>57.50</td>
<td>13.98</td>
</tr>
</tbody>
</table>

Extroversion

<table>
<thead>
<tr>
<th>Extroversion</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.96</td>
<td>10.53</td>
</tr>
</tbody>
</table>

Openness to experience

<table>
<thead>
<tr>
<th>Openness to experience</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50.17</td>
<td>8.54</td>
</tr>
</tbody>
</table>

Agreeableness

<table>
<thead>
<tr>
<th>Agreeableness</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46.28</td>
<td>9.50</td>
</tr>
</tbody>
</table>

Conscientiousness

<table>
<thead>
<tr>
<th>Conscientiousness</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44.11</td>
<td>9.02</td>
</tr>
</tbody>
</table>

Panel B - Participants Summary Performance

<table>
<thead>
<tr>
<th>Group</th>
<th>Semantic Errors/Query</th>
<th>Time Taken/Query</th>
<th>Confidence/Query</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>Standard deviation</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>Group A</td>
<td>3.7153</td>
<td>5.7630</td>
<td>1.3891</td>
</tr>
<tr>
<td></td>
<td>3.4860</td>
<td>5.9647</td>
<td>1.7297</td>
</tr>
<tr>
<td></td>
<td>3.5758</td>
<td>6.4412</td>
<td>1.2698</td>
</tr>
<tr>
<td></td>
<td>4.5952</td>
<td>6.409</td>
<td>1.4303</td>
</tr>
<tr>
<td>Group B</td>
<td>12.6761</td>
<td>7.8827</td>
<td>5.8125</td>
</tr>
<tr>
<td></td>
<td>11.4131</td>
<td>8.1831</td>
<td>5.4693</td>
</tr>
<tr>
<td></td>
<td>12.6934</td>
<td>7.7507</td>
<td>6.0969</td>
</tr>
<tr>
<td></td>
<td>12.6098</td>
<td>6.2585</td>
<td>5.7143</td>
</tr>
<tr>
<td>Group C</td>
<td>10.6579</td>
<td>7.6038</td>
<td>5.9539</td>
</tr>
<tr>
<td></td>
<td>11.4934</td>
<td>6.8310</td>
<td>5.9487</td>
</tr>
<tr>
<td></td>
<td>12.4165</td>
<td>6.9677</td>
<td>5.6234</td>
</tr>
<tr>
<td></td>
<td>14.6475</td>
<td>8.5389</td>
<td>5.5461</td>
</tr>
<tr>
<td>Group D</td>
<td>5.8125</td>
<td>5.9539</td>
<td>1.4111</td>
</tr>
<tr>
<td></td>
<td>5.4693</td>
<td>5.9487</td>
<td>1.2791</td>
</tr>
<tr>
<td></td>
<td>6.0969</td>
<td>6.9677</td>
<td>1.6571</td>
</tr>
<tr>
<td></td>
<td>5.7143</td>
<td>8.5389</td>
<td>1.5604</td>
</tr>
</tbody>
</table>

Panel C - Ambiguity type and Summary Performance

<table>
<thead>
<tr>
<th>Ambiguity type</th>
<th>Clear</th>
<th>Extraneous</th>
<th>Syntactic</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic Errors/Query</td>
<td>2.5263</td>
<td>2.5256</td>
<td>4.6753</td>
<td>5.4605</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>5.2005</td>
<td>5.7434</td>
<td>5.9441</td>
<td>7.0962</td>
</tr>
<tr>
<td>Time Taken/Query</td>
<td>10.6579</td>
<td>11.4934</td>
<td>12.4165</td>
<td>14.6475</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>7.6038</td>
<td>6.8310</td>
<td>6.9677</td>
<td>8.5389</td>
</tr>
<tr>
<td>Confidence/Query</td>
<td>5.9539</td>
<td>5.9487</td>
<td>5.6234</td>
<td>5.5461</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.4111</td>
<td>1.2791</td>
<td>1.6571</td>
<td>1.5604</td>
</tr>
</tbody>
</table>

When group is included as an independent variable, the results of the MANCOVA do not reveal a significant association between group number and the number of errors made and time taken.
Effects of Personality and Ambiguity of Semantic Errors Made By Users During Query Composition

Hypotheses 1(a) through 5(a) predicted that persons possessing various personality characteristics would be more or less successful in formulating queries for information requests containing no, syntactic, extraneous, or both syntactic and extraneous ambiguities. None of the interactions between the five personality dimensions and the four types of query formulations (clear and the three ambiguous) are significant. That is, individuals with various levels of the five different personality dimensions were not significantly more or less successful in resolving ambiguities contained within the information requests. The results of a multivariate analysis of covariance (MANCOVA) (Table 4, Panel A) indicate, however, significant associations between four of the five personality dimensions and number of query errors made. In particular, the results indicate that neuroticism ($F_{1,603}=4.11$, $p=0.0430$, two-tail test), openness ($F_{1,603}=4.75$, $p=0.0297$, two-tail test), and agreeableness ($F_{1,603}=5.23$, $p=0.0226$, two-tail test) significantly affected the number of semantic errors. Conscientiousness has a marginal affect on the number of semantic errors ($F_{1,603}=2.87$, $p=0.0906$, two-tail test).

The parameter estimates for Openness and Agreeableness are in the directions predicted, i.e., persons who exhibited higher levels of openness made fewer errors and persons who were more agreeable made more errors. The directions of the parameter estimates for neuroticism and conscientiousness are opposite to those predicted. For neuroticism, the sign of the parameter estimate indicates that persons with lower levels of neuroticism made more errors. Recall that persons who score low on neuroticism are “calm, even-tempered, and relaxed, and are able to face stressful situations without becoming upset or rattled” (Costa and McCrae, 1992, pg. 15). Low neuroticism individuals in this study made fewer attempts to compose queries and made
more errors. This outcome suggests that the relaxed attitude of lower neuroticism individuals tended to make them assume their queries were correct sooner than higher neuroticism individuals, e.g., as soon as they eliminated all syntax errors.

For conscientiousness, the direction of the parameter estimates indicates that, ceteris paribus, persons with higher levels of conscientiousness made more errors. One possible explanation of this finding is the use of GPA as a covariate. For example, if there are two individuals with the same level of conscientiousness and one has a higher intellect, then the individual with the higher intellect should attain a higher GPA. Equivalently, if two students have the same GPA but different levels of conscientiousness, then the student with the lower level of conscientiousness is likely to possess the higher intellect.\(^\text{14}\)

Post hoc analysis was performed to examine the contribution of each variable toward the number of semantic errors. Seven of the eight independent variables were significant. Of these seven significant variables, complexity and ambiguity, together explained approximately 11.4% of the variance in performance. The five personality variables together explained only approximately 2% of the variance in performance. Thus, while some personality variables did significantly affect the number of errors made, their practical importance is unclear.

<table>
<thead>
<tr>
<th>Source</th>
<th>(R^2)</th>
<th>df</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
<th>Parameter Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0.1501</td>
<td>10</td>
<td>348.99</td>
<td>10.65</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>603</td>
<td>32.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMBIGUITY</td>
<td>3</td>
<td>338.72</td>
<td>10.34</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEUROTICISM</td>
<td>1</td>
<td>134.75</td>
<td>4.11</td>
<td>0.0430</td>
<td>-0.0417</td>
<td></td>
</tr>
<tr>
<td>EXTRAVERSION</td>
<td>1</td>
<td>9.05</td>
<td>0.25</td>
<td>0.6203</td>
<td>-0.0116</td>
<td></td>
</tr>
<tr>
<td>OPENNESS</td>
<td>1</td>
<td>155.52</td>
<td>4.75</td>
<td>0.0297</td>
<td>-0.0545</td>
<td></td>
</tr>
<tr>
<td>AGREEABLENESS</td>
<td>1</td>
<td>171.17</td>
<td>5.23</td>
<td>0.0226</td>
<td>0.0626</td>
<td></td>
</tr>
<tr>
<td>CONSCIENTIOUSNESS</td>
<td>1</td>
<td>94.10</td>
<td>2.87</td>
<td>0.0906</td>
<td>0.0399</td>
<td></td>
</tr>
<tr>
<td>QUERY NUMBER(^\text{1})</td>
<td>1</td>
<td>1957.83</td>
<td>59.77</td>
<td>0.0001</td>
<td>0.6664</td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{14}\) Formally, assume that conscientiousness * intelligence = GPA and that two students have the same GPA, i.e., conscientiousness\(_1\) * intelligence\(_1\) = conscientiousness\(_2\) * intelligence\(_2\) = GPA. For these two students with the same GPA, if student 1 is more conscientious, i.e., if conscientiousness\(_1\) > conscientiousness\(_2\), then intelligence\(_1\) < intelligence\(_2\), i.e., student 1 is less intelligent than student 2 or, equivalently, the less conscientious student is more intelligent than the more conscientious student.
4.2 Effects of Personality and Ambiguity on Time During Query Composition

Hypotheses 1(b) through 5(b) predicted that persons possessing various personality characteristics would be more or less efficient in formulating queries for information requests containing no, syntactic, extraneous, or both syntactic and extraneous ambiguities. None of the interactions between the five personality dimensions and the four types of query formulations (clear and the three ambiguous) were significant. That is, individuals with various levels of the five different personality dimensions did not take significantly more or less time to construct the queries from information requests containing different types of ambiguity. The results of the MANCOVA, reported in Table 4 Panel B, indicate, however, a significant association between one of the five personality dimensions and the time taken to construct the queries. In particular, the results indicate that conscientiousness had a significant effect on the time taken to compose
queries ($F_{1,603}=5.34$, $p=0.0212$, two-tail test). The parameter estimates show that, as predicted, persons exhibiting higher levels of conscientiousness took longer to complete each query. Post hoc analysis was performed to examine the contribution of each variable toward time taken to formulate each query. Four of the eight independent variables, with only one variable related to personality, were significant. Of these four significant variables, the three not related to personality contributed the majority of the $R^2$, i.e., approximately 7%. Thus, while one personality variable, conscientiousness, did significantly affect time taken, its overall contribution to explaining variations in time was minimal (less than 1%).

4.3 Effects of Personality and Ambiguity on User Confidence During Query Composition

Hypotheses 1(c) through 5(c) predicted that persons possessing various personality characteristics would exhibit different levels of confidence in the accuracy of their queries for information requests containing no, syntactic, extraneous, or both syntactic and extraneous ambiguities. None of the interactions between the five personality dimensions and the four types of query formulations (clear and the three ambiguous) were significant. That is, individuals with various levels of the five different personality dimensions were not significantly more or less confident in the queries they produced from information requests containing various types of ambiguity. The results of the MANCOVA, reported in Table 4 Panel C, indicate, however, a significant association between one of the five personality dimensions (neuroticism) and confidence. In particular, the results indicate that neuroticism ($F_{1,603}=9.07$, $p=0.0027$, two-tail test) significantly affected the confidence that users had in the accuracy of their queries. The parameter estimate shows that, as predicted, persons with higher levels of neuroticism were less confident in the accuracy of their queries.
Post hoc analysis was performed to examine the contribution of each variable toward the confidence of the users. Four of the eight independent variables were significant. Of these four significant variables, two contributed the majority of the $R^2$. These two variables were complexity and neuroticism, together explaining 6% of the variation in confidence. Thus, one personality variable (neuroticism) did significantly affect the confidence of the users, however, it only explained approximately 2% of the variation in confidence.

### 4.4 Summary of Results

Table 5 presents a summary of the results. Because none of the interactions between the five personality dimensions and performance were significant, the main effect for each personality trait is depicted in this table.

<table>
<thead>
<tr>
<th>NEO-PI Trait</th>
<th>Performance Measure</th>
<th>Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism $\uparrow$</td>
<td>Accuracy ↓</td>
<td>Significant (opposite predicted)</td>
</tr>
<tr>
<td></td>
<td>Time Taken ↓</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Confidence ↓</td>
<td>Significant</td>
</tr>
<tr>
<td>Extraversion $\uparrow$</td>
<td>Accuracy ↓</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Time taken ↓</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Confidence ↓</td>
<td>Not significant</td>
</tr>
<tr>
<td>Openness $\uparrow$</td>
<td>Accuracy ↑</td>
<td>Significant</td>
</tr>
<tr>
<td></td>
<td>Time taken ↑</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Confidence ↑</td>
<td>Not significant</td>
</tr>
<tr>
<td>Agreeableness $\uparrow$</td>
<td>Accuracy ↓</td>
<td>Significant</td>
</tr>
<tr>
<td></td>
<td>Time taken ↓</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Confidence ↓</td>
<td>Not significant</td>
</tr>
<tr>
<td>Conscientiousness $\uparrow$</td>
<td>Accuracy ↑</td>
<td>Significant (opposite predicted)</td>
</tr>
<tr>
<td></td>
<td>Time taken ↓</td>
<td>Significant</td>
</tr>
<tr>
<td></td>
<td>Confidence ↑</td>
<td>Not significant</td>
</tr>
</tbody>
</table>

### 5. IMPLICATIONS

The results show that various personality dimensions significantly affect users’ abilities to compose accurate queries. Neuroticism, agreeableness, openness to experience, and conscientiousness affected the number of query errors. Conscientiousness affected the length of time taken to compose the queries and neuroticism affected the confidence users had in the
accuracy of their queries. While four of the personality dimensions had a statistically significant effect on the number of query errors, their overall contribution to the explanation of variations in the number of query errors was minimal. The primary factors that contributed to the differences in the number of query errors were ambiguity and complexity. Similar findings were obtained in relation to the time taken. Confidence was the one measure in which one of the personality dimensions, neuroticism, did make a contribution to explaining variation in confidence.

This study set out to determine whether individuals with various levels of the five NEO PI-R personality dimensions were better able to resolve ambiguities. While various personality dimensions did significantly affect query performance, no statistically significant interactions were observed between syntactic or extraneous ambiguities and any of the five personality dimensions. Furthermore, the actual contribution to variations in performance by each personality dimension was usually minimal.

These results have important implications for improving managerial end-user query performance. First, organizations can improve end-user computing by recognizing the impact of personal characteristics on performance (Niederman et al., 1991). Due to the limited size of this impact, however, we do not recommend that organizations radically alter/tailor training programs for users with different personality scores. On the basis of the results of this study, organizations would be better advised to put more effort into training the users to cope with query challenges such as subqueries and joins. Furthermore, when composing information requests, managers and other requestors should examine their requests for syntactic and extraneous ambiguities and resolve such ambiguities prior to query composition. Organizations

---

15 While ambiguity and complexity had a statistically significant effect on the number of query errors, their overall contribution to the explanation of variations in the number of query errors was not exceedingly large. From prior research the effect of complexity has been shown to explain more variance than in this study.
should implement preventive, detective, and corrective procedures to mitigate the possible negative impacts of syntactic and extraneous ambiguities in information requests.

Second, this research has implications for staffing. One of the issues that seems to be making a resurgence in human resource management has been in the area of personality testing as employees are selected on the basis of who best fits the required profile (Toews, 2003; Abernethy, 2005, Van Iddekinge et al., 2005). This research would allow organizations to realise that, while personality may better match organizational needs with appropriate personality types, this impact is likely to be minimal relative to information retrieval. Instead, organizations should ensure that personnel can cope with the challenges of querying and deal with ambiguity. Both of these issues can be mitigated through targeted training programs illustrating to personnel where their “deficiencies” in these areas are and ultimately provide them with mechanisms to mitigate such problems.

6. CONTRIBUTIONS, LIMITATIONS, AND FUTURE RESEARCH

This study makes several contributions to research in the area of human-computer interaction with information repositories. In an examination of the influence of personality characteristics on query performance, this study found statistically significant relationships between personality dimensions and three aspects of query performance (accuracy, time taken, and confidence). These significant relationships, however, contributed minimally to overall variations in performance. This study confirmed that task complexity, ambiguity, and intellectual ability (represented by GPA) significantly affect performance.

The usual caveats associated with laboratory experiments using student participants limit the generalizability of the results. The student participants had, however, received training in information technology (IT) and business-related subjects and, as such, their level of query
proficiency was likely to be typical of managerial users in many organizations. Furthermore, the study only considered the presence or absence of two types of ambiguity. Combinations of various types of other ambiguity may produce different results.

Future research is needed to improve users’ abilities to extract the information they need. First, a more detailed experiment could be conducted to more fully understand the impact of the conscientiousness personality dimension on performance and its interaction with GPA (or similar intellectual measure). Second, a more detailed experiment could be conducted to investigate the effects of varying levels of each of the different types of ambiguities, the possibility of different methods for communicating the information requests/results, and the possibility of a different information retrieval environment (e.g., QBE). Third, future research could examine the impact of different query interfaces to determine what relationships exist between personality, task, and technology. Fourth, experiments could be conducted to examine whether people with particular combinations of the personality dimensions are more effective than people with other combinations. Fifth, the research results could be replicated within a work based environment, e.g., examining the manner in which different personality types detect, communicate, and resolve various ambiguities in an attempt to improve the amount of variance explained by the variables.
REFERENCES

Furnham, A., Chamorro-Premuzic, T. and McDougall, F. 2002. Personality, cognitive ability, and beliefs about
Pre-print version of article

intelligence as predictors of academic performance. *Learning and Individual Differences* (14:1) 47-64.


Toews, B. Matchmaking – Personality profiling key to hiring the right fit for your company, *Motor Truck* (72:4) 32.


Appendix A – ER Diagram

Delivdays
Carrier_code +
City +
State +
Country +
Deliver_days

Carrier
Carrier_code +
Carrier_name
Carrier_type

Customer
Cust_no +
Cust_name
Phone_no
Street
City
State
Post_code
Country
Credit_limit
Outstanding_bal
Pref_carrier_code

Fob
Fob_code +
Fob_name

Employee
Emp_no +
Emp_name

Currency
Currency_code +
Currency_name
Currency_date +
Currency_rate

Invoice
Invoice_no +
Order_date
Cust_no
Ship_date
Want_date
Deliver_date
Paid_date
Fob_code
Discount_pct
Discount_days
Currency_code
Amount_paid
Carrier_code
Emp_no

InvoiceItem
Invoice_no +
Item_no +
Unit_meas
Quoted_unit_price
Agreed_unit_price
Qty_shipped
Qty_accepted
Diff_cause

FK = Currency_code +
[Appropriate Dates]

FK = Invoice_no

FK = Fob_code

FK = Invoice_no

FK = Employee.Emp_no

FK = Customer.Cust_no

FK = Carrier_code

FK = Fob_code

FK = Inventory.Item_no +
Item_name
Item_maker
Item_package
Item_year
Type_of_alc
Alc_category
Alc_content
Avg_unit_cost
Unit_meas
Avg_unit_price
Qty_hand
Qty_ordered

FK = Currency_code

FK = Foreign Key
+ Primary Key
<table>
<thead>
<tr>
<th>No.</th>
<th>Information Request (Clear Formulation)</th>
<th>Model Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>List the item makers of the items we stock. List each item maker only once.</td>
<td>SELECT distinct (item_maker) FROM inventory;</td>
</tr>
<tr>
<td>2</td>
<td>List item number, item name, quantity on hand, and quantity on order for those items where the quantity on hand is greater than 2 times the quantity ordered.</td>
<td>SELECT item_no, item_name, qty_hand, qty_ordered FROM inventory WHERE qty_hand &gt; 2*qty_ordered;</td>
</tr>
<tr>
<td>3</td>
<td>List item number, item name, item maker, item year, and total value, i.e., quantity on hand times average unit cost, where total value is greater than 5000 or item year is less than 1994.</td>
<td>SELECT item_no, item_name, item_maker, item_year, qty_hand*avg_unit_cost FROM inventory WHERE qty_hand * avg_unit_cost &gt; 5000 OR item_year &lt; 1994;</td>
</tr>
<tr>
<td>4</td>
<td>List customer number, customer name, average amount paid, and standard deviation of amount paid. Only include details about customer orders placed after 1 July 2001. Order by average amount paid with highest amount first.</td>
<td>SELECT customer.cust_no, cust_name, avg (amt_paid), stddev (amt_paid) FROM customer, invoice WHERE customer.cust_no = invoice.cust_no AND order_date &gt; '1-Jul-2001' GROUP BY customer.cust_no, cust_name ORDER BY 3 desc;</td>
</tr>
<tr>
<td>5</td>
<td>List customer number, customer name, country, and total number of invoices paid between 1 July 2001 and 30 June 2002. List only those customers having more than five paid invoices.</td>
<td>SELECT customer.cust_no, cust_name, country, count (invoice_no) FROM customer, invoice WHERE customer.cust_no = invoice.cust_no AND paid_date between '1-Jul-2001' and '30-Jun-2002' GROUP BY customer.cust_no, cust_name, country HAVING count (invoice_no) &gt; 5;</td>
</tr>
<tr>
<td>6</td>
<td>List item maker, item number, item name, and the percentage of volume shortages, i.e., 100 times (sum of quantity shipped less sum of quantity accepted)/(sum of quantity shipped). Only include details where wine is the type of alcohol.</td>
<td>SELECT item_maker, inventory.item_no, item_name, 100 * (sum (qty_shipped - qty_accepted) / sum (qty_shipped)) FROM inventory, invoiceitem WHERE inventory.item_no = invoiceitem.item_no AND type_of_alc = 'wine' GROUP BY item_maker, inventory.item_no, item_name;</td>
</tr>
<tr>
<td>7</td>
<td>List customer number, customer name, country, and credit limit of customers located in Japan or of customers with credit limits greater than 15000. In this list include only customers who have placed orders since 1 July 2001.</td>
<td>SELECT customer.cust_no, cust_name, country, credit_limit FROM customer, invoice WHERE customer.cust_no = invoice.cust_no AND order_date &gt; '1-Jul-2001' AND (country = 'Japan' OR credit_limit &gt; 15000);</td>
</tr>
<tr>
<td>8</td>
<td>List invoice number, preferred carrier code, and carrier code where the carrier was not the preferred carrier.</td>
<td>SELECT invoice_no, pref_carrier_code, carrier_code FROM invoice, customer WHERE customer.cust_no = invoice.cust_no AND carrier_code != pref_carrier_code;</td>
</tr>
<tr>
<td>9</td>
<td>List customer number and customer name for all customers, and, if they have ordered anything, a count of unique items ordered.</td>
<td>SELECT customer.cust_no, cust_name, count (distinct (item_no)) FROM customer, invoice, invoiceitem WHERE customer.cust_no = invoice.cust_no (+) AND invoice.invoice_no = invoiceitem.invoice_no (+) GROUP BY customer.cust_no, cust_name;</td>
</tr>
<tr>
<td>No.</td>
<td>Information Request (Clear Formulation)</td>
<td>Model Answer</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 10  | List customer number, name, street, city, state, postcode, and country for customers with credit limits greater than 20000. As part of the same report present the same data for customers who, since 1 July 2001, have total paid invoices of more than 5000. | SELECT customer.cust_no, cust_name, street, city, state, post_code, country 
FROM customer, invoice 
WHERE customer.cust_no = invoice.cust_no 
AND paid_date > '1-Jul-2001' 
GROUP BY customer.cust_no, cust_name, street, city, state, post_code, country 
HAVING sum(amt_paid) > 5000 
UNION 
SELECT customer.cust_no, cust_name, street, city, state, post_code, country 
FROM customer 
WHERE credit_limit >20000;                                                                                                                                                                                                                                                                                                               |
| 11  | List customer number, customer name, number of invoices, and standard deviation of the deliver date minus the want date for customers in Japan. Exclude customers who placed any order between 1 July 2002 and 31 July 2002. | SELECT customer.cust_no, cust_name, count(invoice_no), stddev(deliver_date - want_date) 
FROM customer, invoice 
WHERE customer.cust_no = invoice.cust_no 
AND country = 'Japan' 
AND customer.cust_no NOT IN 
(SELECT cust_no 
FROM invoice 
WHERE order_date between '1-Jul-02' and '31-Jul-02') 
GROUP BY customer.cust_no, cust_name;                                                                                                                                                                                                                                                                                                           |
| 12  | Count the total number of invoices, grouped by country. Next, count the number of late invoices, i.e., where the date delivered was greater than the date wanted. Group by country. List country and the percentage of total orders that were late orders. | CREATE VIEW TotalOrders AS 
SELECT country, count(Invoice_no) Count_Tot_Ord 
FROM customer, invoice 
WHERE customer.cust_no = invoice.cust_no 
GROUP BY country; 

CREATE VIEW LateOrders AS 
SELECT country, count(Invoice_no) Count_Late_Ord 
FROM customer, invoice 
WHERE customer.cust_no = invoice.cust_no 
AND deliver_date > want_date 
GROUP BY country; 

SELECT TotalOrders.country, 100 *(Count_Late_Ord / Count_Tot_Ord) Percent_Late_Orders 
FROM LateOrders, TotalOrders 
WHERE TotalOrders.country = LateOrders.country; |
Appendix C – Error Coding Form

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Question Number</th>
<th>Attempts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Semantic**

*Keywords and Logical Operators*

<table>
<thead>
<tr>
<th>View</th>
<th>Select</th>
<th>From</th>
<th>Where</th>
<th>Group by</th>
<th>Having</th>
<th>Order by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Set Operators*

<table>
<thead>
<tr>
<th>Where</th>
<th>Union</th>
<th>Intersect</th>
<th>Minus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Symbols and Relational Operators*

<table>
<thead>
<tr>
<th>View</th>
<th>Select</th>
<th>From</th>
<th>Where</th>
<th>Group by</th>
<th>Having</th>
<th>Order by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Tables*

<table>
<thead>
<tr>
<th>View</th>
<th>Select</th>
<th>From</th>
<th>Where</th>
<th>Group by</th>
<th>Having</th>
<th>Order by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Attributes*

<table>
<thead>
<tr>
<th>View</th>
<th>Select</th>
<th>From</th>
<th>Where</th>
<th>Group by</th>
<th>Having</th>
<th>Order by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Values*

<table>
<thead>
<tr>
<th>View</th>
<th>Select</th>
<th>From</th>
<th>Where</th>
<th>Group by</th>
<th>Having</th>
<th>Order by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>