Our Problem With Problems: Problem Representations (and Why Truth Matters)

Ronald Dean Balzan

MPhil - Psychology

University of London

Declaration

I hereby declare that the sources of which I have availed myself have been stated in the body of the dissertation and references, and the rest of the work is my own.

_____

Ronald Dean Balzan

Abstract

Problem representation is central to problem solving—solution depends upon sufficiently accurate representation—yet its study has been relatively neglected. A *representation* models one's understanding of a problem: the problematic situation, goal and permissible steps to solution (Newell & Simon, 1972). Most problems are ill defined (Reitman, 1965), yet their study has been relatively neglected too. In their ill definition, ill-defined problems are susceptible to misrepresentation. In eight correlational and experimental internet-based studies and one cross-sectional face-to-face study, this dissertation examined ill-defined problem representations with the aim of bettering them and therefore their solution rates. The focus on most problems, so those of most people, made "average" adults the target, and participation criteria were minimised. Problem representations being unobservable, misrepresentations were inferred. Accuracy, or truth, which is necessary to problem representation, follows from quality argumentation. The integrity of an argument's iterative process of claims and counterclaims, each of which must be supported by genuine evidence, determines that quality. In genuine evidence, correlations exist between antecedents and outcomes. Therefore, correlation misjudgements imply poor evidence, poor arguments, inaccuracies and problem misrepresentations. This dissertation first investigated the frequency of correlation misjudgement and poor argumentation to determine the extent of the problem of problem misrepresentation. To explain the phenomenon, the dissertation then studied errors common to those findings, namely an emphasis on irrelevant factors and an over- and under-emphasis on relevant factors, as well as overconfidence, self-awareness, empathy, awareness of problem representation, age and undergraduate education. In general, the findings suggest participants misrepresent ill-defined problems more than half of the time, and none of

the factors investigated explain this phenomenon. This dissertation proposes that correlation misjudgements and poor argumentation overwhelm these investigations. Until people learn to evaluate correlations and argue better, the impact of other, de facto secondary factors will be undiscoverable.

Table of Contents

List of Tables and Figures

Our Problem With Problems:
Problem Representations (and Why Truth Matters)

> "The formulation of a problem is often more essential than its solution …" (Einstein & Infeld, 1938, p. 92).

1. Introduction

While we often look at what people think, we neglect to study why they think it, and reasoned views are often incorrectly assumed (Kuhn, 1991). We know, for instance, that in the 2016 EU Referendum, 52% of voters voted for the UK to leave the EU, and 48% voted for it to remain. We also know that in the week of June 18-24, 2018, Gallup reported that 41% of those polled approved of the US President's job performance. Yet we know very little about precisely why individuals voted as they did or approved of the President's performance. In their count, however, every view is valued equally, the implicit assumption being that each is sound.

Similarly, while we often study problem solving, we neglect to study the problems being solved. We seemingly either assume that our problems have been *represented*, or modelled, correctly, or we do not think about their representations much, if at all. Regardless, whenever we neglect to adequately model a problem, we end up trying to solve a problem that is essentially different from the one we intended to solve, and the original problem continues, sometimes getting worse. Despite vast amounts of effort and investment, for example, obesity rates are at unprecedented levels in much of the developed world.

In short, problem representation is central to problem solving, and to solve problems is to progress. Barring luck, whenever we misrepresent a problem, it will go unsolved, and progress will elude us—we will move sideways through life at best and sometimes backwards.

Despite the importance of problem representations, however, their study has been relatively neglected, and they are little understood. In addition, of the research that has been conducted, little is recent, as evidenced by many of the citations that follow.

1.1 Problems and Their Representation

What is a *problem*? Duncker's (1945) definition, having stood the test of time, is often cited:

> A problem arises when a living creature has a goal but does not know how this goal is to be reached. Whenever one cannot go from a given situation to the desired situation simply by action, then there has to be recourse to thinking. (p.1)

Nearly 30 years on, Newell and Simon (1972) agreed: "A person is confronted with a problem when he wants something and does not know immediately what series of actions he can perform to get it" (p. 72).

The thinking to which Duncker refers is *problem solving*. Problem solving requires devising a plan of action, typically comprised of a series of decisions, to achieve a goal, and it begins with a problem's representation.

1.1.1 Problem Representations

A *problem representation* is a model that summarises the would-be solver's understanding of a problem's critical components—the problematic situation, the goal, and the permissible steps to solution (Hayes & Simon, 1977; Newell & Simon, 1972; Novick & Bassok, 2005). Say, for instance, that I cannot find the key to my car. My problematic situation is my missing key, my goal is to find it, and permissible steps to solution include a retracing of my steps since I last remember having it.

In *Human Problem Solving*, Newell and Simon (1972) studied problem solvers' step-by-step movements toward their goals with the aim of identifying general problem-solving strategies. Their theory of problem solving has two parts: 1) the formulation of a problem's representation and 2) the problem's solution in light of that representation. Ironically, however, given the dependency of solution on representation, Newell and Simon focus on solution, simply stating, "In any event, the first part of the theory … has not received much attention as yet" (p. 91).

In response, hearkening back to the Gestaltists and their emphasis on the achievement of understanding in problem solving, Greeno (1977) emphasised the importance of representations to problem solving. Indeed, a problem's sufficient representation is a prerequisite for its solution (Goel & Grafman, 2000).

While representations vary in nature (e.g., visuospatial, linguistic, etc.), each is a cognitive translation of external and internal stimuli and situations (Newell, 1990; B. Tversky, 2005). Those translations are subjective by definition—they may be edited or embellished according to the solver's knowledge, purposes and dispositions. By setting implicit boundaries, representations restrict and determine solutions. If a representation is sufficiently inaccurate, it will prevent the solver from solving their problem.

1.1.2 Well-Defined and Ill-Defined Problems

Well-defined problems. In classical accounts of problem solving, every problem is either well defined or ill defined. *Well-defined problems* have clearly defined components and systematic solution processes (Minsky, 1961). In chess, for instance, the problem is well defined by the game's setup together with its objective and rules. The problematic situation is established by the board's organisation in light of the goal of capturing the opponent's king, and the moves permitted to achieve the

goal are prescribed. Essentially, we are given everything we need to solve well-defined problems. Unfortunately, outside of school, university, labs and games, well-defined problems are relatively rare.

Ill-defined problems. Most problems in life, it turns out, are ill defined. In an *ill-defined problem*, at least one critical component is unclear or missing, and solution strategies are largely non-systematic and ad hoc (Reitman, 1965). We need to decide, for instance, whether or not to continue our education, and if so, how and where. Those decisions will depend on many things, including what it is we want to do with our lives, and we just may not know.

Duncker's (1945) definition of a problem, therefore, falls short. Whereas Duncker assumes that we have a goal, problems also arise when our goal, the permissible steps to solution or the problematic situation itself is unclear: ill-defined problems.

Galotti (1989) refers to well-defined problems as formal problems; ill-defined problems, such as finding a job, educating a child, or saving a marriage, are informal, or everyday, problems. Perhaps *everyday* refers to ill-defined problems' prevalence; *informal* may refer to their loose natures. *Formal*, on the other hand, may refer to well-defined problems' structured and self-contained natures. These terms—ill-defined, informal and everyday; well-defined and formal—will be used interchangeably throughout this dissertation.

Calling ill-defined problems everyday problems, however, is not to belittle them. In the words of Fromm (1941):

> Modern man lives under the illusion that he knows what he wants, while he actually wants what he is *supposed* to want. In order to accept this it is necessary to realize that to know what one really wants is not comparatively

> easy, as most people think, but one of the most difficult problems any human being has to solve. It is a task we frantically try to avoid by accepting ready-made goals as though they were our own. (pp. 251-252)

In general, in adequately solving everyday problems, we succeed in navigating life, but lacking in clarity and sometimes authenticity, these problems can be difficult to solve.

Well-defined versus ill-defined problems. Unfortunately, the study of ill-defined problems has been neglected in favour of well-defined problems. Table 1.1 summarises the differences between the two:

Table 1.1

Differences Between Well-Defined and Ill-Defined Problems

Well-Defined	Ill-Defined
All premises are supplied.	Some premises are implicit; some are not supplied.
Problems are self-contained.	Problems are not self-contained.
There is typically one correct answer.	There are typically several possible answers.
Established methods of inference often apply.	Established procedures for solution are rare.
Solution is typically objective.	It is often unclear whether solution is good enough.
Content is often of limited, academic interest.	Problem content typically has personal relevance.
Problems are solved for their own sake.	Problems are often solved in achieving other goals.

Note. Adapted from "Approaches to Studying Formal and Everyday Reasoning," by K. M. Galotti, 1989, *Psychological Bulletin, 105*(3), p. 335. Copyright 1989 by the American Psychological Association.

The relatively clear structure of well-defined problems versus ill-defined problems leads to superior reliability and validity (Linn, 1994). As a consequence, ill-defined problems have attracted relatively little empirical attention, and no clear methodology has evolved for the study of their representation and solution (Garnham & Oakhill, 1994).

Historically, the focus on well-defined problems has been justified under the assumption that knowledge gained from the study of well-defined problems will transfer to advance our understanding of and predict performance regarding ill-defined problems (Galotti, 1989). However, three predominant views of the relationship between formal and informal problem solving, or reasoning, exist. The first is that formal reasoning is a subset of everyday reasoning, formal being easier than informal (Halpern, 2013; Wason & Johnson-Laird, 1972). The second is that the two are different, but share similar processes (Johnson-Laird, 1982). The third view asserts that the two are distinct and have few similarities (Perkins, 1986).

Perhaps with the exception of confirming their difference, knowledge gained from the study of well-defined problems has not informed our understanding of ill-defined problems nor has it enhanced our ability to solve them (Galotti, 1989; Hambleton & Murphy, 1992; Rogers, Maguire, & Leighton, 1998). Therefore, the only way to understand ill-defined problem solving may be to study ill-defined problems (Perkins, 1986).

1.1.3 Ill-Defined Problem Misrepresentation

In their ill definition, ill-defined problems are susceptible to *misrepresentation*. In other words, at least one of the problem's critical components—the problematic situation, goal or permissible steps to solution—may be modelled inaccurately. Ill-defined problems may be misrepresented for two reasons and for either reason or both.

On one hand, a problem may be misrepresented due to the solver's poor representation. Facing an ill-defined problem, the solver invariably faces a sub-problem: having to clarify the problem, which often requires determining the data

relevant to its solution (Galotti, 1989). That clarification attempt, of course, may be unsuccessful.

Let us say, for instance, that I am struggling with my weight, and I have determined that I have gained weight due to a lack of exercise. I have neglected to acknowledge, however, that I let exercise excuse my diet—as long as I exercise, I allow myself to eat whatever I want, whenever I want. In failing to acknowledge my diet's role, I have misrepresented the problem, and all other things being equal, I will fail to lose the weight I had intended.

On the other hand, a problem may be misrepresented due to its nature. A cure for cancer, for example, may have gone undiscovered because researchers have failed in their representation of the problem, but they may not yet have discovered the knowledge necessary to sufficiently represent it.

The literature and we, however, often fail to make a distinction between these two reasons. In failing to do so, we may unjustifiably excuse the continuance of a problem and a lack of progress by attributing it to a problem's nature without first sufficiently examining the quality of the solver's representation.

In observing the poor quality and misdirected nature of cancer research 46 years ago, for instance, Watson (1973) predicted the propagation of well-intentioned, but mediocre work and limited progress. Similarly, in "The Trouble With Brain Science," Marcus (2014) attributed our relatively few answers in part to a lack of agreement regarding the questions being asked. Lastly and more pointedly, we have not become markedly better at solving our ill-defined problems (Galotti, 1989; Hambleton & Murphy, 1992; Rogers et al., 1998). In light of efforts made, is a lack of progress with regard to these issues due to the problems' natures or researchers' poor representations of the problems?

After a reasonable effort has been made to solve a problem without success, but before we can legitimately chalk up the lack of progress to the problem's nature, we must ask ourselves if we are representing a problem to the best of our knowledge. More proactively, if our goal is to solve a problem, it is in our best interest to represent it to the best of our knowledge at the outset, despite fear or inertia. Perhaps too often we are trying to solve the wrong problem, if any problem, due to our own misrepresentation of the problem.

1.2 The Frequency of and Proxies for Problem Misrepresentations

1.2.1 The Frequency of Problem Misrepresentations

While ill-defined problems may be susceptible to misrepresentation, in a sense misrepresentation matters to the extent it occurs. An exclusive focus on frequency, however, implies that all ill-defined problems are of equal importance. The consequences of different ill-defined problems going unsolved, however, make it clear that some are more important than others. Addressing one's smoking habit, for instance, is more important than deciding what to make for dinner one evening.

Proponents of *System 1* and *System 2* thinking—System 1 being relatively effortless and automatic; System 2 being effortful and associated with relatively important decisions (for an overview, see Kahneman, 2011)—may argue that the misrepresentation of System 1 problems, so to speak, is more likely as people pay less attention. Levitt (2016), however, suggests that important problems are misrepresented often too: He found that people may make better decisions regarding important matters by tossing a coin, or leaving matters to chance, as measured by their self-reported happiness six months after their decision to, say, leave a job or end a relationship.

In light of this apparent contradiction, perhaps the frequency of problem misrepresentation should be considered as independent of problem importance, and therefore the study of problem misrepresentation may benefit from an examination of its frequency exclusively.

The frequency with which we observe the continuance, reoccurrence and growth of many ill-defined problems, such as obesity and job dissatisfaction, suggests that people may be poor at solving them. One reason people may be poor at solving ill-defined problems may be that they often misrepresent them. Unfortunately, as problem misrepresentations cannot be observed, their frequency must be inferred.

1.2.2 Proxies for Problem Misrepresentations

Just as problem representation is central to problem solving, an understanding of correlational and causal relationships is central to problem representation. In the words of Buehner and Cheng (2005), "Causation, and only causation, licenses the prediction of consequences of actions" (p. 144). In life, however, causation is often difficult to prove, so correlation becomes a meaningful substitute. To facilitate discussion, when this dissertation refers to correlations, it refers to both causal and non-causal correlational relationships.

Only causation and correlation enable goal-directed behaviour, and the objective in problem solving is always goal achievement. With a correct understanding of causal and correlational relationships, we are able to create plans to achieve goals; without it, we are not.

Causal and correlational relationships also determine evidence quality, and evidence of high quality, or sound evidence, is necessary to accurate, or truthful, problem representations. Correlational or causal relationships between antecedents and outcomes underlie sound evidence.

Accuracy and truth are ascertained through sound argumentation, an iterative process of claims and counterclaims, each supported by sound evidence. In the absence of such evidence, arguments fail to exist—claims and counterclaims are merely opinions, and accuracy and truth are forsaken.

In sum, problem representation is dependent upon our understanding of causal and correlational relationships and the integrity of our arguments. Errors in causal and correlational inference and weaknesses in our arguments, therefore, should correlate with the frequency of problem misrepresentation.

Causal and correlational relationships. Causal and correlational inference studies commonly rely on antecedent (e.g., dieting) and outcome (e.g., weight loss) pairings. Such studies present subjects with some or all of the following four pairings: 1) *both* the antecedent and outcome are *present*, *only* 2) the antecedent *or* 3) outcome is *present*, and 4) *both* the antecedent and outcome are *absent*. Subjects are asked to make an inference regarding a relationship between the antecedent and outcome based on the pairing frequencies provided.

For instance, Kuhn, Phelps, and Walters (1985) used such pairings in their examination of subjects' willingness to infer everyday causal relationships from incomplete data, their hypothesis being that most of the inferences people make in life are causal in nature, and the notion of correlational relationships is largely absent from our thinking. In other words, if an individual makes any inference whatsoever regarding the relationship between an antecedent and outcome, it will be causal in nature.

Indeed, their experiment found that although there were some improvements with age, undergraduate participants, the oldest group studied, remained remarkably willing to infer causality based on minimal and insufficient frequency data as well as

from only correlational and uncorrelated antecedent-outcome data. In addition, no subjects indicated a need for critical information that was missing. In general, Perkins (1985) found that post-primary education has little impact on such reasoning skills.

These findings corroborate those of others (e.g., Kuhn, 1991; Nisbett & Ross, 1980) that suggest adults regularly fail to make correct correlational inferences, even under optimal experimental conditions in which complete and organised frequency data are provided. Rarely, however, is frequency data complete and organised, the implication being that in life we will be poorer at identifying causal and correlational relationships than these studies indicate.

Argument integrity. Billig (1987) proposes that argumentation is central to everyday thinking, problem representation and solving therefore included. In the words of Kuhn (1991):

> In other words, Billig suggests, much of the thinking we do, certainly about issues that are important to us, involves silently arguing with ourselves—formulating and weighing the arguments for and against a course of action, a point of view, or a solution to a problem.…It arises every time a significant decision must be made. Hence, it is at the heart of what we should be interested in and concerned about in examining people's thinking. (pp. 2-3)

Despite the centrality of argumentation to everyday problem representation and solving, however, our arguments likely suffer from weak evidence, a consequence of our poor ability to discern causal and correlational relationships.

In addition, people commonly fail to construct two-sided arguments in informal reasoning (Kuhn, 1991; Perkins, 1985), yet without two sides, an argument does not exist. People may fail to construct two-sided arguments due to unfounded certainty, or overconfidence, in incorrect causal explanations of phenomena, even

when they have relatively low levels of personal experience with a topic, topics about which even experts remain uncertain (Kuhn, 1991). In certainty, the need for argumentation's claim-counterclaim volleying and sound evidence disappears as does the incentive for developing argumentation skills in general.

Together and apart, the findings discussed above as well as life itself, referring to the sheer number of solvable everyday problems that go unsolved, suggest that we may misrepresent our ill-defined problems, or most problems, with significant frequency.

1.3 Factors Impacting Problem Representations and Not

The neglect of problem representations leaves us wanting. Possibly misrepresenting many, if not most, of the problems in our lives, we stand to benefit significantly from greater knowledge of problem representations, including the factors impacting them.

Advances are more likely to occur, however, once we dispel common misconceptions regarding the importance of factors that are believed to matter to everyday problem representation, but actually may matter little, if at all: cognitive intelligence and "emotional intelligence," in its popular form. These factors are in addition to general post-primary education, as discussed above. Then, without distraction, we can attend to factors that may indeed matter, such as background knowledge, domain expertise and context, and encourage investigation with regard to others.

1.3.1 Factors Not Impacting Problem Representations

While a significant amount of research on intelligence has been conducted over the past century, the definition of *intelligence* remains largely in dispute (Sternberg & Kaufman, 2011). Naturally, disagreement regarding intelligence's

definition undermines our ability to measure it. Two researcher's intending to measure intelligence could be measuring two different constructs (Galotti, 1989). As a consequence, our knowledge of intelligence is limited at best. Nevertheless, evidence suggests that intelligence, as typically defined, does not affect ill-defined problem-solving ability.

Cognitive intelligence. When we speak of intelligence, we are usually referring to the cognitive skills employed in well-defined problem solving, not that we realise it. Gardner (1983) cites them as the linguistic and logical-mathematical skills typically emphasised in school and university; Sternberg (1996) refers to them more generally as the skills involved in valid abstract reasoning.

Whatever the case, it is the type of intelligence commonly measured by IQ tests, such as the Wechsler Adult Intelligence Scale (Wechsler, 2008) and Raven's Progressive Matrices (Raven, Court, & Raven, 1993), and often reliably so. The positive correlation of scores amongst such tests is often referred to as *Spearman's* (1904) *g*, a factor often interpreted as a measure of general intelligence.

While IQ test performance helps us predict performance with regard to well-defined problem solving and therefore school and university, the majority of real-world performance does not seem to be accounted for nor predicted by IQ test performance (e.g., Sternberg, Wagner, Williams, & Horvath, 1995; Wigdor & Garner, 1982). In other words, cognitive intelligence appears to be largely unhelpful in representing and solving ill-defined problems.

The claim that post-primary education does not matter when it comes to the representation and solution of our ill-defined problems (Perkins, 1985) also becomes less surprising. Indeed, we may be remiss in emphasising cognitive intelligence's development in general education. As these skills do not seem to significantly benefit

us when it comes to solving most of life's problems, serious questions regarding school and university curricula are begged.

Theory of multiple intelligences. Binet thought intelligence was something more than just cognitive intelligence. One of intelligence's earliest measurers, Binet was reluctant to quantify a child's intelligence in part because he noticed different children might achieve the same total score on his tests, but with a different pattern of correct and incorrect answers (as cited in Mackintosh, 2011). This pattern variance confirmed Binet's belief that intelligence involves somewhat independent abilities, including memory, common sense and imagination.

To Binet, *intelligence* referred to one's ability to successfully navigate the world or, in other words, solve everyday problems. Wechsler (1940) himself, the creator of the first adult intelligence test, noted "individuals with identical IQs may differ very markedly in regard to their effective ability to cope with the environment" (p. 444). Indeed, cognitively intelligent people themselves fail to solve ill-defined problems often enough to suggest something else matters.

In acknowledging that something other than cognitive intelligence matters, yet within a context of intelligence and analogous to Thurstone's (1934) general factor theory, Gardner's (1983) theory of multiple intelligences is perhaps the best known of the pluralist theories. While including the linguistic and logical-mathematical intelligences primarily associated with well-defined problem solving, the explicit aim of the multiple intelligences is the solution of everyday problems.

Broadening intelligence's definition with a set of criteria, Gardner (1999a) posits that an individual possesses at least eight relatively independent intelligences:

- *Linguistic* – Ability to analyse information and create products involving oral and written language
- *Logical-mathematical* – Ability to develop equations and proofs, make calculations and solve abstract problems

- *Spatial* – Ability to recognise and manipulate large-scale and fine-grained spatial images
- *Musical* – Ability to produce, remember and make meaning of different sound patterns
- *Bodily-kinaesthetic* – Ability to use one's own body to create products or solve problems
- *Naturalistic* – Ability to identify and distinguish among different types of plants, animals and weather formations found in the natural world
- *Interpersonal* – Ability to recognise and understand other people's moods, desires, motivations and intentions
- *Intrapersonal* – Ability to recognise and understand one's own mood, desires, motivations and intentions

In accordance with Binet's pattern variance, the relative strength and weakness of each intelligence will vary from person to person, and an individual with a particular strength in one will not necessarily demonstrate a similar strength in another.

Unfortunately, Gardner did not focus on psychometrics in conceptualising his theory, despite listing psychometric strength as a criterion, nor are his criteria for intelligences widely agreed. While others since have investigated the theory from a psychometric perspective, the results have been mixed. An experiment by Visser, Ashton, and Vernon (2006), for example, showed a strong loading of the theory's eight intelligences on Spearman's (1904) *g* as well as inter-correlations amongst them to effectively undermine a multiple intelligences theory.

Chen and Gardner (1997), however, using separate tests for each of the eight intelligences, successfully identified strengths across a number of intelligences in a group of at-risk first graders. The value of identifying a child's strengths versus overlooking them under a narrower definition of intelligence is obvious. Stressing that intelligences can be shaped by our cultures and educations, Chen and Gardner suggest that their study benefits from naturalistic observations—the intelligences cannot be observed in isolation, but only in the performance of tasks in a cultural context.

While a theory of multiple intelligences seeks to expand the definition of intelligence, perhaps correctly and wisely, it does so without advancing clarity (and

therefore measurability) or gaining consensus. Therefore, the greatest value of the theory of multiple intelligences may lie in its challenge to the common conception of intelligence and its suggestion of a definition of intelligence that is consistent with the one Binet intended.

Emotional intelligence. Gardner (1983) did not identify emotional intelligence per se as one of the multiple intelligences, for it failed to meet his criteria. Rather, he denied the possibility of its existence (Gardner, 1999b). Gardner did, however, identify the previously mentioned interpersonal and intrapersonal intelligences. In all likelihood, emotional intelligence and those interpersonal and intrapersonal intelligences are related concepts.

On the whole, it is Salovey and Mayer (1990) who are credited with proposing the first theory of emotional intelligence. Mayer and Salovey (1997) define *emotional intelligence* as "the ability to perceive and express emotion, assimilate emotion in thought, understand and reason with emotion, and regulate emotion in the self and others" (p. 11). Emotional intelligence also refers to one's ability to solve problems having recognised the possible meaning of emotional patterns (Mayer & Salovey, 1997; Salovey & Mayer, 1990). This ability is to be evaluated with regard to how successfully an individual navigates their world (Salovey & Mayer, 1990). In other words, emotional intelligence should matter to ill-defined problem solving.

Daniel Goleman's (1995) popular book, *Emotional Intelligence*, a lay work based on Salovey and Mayer's (1990) theory, introduced emotional intelligence to the general population. At the time of its publication, however, little emotional intelligence research had been conducted (Goleman, 2005), and no means of measuring emotional intelligence existed (Salovey, 2011). Regardless, strong claims were made as to its power (Goleman, 1995). *Time* magazine's (1995) cover claimed

that it "may be the best predictor of success in life, redefining what it means to be smart." Describing the situation, Salovey (2011) states, regrettably, "The genie is out of the bottle, I'm afraid, and stuffing it back in is futile."

Indeed, the heuristic value of emotional intelligence may have undermined its proper study (Barrett & Salovey, 2002). Since 1995, a great deal of emotional intelligence research has been conducted, and several tools that profess to measure emotional intelligence have been developed. Yet important disagreements over emotional intelligence's definition remain.

Once again, if it is unclear as to what emotional intelligence is, it will remain difficult to measure, hence, for instance, the controversy surrounding its measurement in US schools (Zernike, 2016). Angela Duckworth, Christopher H. Browne Distinguished Professor of Psychology at the University of Pennsylvania, resigned from the board of a group overseeing a related project in California citing no reliable way to measure the construct.

In short, emotional intelligence's importance in ill-defined problem solving remains unclear, despite intentions and claims to the contrary.

1.3.2 Factors Impacting Problem Representations

Having highlighted factors often assumed to matter to problem representation that may not, this dissertation now turns to factors that research suggests matter to problem representation (for an overview, see Novick & Bassok, 2005).

Background knowledge and domain expertise. A solver's background knowledge affects whether and to what extent they focus their attention on factors that are relevant to their problem's representation. This knowledge may refer to their experience with analogous problems (e.g., Gick & Holyoak, 1980), general schemas in memory (e.g., Bassok & Holyoak, 1989; Gick & Holyoak, 1983), or domain

expertise (e.g., Chase & Simon, 1973). Galotti (1989) highlighted the importance of this knowledge when she suggested every ill-defined problem's sub-problem of its need for clarification, which often requires determining the data relevant to its solution.

With regard to analogous problems, such problems may be helpful to the solution of another problem only to the extent they are structurally similar; superficially similar problems, such as those that only share themes or objects, will not be helpful (Gick & Holyoak, 1980). Similar structure, however, is a necessary, but insufficient condition: Solver's must *see* the structural analogy and *apply* it to the current problem for it to be useful. Specific analogous problems contain specific content that could become a red herring (Holyoak, 1985), and the likelihood of not seeing or applying a structural analogy increases with attention to solution-irrelevant differences.

In addition to analogous problems, solvers may rely on general schemas in memory. These schemas are abstract in that they include information that is common to multiple problems of a similar type and exclude information that is unique to the individual problems that have been abstracted. Such schemas can be induced by comparing two or more analogous problems or by solving one problem by analogy to another (Bassok & Holyoak, 1989; Gick & Holyoak, 1983).

As a solver's background knowledge affects whether and to what extent they focus their attention on solution-relevant factors, solvers with domain expertise, or experts, may be advantaged in identifying and solving problems that are structurally similar. Experts' representations tend to highlight solution-relevant structural features, whereas novices' representations are more susceptible to highlighting non-structural

features that are irrelevant to solution, such as common objects (Chase & Simon, 1973; de Groot, 1966).

Context. Context also has been suggested to influence a problem's representation. A problem's context may refer to its perception (e.g., Weisberg & Alba, 1981), story content and text (e.g., Hayes & Simon, 1977; Kotovsky, Hayes, & Simon, 1985) or objects (e.g., Duncker, 1945). Even so, context shares the attention risks inherent in the use of specific analogous problems to solve current problems: The solver must discern between the solution-relevant and solution-irrelevant features, then focus on the former and ignore the latter in representing their problem.

With regard to perceptual form, Weisberg and Alba's (1981) nine-dot problem (Maier, 1930) study suggests that relaxing constraints implied by the perceptual form leads to superior solution performance. Object-based inference includes the phenomenon of *functional fixedness*, or when it becomes difficult for people to see an object used for one purpose, perhaps habitually, as having properties that would enable it to be used for a different purpose (Duncker, 1945). Finally, and perhaps unsurprisingly, it has also been suggested that the content of a story and its text may influence how the solver represents a problem by bringing a solver's focus to specific factors (Hayes & Simon, 1977; Kotovsky et al., 1985).

1.4 The Implications of Problem Representation on Solution

Just as a misrepresented problem leaves us solving a problem essentially different from the one we intended to solve, a problem's representation determines one's choice of solution strategy (Novick & Bassok, 2005; Wertheimer, 1961). In other words, one problem represented in two different ways in effect becomes two distinct problems, and each may have a distinct solution strategy. Of course, if a problem has been sufficiently misrepresented, no solution strategy will be effective.

1.4.1 The Limitations of Deductive, Inductive and Abductive Reasoning

Deductive reasoning. Reasoning emphasises the drawing of inferences, or conclusions, from given information, or premises. An inference is *deductive* when given true premises, the structure of the argument guarantees a logically true conclusion (Holyoak & Morrison, 2005). Evans (2005) illustrates deductive reasoning with a syllogism, a valid form of logical argument:

> All C are B.
> No A are B.
> Therefore, no A are C.

Its conclusions also can be logically true, however, even if its premises are false:

> All dogs are animals.
> No cats are animals.
> Therefore, no cats are dogs.

Inductive reasoning. Reasoning is *inductive* when true premises enhance the odds of a conclusion being true, but they do not guarantee it (Holyoak & Morrison, 2005).

Both similarity (e.g., Goldstone & Son, 2005; Quine, 1977) and causality (e.g., Lipton, 1991; Miller, 1987) are suggested to be important to effective inductive reasoning. Induction typically involves the relevant application of similar and causal relations to new stimuli and representations (Sloman & Lagnado, 2005). Such application underscores the importance of background knowledge and domain expertise with regard to similarity and highlights the risks that accompany incorrect assumptions of causality and correlation.

A classic example of inductive reasoning would be to infer that all swans are white after having observed only white swans, many swans indeed being white. As we know, however, swans are sometimes black. Therefore, while swans often are

white, a true premise, the conclusion that all swans are white is not guaranteed, and in this case the conclusion is known to be false.

Abductive reasoning. In the interest of completeness, there is also *abductive* reasoning, which is reasoning from an observation to its most likely cause. While such reasoning may be helpful to narrowing the set of possible causes, abductive reasoning does not permit confirmation of the cause.

1.4.2 Regarding Algorithms, Heuristics and Insight

Two general step-by-step problem-solving processes exist: algorithms and heuristics (see Novick & Bassok, 2005). *Algorithms* guarantee solutions, implying that the premises or components must be true. In other words, algorithms are helpful in solving well-defined problems. *Heuristics* are efficient problem-solving shortcuts that acknowledge limited knowledge, limited time and the limited capacity of working memory—Goldvarg and Johnson-Laird (2001) suggest people represent as little as possible in order to minimize load on working memory. Heuristics promise only that a solution is likely and are often used in solving everyday problems (Evans, 1989; Newell & Simon, 1972; A. Tversky & Kahneman, 1982).

A. Tversky & Kahneman (1974), for example, suggest a person may employ an *availability* heuristic in assessing the risk of a small business failing. Here, the person will determine a frequency and probability of failure based on the ease with which cases of small business failures come to mind. As availability is useful for assessing frequency and probability, that approach may indeed lead to an accurate prediction of failure.

However, certain instances of failure may not be retrieved, and other instances may be unknown to the retriever altogether. Many entrepreneurs, for instance, hire people like themselves rather than people with complementary skills, skills often

crucial to a business's growth and success, and this tendency may not be known by the retriever. In any case, the set for determining frequency and probability may be incomplete, rendering one's frequency and probability assessments incorrect.

Although sometimes compromised by systematic error in the form of biases, the effectiveness of these heuristics is largely assumed (e.g., Gilovich, Griffin, & Kahneman, 2002; Nisbett & Ross, 1980; A. Tversky & Kahneman, 1974). Algorithms and heuristics are reliable, however, only to the extent a problem's representation is sufficiently accurate, or true. Therefore, algorithms and heuristics are secondary problem-solving factors that are dependent upon and subordinate to a problem's representation.

In addition to the strategies that fall into one of these two relatively linear processes, there is the matter of *insight*. For Gestaltists, problem solving with insight focuses on problem representation to the virtual exclusion of any step-by-step process (e.g., Duncker, 1945; Maier, 1930; Novick & Bassok, 2005). Following an initial period of work, the solver reaches an impasse that is followed by a sudden remodelling of the problem after which a solution instantly appears.

More recent findings, however, are inconsistent with the Gestaltist view. Solution instantaneity studies suggest "instant" solutions evolve from cumulative information (Durso, Rea, & Dayton, 1994; Novick & Sherman, 2003). Similarly, impasse research suggests an impasse may be reached if the problem is originally misrepresented. The impasse is overcome through the relaxation of original constraints, effectively leading to a recasting of the problem representation itself (Knoblich, Ohlsson, Haider, & Rhenius, 1999; MacGregor, Ormerod, & Chronicle, 2001). These findings suggest that insight may be similar to algorithms and heuristics, and its instantaneousness may be more perception than reality.

1.4.3 Ill-Defined Problem Solving

Under deductive and inductive reasoning, algorithms and heuristics, the veracity of the representation is all-important. It is only with true premises or components that deductive reasoning and algorithms can be used reliably, and it is only with true premises that inductive reasoning and heuristics can be used effectively, if less reliably.

As mentioned, in well-defined problems, all critical components are present and accurate, or true. In ill-defined problems, at least one critical component is inaccurate or missing, even if unintentionally so, making ill-defined problems susceptible to misrepresentation. With their inaccurate components, misrepresented problems cannot be solved by inductive or deductive reasoning, heuristics or algorithms.

In short, the onus is on the problem solver to ensure a problem's representation is as accurate, or truthful, as possible before attempting to solve an initially ill-defined problem.

1.5 Regarding Ethics and Risk

In studying problem misrepresentation, this dissertation reports a series of investigations involving human participants. The *Code of Ethics and Conduct* (British Psychological Society, 2018) and the *Code of Human Research Ethics* (British Psychological Society, 2014) were adhered to throughout these investigations. A Birkbeck, University of London, Department of Psychological Sciences' *Ethical Approval Form for Research Involving Adults* was approved prior to initiating each investigation contained herein with the exception of the first study, in which case an *Ethical Approval Form for Research Involving Secondary Data Analysis* was

approved. No significant ethical concerns or study risks were identified for any of the studies comprising this dissertation.

In accordance with these codes, in both letter and spirit, the rights and dignity of each participant were respected throughout these studies. The activeness of their participation was stressed, including their right to withdraw from each study at any time without consequence, the exception, again, being the first study, in which no participants were recruited. Solicitation essentially took the form of an information sheet explaining each study's aim and nature (see Appendix A for a sample information sheet).

This information was then presented again to each participant immediately prior to participation. Individuals were reminded of their agreement to participate, their right to withdraw from the study without explanation, and the confidentiality and anonymity of their participation in a voluntary informed consent form to which they agreed (see Appendix B for a sample consent form).

In the interest of anonymity, no identifying details were captured, and each participant received or created a unique identification number. If a participant wished to withdraw from the study post participation, their data could be located only by their providing the principal investigator with their unique identification number. Should a participant wish to withdraw or have any questions or concerns, contact information for the primary investigator and his supervisor was provided.

As stated in the *Code of Ethics and Conduct* (British Psychological Society, 2009), "No code can replace the need for psychologists to use their professional and ethical judgement….Thinking is not optional" (pp. 4-5). Therefore, the aforementioned codes were influencers as well as explicit guides and established parameters. Their internalisation impacted all aspects of this dissertation.

1.6 Dissertation Summary

On one hand, most problems are initially ill defined (Reitman, 1965) and therefore risk misrepresentation. On the other hand, accurate problem representation is prerequisite for solution (Goel & Grafman, 2000), and therein lies the rub: Ill-defined problems are inherently susceptible to misrepresentation, yet their solution depends on their accurate representation.

The sub-problem that Galotti (1989) cites with regard to ill-defined problems, or their need for clarification, however, implies that problems lie on a spectrum with ill-defined problems on one end and well-defined problems on the other. In other words, ill-defined problems are not necessarily eternal, but beginnings rather than ends, and Galotti's sub-problem, in effect, becomes primary.

Therefore, in order to solve an ill-defined problem, the solver's primary goal becomes to sufficiently move its representation from the ill-defined end of the spectrum toward the well-defined end. More specifically, the primary goal is to represent an initially ill-defined problem well enough, meaning that what is essential to its representation is reflected.

Not only does Galotti's (1989) sub-problem suggest this goal, but the problem-solving strategies discussed above require it—deductive and inductive reasoning, algorithms and heuristics are unreliable in the absence of sufficiently accurate components, or in other words, a sufficiently represented problem.

If the primary objective when confronting an ill-defined problem is its sufficient representation, then perhaps the only difference between well-defined and ill-defined problems is their initial representations. Whereas well-defined problem representations are complete and accurate from the start, ill-defined problems are initially inaccurate or incomplete. As no problem is solved until it is sufficiently

represented, the general strategies by which all problems are solved therefore may be the same—there are not distinct ill-defined and well-defined problem-solving strategies, only initially ill-defined and well-defined problems and representations.

Despite its importance, however, we know relatively little about ill-defined problem representation. The study of well-defined problems has failed to enlighten us in this regard.

As argued above and contrary to popular beliefs and sometimes intentions, cognitive intelligence, post-primary education, age and emotional intelligence do not seem to matter to the representation of everyday problems. Background knowledge, domain expertise and context may matter, but everyday problem misrepresentation still may occur with significant frequency.

In light of their prevalence and importance, yet neglect, the need to study ill-defined problems is clear, and they are the focus of this dissertation. To be specific, the focus will be on the everyday problems of average adults that require effortful thinking, the kind Kahneman (2011) attributes to System 2, rather than System 1, or those that are relatively effortless and automatic. The achievement of peace in the Middle East, for instance, while requiring effort, is not the responsibility of an average adult.

While the study of ill-defined problems may be challenging from a psychometric perspective, as highlighted in Table 1.1, the implications of this research are practical and vast and make the challenge one worth accepting. By advancing our understanding of everyday problem representations, the research could lead to progress on virtually every front.

The problem of problem misrepresentation, however, is itself an ill-defined problem. Indeed, the lack of progress to date in our ability to solve ill-defined

problems may suggest the problem has been misrepresented, assuming, that is, we have represented it at all.

This dissertation is, in essence, an examination of the ill-defined problem of everyday problem misrepresentation. As such, we must first study its components for accuracy. The problem's goal is clear: better representation of our ill-defined problems; the problematic situation, however, is less so, and therefore, so are the permissible steps to solution. Once we have represented this problem to the best of our abilities, should it continue, then and only then can we attribute its possible further misrepresentation to its nature.

The endless variety of ill-defined problems may make a general theory of ill-defined problem solving unattainable (Ohlsson, 2012); a general theory of ill-defined problem representation, however, may not be. Given that accurate problem representation is necessary for solution, generalisable insights into better ill-defined problem representation may lead to less wasted effort and enhanced odds of solution—in other words, progress. While this dissertation may stop short of providing a general theory of ill-defined problem representation, it hopes to at least suggest its possibility.

1.6.1 Dissertation Outline

More specifically, this dissertation first attempted to establish the problem of ill-defined problem misrepresentation via its frequency. Its prevalence was approximated indirectly through a series of investigations, all but one as a function of age and years of undergraduate education.

Investigations 1 and 2. First, a misrepresentation is in some critical way inaccurate, perhaps being incomplete; conversely, a representation must be accurate, or truthful. Truths are arrived at through argumentation, and sound argumentation

relies, in part, on sound evidence. Therefore, the frequency of everyday problem misrepresentation was inferred through an examination of the frequency with which subjects failed to create arguments (Investigation 1) and identify sound evidence (Investigation 2).

Investigations 3 and 4. Then, this dissertation sought to confirm these initial frequency findings by examining subjects' understanding of correlational relationships between antecedents and outcomes under both relatively naturalistic (Investigation 3) and optimal (Investigation 4) conditions. An understanding of such relationships determines one's ability to discern sound evidence. In addition, as only causation and correlation enable goal-directed behaviour, it is only with a correct understanding of causal and correlational relationships that we are able to create plans to achieve goals. Therefore, by examining the frequency with which subjects failed to identify such relationships, here, too, we can infer a frequency of everyday problem misrepresentation.

Investigations 5, 6 and 7. Having established a notion of everyday problem misrepresentation frequency, this dissertation next investigated the impact of specific problem representation information on subjects' representations. As problem representation is a topic not typically taught in school or university, most participants were likely unfamiliar with it. It is unreasonable, of course, not to mention unfair, to evaluate someone's competence in something of which they have no prior knowledge and with which they ostensibly have no prior experience. Therefore, this dissertation examined the impact of awareness of the importance of problem representation information specifically (Investigation 5) as well as awareness of the risks associated with emphasis on irrelevant information (Investigation 6) and the over- and underweighting of relevant data (Investigation 7) in ill-defined problem

representation, all perhaps common errors in everyday problem representation as suggested by the findings of this dissertation's third and fourth studies.

Investigations 8 and 9. Once the impact of "problem representation education," so to speak, was evaluated, this dissertation then turned to an investigation of three other factors that may impact the quality of problem representations, the first being overconfidence (Investigation 8), followed by empathy and self-awareness (Investigation 9). In light of Kuhn's (1991) findings in which people commonly fail to construct two-sided arguments in informal reasoning and may do so due to certainty in their incorrect causal explanations of phenomena, overconfidence with regard to problem representation seemed worthy of investigation.

While it seems that something other than cognitive intelligence matters in representing and solving everyday problems, the general construct of intelligence seems to have inhibited our progress in identifying what that something is. There may be multiple intelligences, but in tying themselves to the construct of intelligence, a construct that's poorly defined, their measurement becomes problematic as does understanding the role they in turn may play in everyday problem representation.

A number of researchers have suggested the importance of emotion in good decision-making (e.g., Damasio, 1994; DeSousa, 1987). Indeed, Salovey and Mayer (1990) suggest as much with regard to emotional intelligence. Intuition suggests self-awareness and empathy, often cited as dimensions of emotional intelligence, could play a role in everyday problem representation. *Self-awareness*, or our understanding of how we ourselves feel, and *empathy*, the ability to feel what others may feel, could enhance our understanding of a problem's components, particularly in those problems involving the self and others, respectively.

In light of our limited knowledge of intelligence in general and the lack of cognitive intelligence's role in everyday problem representation specifically, this dissertation sought measures of emotion that are independent of these constructs. Therefore, Investigation 9 defined emotional intelligence as trait emotional intelligence, a personality trait, and studied empathy and self-awareness specifically and accordingly.

1.6.2 Dissertation Hypotheses

This dissertation hypothesised that we frequently misrepresent ill-defined problems. In addition, it hypothesised that knowledge of problem representation and its importance to problem solving as well as of errors common to problem representation will influence the quality of representations. This dissertation also hypothesised that overconfidence negatively correlates with and that levels of empathy and self-awareness positively correlate with the quality of our representations, and with regard to empathy and self-awareness, of problems involving others and the self, respectively. Lastly, this dissertation hypothesised that the performance examined herein would not correlate with participants' age and years of undergraduate education.

2. The Importance of Truth in Problem Representation

When you have consciously modelled problems, how many times have you solved a problem that you have mismodeled? While perhaps self-evident, only sufficiently accurate problem representations permit solution, and while perhaps tautological, a sufficiently accurate problem representation comprises sufficiently accurate components: the problematic situation, goal and permissible steps to solution (Hayes & Simon, 1977; Newell & Simon, 1972; Novick & Bassok, 2005). By definition, a misrepresentation occurs when at least one of those components is inaccurate or missing.

According to the *American Heritage Dictionary* (2016), *truth* is something that conforms to fact or reality, "a statement proven to be or accepted as true"; it defines *true* as something that is "not false or erroneous," something that is "reliable," that is "accurate." In other words, truth is synonymous with accuracy and therefore central to problem representation. The terms truth and accuracy will be used interchangeably throughout the rest of this dissertation.

Despite its centrality to problem representation, however, truth is often ignored. In the 2015 Volkswagen emissions scandal, for instance, company employees programmed emission controls to activate during testing in order to meet regulatory requirements. In real-world driving, emission levels exceeded those requirements. The goal was regulatory compliance, but while the true problematic situation was illegal emission levels, Volkswagen employees misrepresented it as one of software. The emission problem was not solved, of course, because it was not addressed.

We may ignore truth with regard to personal problems too. For example, we may be unhappy in our work, but we may not be willing to acknowledge the reason or

reasons why, or the real problematic situation. Perhaps we simply do not fit in—the organisation's and our colleagues' values, it turns out, are just too different from our own. Rather than acknowledge that truth, however, and with it that leaving is the likely solution, we in effect choose unhappiness—we may lack the courage, prefer to avoid the work or be unwilling to make the sacrifices that change would require, at least for now, at least until our unhappiness becomes too great.

If our goal is to solve problems and progress, then a prevalence of untruths suggests a frequency of problem misrepresentation that is at odds with that goal. Indeed, the prevalence of untruths may be so great that the cover of *Time* (2017) magazine asked, "Is Truth Dead?"

Levine, Park, and McCornack (1999) speak of a "veracity effect," or a tendency to accurately identify truths at greater than chance odds, but accurately detect lies at below chance odds. Similarly, in their meta-analysis of deception judgments, Bond and DePaulo (2006) suggest that we correctly classify truths as nondeceptive more than half of the time (61%), but we also incorrectly classify lies as nondeceptive more than half of the time (53%)—as well as incorrectly classify truths as deceptive 39% of the time.

Therefore, while we often may know truths when we see them, we are less skilled regarding falsehoods. With the ability to represent and solve problems at stake, the nature of truths and untruths and our skill in arriving at and discerning them are worthy of our attention. In light of the above, this chapter first examines what truths are not.

2.1 The Nature of Falsehoods

While the presence of any inaccurate component results in a problem's misrepresentation, the natures of falsehoods vary. In general, however, falsehoods can

be classified as one of two types: intentional and unintentional. It is the relative prevalence of these categories that may be important to furthering our understanding of truth's role in problem misrepresentation.

2.1.1 Intentional Falsehoods

Intentional falsehoods are lies, acts of deception. To *deceive* is to cause someone to believe something that is not true, typically for the perpetrator's personal advantage or gain.

While there are newsworthy deceptions, such as the one involving Volkswagen discussed above, lying seems to be a fact of daily life, and most lies are not newsworthy. Often they are *white lies*, or social lies, viewed as essential to polite interactions (Agosta, Pezzoli, & Sartori, 2013). For example, when asked, someone may tell a waiter that their food is "fine" when they are actually somewhat dissatisfied with it. Most lies also seem to be perceived as little, of little consequence and as not intended to harm others (DePaulo, Kirkendol, Kashy, Wyer, & Epstein, 1996).

DePaulo et al. (1996) suggest that a lie takes place in a third of interactions; Serota, Levine, and Boster (2010) suggest lying occurs in 40% of interactions. The prevalence of lying seems highest amongst teenagers, and its frequency decreases with age and education (Levine, Serota, Carey, & Messer, 2013). Lying also seems to happen less frequently the closer the relationship between the liar and the person to whom they are lying (Serota et al., 2010; Smith, Hancock, Reynolds, & Birnholtz, 2014). While the reliability of much of the research concerning lying needs to be considered due to its self-report nature, a number of findings suggest that most lies are told by a few prolific liars (Halevy, Shalvi, & Verschuere, 2014; Serota & Levine, 2015; Serota et al., 2010).

Therefore, if relatively few people lie, lying occurs in less than half of our interactions, lying decreases with age and education, and many lies are white lies, it may not be lies, or intentional falsehoods, that should be of primary concern when it comes to everyday problem misrepresentation. This conclusion may run counter to many people's expectations as sensational lies may receive a disproportionate amount of media coverage, thereby affecting frequency perceptions.

2.1.2 Unintentional Falsehoods

If lying occurs relatively infrequently, most falsehoods may be unintentional in nature. Therefore, if problem misrepresentation happens frequently, unintentional falsehoods may play a significant role in explaining that frequency. So, the question begged is an important one: How prevalent are unintentional falsehoods?

Unintentional falsehoods are largely of two types: 1) those that are due to *negligence*, or the failure to take proper care in doing something, and 2) those that are simply *mistakes*, or inaccurate judgments or facts. Whereas lies are relatively discernible, unintentional falsehoods are less so due to their indiscrete nature. While a liar is obviously dishonest, a person remains honest despite producing or relying on unintentional falsehoods, hence "honest mistakes." If a few prolific liars tell most lies (Halevy et al., 2014; Serota & Levine, 2015), then most people are perceived as being honest most of the time.

Regardless, as a solver needs to accurately represent their problem in order to solve it, it becomes important to understand how one ascertains truths as well as the average person's ability to do so. A weakness in this ability may lead to unintentional falsehoods and therefore problem misrepresentations. It is through argumentation that truths are arrived.

2.2 The Role of Argumentation in Arriving at and Discerning Truths

2.2.1 Argument

The *American Heritage Dictionary* (2016) defines *argument* as "a course of reasoning aimed at demonstrating the truth or falsehood of something." In 1958, Toulmin suggested the limitations of logic in our thinking and put forward a case for studying how people argue in natural settings.

Since then, various attempts have been made at developing tools to assess quality of argumentation (von Aufschnaiter, Erduran, Osborne, & Simon, 2008). Most of these tools, however, have relied on Toulmin's framework for everyday argumentation, known as Toulmin's Argument Pattern (TAP). As such, the tools typically analyse an argument's content for evidence of TAP's elements. Toulmin (1958) identified four basic argument elements:

1) *Claim* – conclusion whose merits are to be established; a thesis or goal
2) *Data* – facts appealed to in support of a claim; specific and appealed to explicitly
3) *Warrant* – reasons suggested to justify a connection between the data and a claim; general and appealed to implicitly
4) *Backing* – basic assumptions justifying particular warrants; theoretical or historical supporting statements

To help us better understand the basic TAP model, Driver, Newton and Osborne (2000) conveniently summarised the framework in applied form: "Because (data)…since (warrant)…on account of (backing)…therefore (conclusion)" (p. 193).

In addition, Toulmin identified two elements found in more complex arguments:

1) *Rebuttal* **–** specific conditions under which the claim will be false; acknowledges limitations of a claim or supports an opposing thesis or claim
2) *Qualifier* – specific conditions under which a claim can be taken as true

Kuhn (1991) proposes that the incremental complexity of arguments incorporating rebuttals comes from the requirement of having to integrate original and alternative theories to argue the original theory's relative correctness.

In focusing on elements, TAP permits analysis at the individual argument level independent of interactive discourse (Garcia-Mila, Gilabert, Erduran, & Felton, 2013). While TAP can be used to assess the quality of argumentation, however, it cannot be used to make judgments about an argument's correctness. That judgment requires the incorporation of subject knowledge (Driver et al., 2000).

TAP poses challenges in operationalisation, however, chiefly with regard to ambiguities concerning its elements (Duschl, 2008; Kelly, Drucker, & Chen, 1998; Osborne, Erduran, & Simon, 2004). A lack of clarity regarding the element definitions leads to a lack of clarity regarding the differences between them. For instance, few studies seek to identify qualifiers, perhaps suggesting that while indicative of complex arguments, the nuances between the definitions of qualifier, warrant and backing make the identification of qualifiers too difficult.

Regardless, the general structure of an argument is relatively standard (e.g., Kuhn, 1991; Toulmin, 1958):

1) A *claim* is made, or a *causal theory* is proposed, in which a claimant states their belief regarding a matter.
2) The claimant then provides *evidence* in support of their claim. The evidence explains how the claimant knows the cause to be true.
3) Someone, perhaps even the claimant themselves, may then suggest an *alternative theory*. It is what someone might say to counter the original claim.
4) *Counterevidence*, or data that suggests that the original claim may be wrong and that the alternative theory is true, is then provided.
5) A *rebuttal* may be made. It is what the original claimant could say in response to the alternative theory to suggest that the alternative theory is wrong.

And on it can go, an iterative process of claims and counterclaims supported by evidence through which one arrives at truths or more likely their approximations.

The success of that process, however, is a function of its quality. If our arguments are weak, we will arrive at falsehoods, even if unwittingly. If those falsehoods enter into our everyday problem-solving efforts, we will misrepresent and thereby fail to solve our everyday problems.

That quality is determined by 1) adherence to the argumentation process itself, meaning that there are a satisfactory number of claim-counterclaim iterations, and 2) the quality of the evidence introduced in support of each claim and counterclaim. In a sense, the claim is the least important part of an argument, although you have to know it in order to provide evidence in support of it (Kuhn, 1991). In the absence of sound evidence, a claim is merely an opinion.

Populism demonstrates the quality of argumentation's importance. In populism, all claims are assumed to be of equal value, yet therein lies populism's weakness and danger (Bloom, 1988; Kuhn, 1991). While the claims are equal in the sense that all human beings are of equal value, and each claim reflects an individual human being's view, that is where the equality ends. Claims supported by sound evidence are superior to claims that are not, opinions.

2.2.2 Evidence

In light of evidence's importance, it behoves us to parse evidence types. Yet Toulmin (1958) simply equates evidence to data and makes no attempt to distinguish between types. As a consequence, his definition of evidence comes under frequent criticism (e.g., Duschl, 2008; Kelly et al., 1998; Osborne et al., 2004). In the absence of additional criteria and therefore independent of any objective measure of quality, Toulmin's data become only what the arguer perceives evidence to be. In short, the "facts" appealed to in support of a claim become subjective beliefs instead of objective points of validation.

In 1991's *The Skills of Argument*, considered by some to be "a landmark investigation of people's ability to engage in real-world argument" (Hahn, Bluhm, & Zenker, 2017, p. 487), Kuhn addresses Toulmin's shortcomings with regard to evidence. Therefore, Kuhn's definitions of evidence will be relied upon in the investigations that follow. According to Kuhn, there are three general types of evidence:

1) *Genuine evidence* – primarily co-variation evidence of varying degrees of strength, but all suggesting at least a correlational relationship and at best a causal relationship between an antecedent and outcome
2) *Pseudo-evidence* – typically a single case or generalised summary that depicts how the phenomenon might occur (i.e., evidence chosen to illustrate a claim, rather than selected at chance)
3) *Non-evidence* – typically unnecessary and unconnected to the causal claim or citing the outcome as evidence of the antecedent

Genuine evidence is the only type that matters to sound argumentation and therefore problem representation.

To illustrate, in trying to assess a particular diet's effectiveness, the findings of a correlational study of the weight-loss experiences of a group of people on the diet with those of a similar-sized group of people not on the diet would constitute genuine evidence. Just one person's weight-loss experience with the diet, however, would be pseudo-evidence. Unlike genuine evidence, pseudo-evidence does not permit sound correlational inference regarding an antecedent and outcome. Lastly, to cite weight loss as evidence of being on the diet would be non-evidence as it is simply citing the outcome as evidence of the antecedent.

It seems, however, that only a minority of people have an understanding of argumentation and an appreciation of its value. In Kuhn's 1991 study, less than half of the subjects produced genuine evidence in support of causal claims. There was, however, a prevalence of pseudo-evidence and a perception by those who relied on it that it is as powerful as genuine evidence in establishing the correctness of their

causal theories. In addition, as mentioned in Chapter 1, people commonly failed to construct two-sided arguments, and without two sides, an argument does not exist.

This dissertation begins by investigating the strength of our abilities to ascertain and discern truths through argument and therefore our propensity for accurate problem representations. If most of the arguing we do is done silently, within our minds (Billig, 1987), however, we must measure these abilities by proxy.

As such, this chapter first analyses a natural forum for argumentation to assess whether people, unprompted, will attempt to argue, regardless of evidence quality. Then the chapter investigates participants' abilities to discern genuine evidence from other types. If participants fail to argue their claims or discern genuine evidence, they may also fail in their evaluations of others' claims. In other words, participants may be inclined to rely on unreliable sources and falsehoods and end up misrepresenting their problems.

2.3 Investigation 1

In this study, online comments related to three national votes were analysed. The study focused on the presence of argument components only, namely claims and counterclaims supported by evidence; evidence quality was ignored. The primary hypothesis was that participants will fail to construct an argument with significant frequency when given a natural, unprompted opportunity to persuade others. The secondary hypothesis was that there will be a positive association between the incidence of argument and the personal relevance of the topic.

2.3.1 Method

2.3.1.1 Participants

In this naturalistic study, participants comprised those individuals who provided the first 100 comments in response to six online articles (described below),

three from *The New York Times* website and three from *The Guardian* website. Therefore, the sample analysed was 600 comments in total.

As some individuals provided comments more than once in response to each article, the number of participants was lower than 100 per article. Table 2.1 summarises the number of participants that provided one comment versus multiple comments by article (see Section 2.3.1.2 for the article coding key):

Table 2.1

Participants by Number of Comments per Article – Investigation 1

	Article						
	A.1	A.2	B.1	B.2	C.1	C.2	Total
1 Comment	52	27	68	51	60	30	288
> 1 Comment	18	22	12	17	18	22	109
Total	70	49	80	68	78	52	397

While there were 600 comments, there were only 397 unique participants. Of the 397 participants, 288 provided one comment per article; 109 participants provided more than one comment. For those participants providing more than one comment, each commented an average of three times per article, a number consistent across articles and between *The New York Times* and *The Guardian* readers with the exception of article C.1, in which the average was two.

The only criterion for participation was having commented on an article in question. No participant demographic data were captured.

2.3.1.2 Materials and Procedure

One article on each of the following three topics was chosen from *The New York Times* and *The Guardian* websites: the 2016 "Brexit" vote, the 2016 US presidential election, and the 2017 French presidential election. The articles were:

A. 2016 Brexit Vote

1. "'Brexit,' a Feel-Good Vote That Could Sink Britain's Economy," *The New York Times* (Goodman, 2016)
2. "Don't Vote Remain for Us Europeans—Do It for Yourselves," *The Guardian* (Renzi, 2016)

B. 2016 US Presidential Election

1. "Why This Election Terrifies Me," *The New York Times* (Bruni, 2016)
2. "*The Guardian* View on America's Choice: Don't Vote for Trump. Elect Clinton," *The Guardian* (2016)

C. 2017 French Presidential Election

1. "Le Pen and Macron Clash in Vicious Presidential Debate in France," *The New York Times* (Nossiter, 2017)
2. "*The Guardian* View on France's Election: A Win for Macron and Hope," *The Guardian* (2017)

The articles will be referred to as A.1, A.2, B.1, B.2, C.1 and C.2 going forward. As each topic concerned an election or referendum, the articles chosen were published and the comments analysed were made before votes were cast to avoid any potential influence of the election or referendum outcomes on the comments made.

In addition, topics and sources were chosen in order to assess whether the degree of familiarity with an issue and a vested interest in the election or referendum outcome influenced the quality of argumentation. In other words, presumably a predominantly US-interested audience will know more about the US presidential election and care more about its outcome than a non-US-interested audience. As a consequence, *The New York Times*, a US paper, was chosen as a source of articles. Similarly, presumably a predominantly UK-interested audience will know more about the Brexit referendum and care more about its outcome, so *The Guardian*, a UK paper, was chosen. The French presidential election was chosen as a control topic—

given the assumed predominantly US- and UK-interested readerships of *The New York Times* and *The Guardian*, respectively, no bias was expected.

Scoring. As mentioned, each comment was analysed for the presence of argument components only: claims, counterclaims and evidence. While the number of claim-counterclaim iterations and the quality of evidence determines argument quality, argument quality was not considered. The quality of evidence and the ability to discern between genuine, pseudo- and non-evidence, specifically, was the focus of Investigation 2. In this investigation, if a claim or counterclaim was substantiated with *any* evidence, it was classified as an "Argument." If a claim or counterclaim was unsubstantiated or if no claim was made, it was classified as "Opinion."

For example, one of the comments made in response to *The Guardian's* Brexit article, Article A.2, states, "I disagree. There's nothing that can't be fixed." As this claim is unsubstantiated, it is merely an opinion, so it fails to argue. An argument begins, however, when a commentator in response to the same article states, "The answer to [the] 'migration' [problem] is to raise the standard of living in those EU countries which are poorer. Strangely there is not much talk of Irish migrants, which historically made up the largest number of people moving to Britain." In this case, evidence is introduced. Ireland was once a relatively poor country, and at that time, migration from Ireland to the UK was relatively high. Ireland is now a relatively prosperous country, and migration from Ireland to the UK has dropped.

2.3.2 Results

The frequency of Argument versus Opinion in the first 100 comments with regard to each of the six articles is summarised in Table 2.2:

Table 2.2

Frequency of Argument Versus Opinion by Article – Investigation 1

	Article						
	A.1	A.2	B.1	B.2	C.1	C.2	Total
Argument	27	31	15	2	3	3	81
Opinion	73	69	85	98	97	97	519

Given the sample size of 100 comments per article, the frequencies reported by article in Table 2.2 also represent the percentage of comments that try to argue a point versus share an opinion at best.

Ideally, a Pearson's chi-square test or Fisher's exact test would have been performed to assess how often participants argued versus chance and whether there was a positive association between the incidence of argument and the personal relevance of a topic. Unfortunately, as some participants commented more than once with regard to an article (see Table 2.1), the data were not independent in nature, so a condition necessary for performing those tests was not met.

In general, commentators seldom argued, or more specifically offered evidence in support of their claim, doing so in only 81 of 600 comments, or 13.5% of the time. Across topics, *The New York Times* readers offered evidence 15% of the time; *The Guardian* readers did so 12% of the time.

With regard to topic relevance, *The New York Times* readers out-argued *Guardian* readers with regard to the US presidential election by 15:2. *The Guardian* readers, on the other hand, out-argued *The New York Times* readers with regard to the Brexit referendum only marginally, 31:27. Regarding the French presidential election, an equal and small number of *The New York Times* and *The Guardian* readers, three, argued their claims.

2.3.3 Discussion

In the absence of a statistical test, the sheer number of Opinions versus Arguments suggests that participants do not construct an argument when given an opportunity. While readers were not asked to argue per se when commenting in response to an article, it is not unrealistic to assume that persuading others is what those who commented were trying to do; still, some simply may have been venting.

Support for the secondary hypothesis of a positive association between the incidence of argument and the personal relevance of the topic was mixed. Whereas *The New York Times* readers out-argued *Guardian* readers with regard to the US presidential election by 15:2, both groups of readers argued with regard to Brexit almost equally. The low and equal presence of argument in comments regarding the French presidential election seemingly confirmed the topic's control nature.

The value of any study is first and foremost a function of its reliability and validity (Dunbar, 2005). While its value increases in proportion to both properties, research must be reliable for validity to matter (Field, 2009).

Reliability. Given the naturalistic nature of Investigation 1, needless to say, reliability had not been previously assessed. Inter-rater reliability, however, was measured using the scoring methodology described above. The assessment was conducted with a fellow Birkbeck psychology PhD student on a subset of data consisting of the first 60 comments given in response to article B.2, *The Guardian's* 2016 US presidential election article. Cohen's kappa was high at .94, indicating nearly "perfect" agreement.

Validity. With regard to external validity, or the case for the findings' generalisability, population validity may have been jeopardised as *The New York Times* and *The Guardian* readers and those readers who comment on articles, in

particular, may not be representative of the average adult. The study's utterly naturalistic nature, on the other hand, makes its ecological validity relatively certain.

That uncontrolled nature, however, renders its internal validity questionable. Construct and content validities, for instance, are difficult to assert, as commentators were not asked to argue. Predictive and concurrent validities are more likely as the naturalistic nature of the investigation suggests that "what you see is what you get," meaning that we have no reason to assume responses are not reflective of the participants' current understanding of argumentation and therefore are not predictive.

These findings will be discussed further in the General Discussion and Conclusion section.

While this study may have been disadvantaged from a validity perspective by its naturalistic nature, the sheer number of comments, the overwhelming absence of Argument in favour of Opinion and a Cohen's kappa of .94 give weight to these findings nevertheless. In an attempt to address this study's shortcomings and investigate quality of argument with regard to evidence specifically, this dissertation now turns to a more controlled study.

2.4 Investigation 2

Investigation 1 suggests that people offer opinions but fail to argue more often than not. In other words, they make claims but offer no evidence to support them.

Regardless, it is only through arguments of quality that truths are ascertained. As discussed, an argument's quality is a function of 1) the integrity of its iterative claim-counterclaim process and 2) the quality of evidence provided in support of those claims and counterclaims. This chapter's second study investigated participants' ability to discern genuine evidence, or evidence of quality, from pseudo- and non-evidence.

In this study, the primary hypotheses were that participants will identify genuine evidence less than half of the time and no better than chance. Secondarily, it was hypothesised that there would be no association between the ability to identify genuine evidence and either age or years of undergraduate education. Because these secondary hypotheses are null hypotheses, standard NHST (null hypothesis statistical testing) was inappropriate. The results were analysed using Bayesian statistics in order to quantify the extent to which they increase our confidence in the null hypotheses.

2.4.1 Method

2.4.1.1 Participants

Participation criteria were minimised in the hope of capturing a sample representative of the "average" adult. As everyday problems include the problems of living, they are the problems of every person, making average adults a meaningful target. The only criterion for participation was that a participant be at least 17 years of age.

Participant data were collected between 14 and 28 August 2017. Fifty-three participants were recruited online through the Birkbeck, University of London, participant database; from Hanover College's *Psychological Research on the Net* website; and on an ad hoc basis.

Due to seven participants' failure to provide required data, however, data from only 46 participants were used (24 women, 21 men, one other). As summarised in Table 2.3, M_{age} = 43.39 years, SD_{age} = 14.25 years, age range: 17 – 75 years, $M_{undergrad\ years}$ = 3.39 years, $SD_{undergrad\ years}$ = 1.24 years, and undergraduate education range: 0 – 5 years:

Table 2.3

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 2

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Age	46	43.39	14.25	17-NM	17-75	-0.20	-0.61
Undergrad	46	3.39	1.24	0-NM	0-5	-1.69	2.10

Note. NM = not meaningful.

Using the Kolmogorov-Smirnov test, the ages of participants were normally distributed, $D(46) = .12$, $p = .085$; participants' years of undergraduate education, however, were not normally distributed, $D(46) = .38$, $p < .001$. Years of undergraduate education tended to be high and clustered around the mean.

Thirty-two of the 46 participants resided in the US, 12 in the UK, and two elsewhere. Forty-five of the 46 participants cited English as their primary language. Certain Birkbeck undergraduate participants received one credit toward a study participation requirement; otherwise, no compensation was given.

2.4.1.2 Materials and Procedure

While no time limit was enforced, participants committed to approximately 10 minutes of testing time. Data were collected online at a time and place of each participant's choosing.

Each participant received the same two hypothetical problem scenarios in sequential order. The scenarios can be found in Appendix C. In order to minimise systematic variation, the order in which the scenarios were presented was counterbalanced, half of all participants receiving Scenario 1 first and the other half receiving Scenario 2 first. In each scenario a causal claim was made.

Scenario 1 described a situation in which 25% of secondary, or high, school students in the country of Imaginaria, a developed country, were failing to graduate. The average percentage of students that fail to complete secondary school in developed countries is given as 10%. A group of secondary school heads in Imaginaria claim that the reason for its higher-than-average failure rate is a shortage of good teachers. Scenario 2 described a friend who is purchasing a new car. Her car salesperson suggested that she purchase a treatment that the salesperson claims will prevent rust. The friend is on a budget, but she will buy the treatment if she believes it will work.

After each scenario, participants were asked to choose from a list the item or items that served as standalone evidence, meaning in the absence of any other information, in support of the school heads' claim in Scenario 1 and the car salesperson's claim in Scenario 2. If participants believed no item or items served as standalone evidence, they could choose "None of the above," the last item on each list. In an additional attempt to minimise systematic variation, the order of the items on each list, with the exception of "None of the above," was randomised between participants.

In each of the two scenarios, only one item was genuine evidence. In other words, there was only one correct answer; the other five items were incorrect.

In Scenario 1, the correct response read, "In Imaginaria, teaching jobs attract few applicants. Studies show that countries with many applicants per teaching job have better teachers and higher completion rates." This item constituted genuine evidence due to the suggested covariation between number of applicants per teaching job, an antecedent, and the quality of teachers and completion rates, both outcomes. In Scenario 2, the correct response read, "An independent survey found that of 500

people who purchased the treatment, 6 of their cars rusted, and of 500 who did not, 300 rusted." Here, too, covariation is suggested between the treatment, an antecedent, and the occurrence of rust, an outcome.

In addition to "None of the above," each scenario had two list items that constituted pseudo-evidence and two that constituted non-evidence. Each pseudo-evidence item took the form of a single example depicting what could occur. For example, in Scenario 2, one of the pseudo-evidence items read, "Sabi, Joan's best friend, purchased a new car—a convertible—last year, and he bought the treatment. So far, his car has not rusted." As described above, a single occurrence fails to provide the evidence necessary to determine covariation. Purchaser and non-purchaser outcome data from a larger sample is necessary to assess covariation.

The remaining two items on each scenario's list were non-evidence items. In Scenario 1, for example, one of the non-evidence items read, "The percentage of students that fail to complete secondary school in Imaginaria is so high because there aren't enough good teachers in the country." This item simply restates the causal claim for which this item is meant to be evidence.

2.4.2 Results

2.4.2.1 Descriptive Statistics

Correctness of response. Table 2.4 summarises the frequency of correct and incorrect responses by scenario:

Table 2.4

Frequency of Correct and Incorrect Responses by Scenario – Investigation 2

Scenario	Response	*n*
1	Correct	17
	Incorrect	29
2	Correct	23
	Incorrect	23

When a participant selected the genuine evidence item in addition to one or more incorrect items, their response was classified as incorrect as the alternative would have been meaningless—a participant simply could have chosen every item other than "None of the above," or exercise virtually no discernment, and have been considered correct.

The frequency of incorrect responses by evidence type is summarised in Table 2.5:

Table 2.5

Frequency of Incorrect Responses by Evidence Type – Investigation 2

Scenario	Response	*n*
1	Pseudo-evidence	6
	Non-evidence	11
	None of the above	14
2	Pseudo-evidence	7
	Non-evidence	7
	None of the above	11

Of the 29 participants who responded to Scenario 1 incorrectly, six chose pseudo-evidence items, 11 chose non-evidence items, and 14 chose "None of the above." Those numbers total more than 29 as some participants chose more than one evidence type or one evidence type in addition to "None of the above." Of the 23 participants who responded incorrectly to Scenario 2, seven chose pseudo-evidence items, seven chose non-evidence items, and 11 chose "None of the above." As with Scenario 1, these numbers total more than the number of participants that responded incorrectly.

2.4.2.2 Inferential Statistics

Correctness of response—frequency. To assess how often participants cited genuine evidence in support of the causal claims versus chance, a one-sample Pearson's chi-square test was performed. As the correctness of response data were independent in nature, and the expected frequency of each outcome was greater than 5, the assumptions necessary for performing the chi-square analysis were met.

Correctness of response is a categorical variable with two possible outcomes: correct and incorrect. Therefore, assuming each of the six list items in each scenario would be chosen with equal probability, there was a one-out-of-six chance of a participant choosing the correct response (i.e., genuine evidence) and a five-sixths chance of a participant choosing an incorrect response (i.e., pseudo-evidence, non-evidence, "None of the above," or some combination of the six items). As participants were able to choose more than one list item, the single correct item in effect prescribed the odds: If there is a one-out-of-six chance of a participant choosing the correct response, there must be a five-sixths chance of a participant giving an incorrect response. The observed versus expected outcomes for each scenario are summarised in Table 2.6:

Table 2.6

Observed Versus Expected Responses by Scenario – Investigation 2

Scenario	Response	*n*	
		Observed	Expected
1	Correct	17	7.7
	Incorrect	29	38.3
2	Correct	23	7.7
	Incorrect	23	38.3

Analysis of the response frequencies for Scenario 1 indicated that participants chose genuine evidence more often than chance with large effect size ($X^2(1) = 13.63$, $p < .001$; $r = .54$). Cohen (1988, 1992) suggests effect sizes be interpreted as follows: $r = .10$, small effect; $r = .30$, medium effect; and $r = .50$, large effect. Participants also chose genuine evidence more often than chance with large effect size regarding Scenario 2 ($X^2(1) = 36.79$, $p < .001$; $r = .89$). Whereas 37% of participants responded correctly in Scenario 1, 50% responded correctly in Scenario 2.

Correctness of response versus age and years of undergraduate education—correlations. It was secondarily hypothesised that there would be no association between correctness of response and either age or years of undergraduate education. Again, because this hypothesis is a null hypothesis, the results were analysed using Bayesian statistics in order to quantify the extent to which they increase our confidence in the null hypothesis. The results are summarised in Table 2.7:

Table 2.7

Bayesian Pearson Correlations: Correctness of Response Versus Age and Years of Undergraduate Education – Investigation 2

		Correctness of Response	
		Scenario 1	Scenario 2
Age	BF_{10}	0.32	0.43
Undergraduate Years	BF_{10}	0.18	0.20

With regard to Scenario 1, a Bayesian Pearson correlation between age and correctness of response yielded a Bayes factor (BF_{10}) of 0.32, suggesting evidence in moderate support of the null hypothesis of no association (see Lee & Wagenmakers', 2013, classification scheme for interpreting Bayes factors). A Bayes factor of 0.18 also suggests evidence in moderate support of the null hypothesis of no association between correctness of response and years of undergraduate education.

Regarding Scenario 2, a Bayes factor of 0.20 similarly suggests evidence in moderate support of the null hypothesis of no association between correctness of response and years of undergraduate education. Regarding Scenario 1, a Bayes factor of 0.43 suggests evidence in anecdotal support of the null hypothesis of no association between correctness of response and age.

2.4.3 Discussion

In both scenarios, counter to the primary hypothesis, Pearson's chi-square findings suggest that participants discern genuine evidence at *better* than chance for both scenarios and half of the time regarding Scenario 2. While that finding sounds promising, chance can be a misleading benchmark, as discussed in the General Discussion and Conclusion. Indeed, participants failed to identify genuine evidence 63% of the time in Scenario 1 and 56% of the time on average.

With regard to the secondary hypotheses, Bayes factors suggest evidence in moderate-to-anecdotal support of the null hypotheses of no association between correctness of response and either age or years of undergraduate education.

Reliability. As the hypothetical scenarios were created specifically for this study, no pre-existing reliability assessments of them existed. Reliability was assessed, however, by identifying the percentage of participants who responded to both scenarios correctly and both scenarios incorrectly. Of the 17 participants who responded to the first scenario correctly, 10 (59%) responded to the second scenario correctly. Of the 29 who answered the first scenario incorrectly, 16 (55%) responded to the second scenario incorrectly. Therefore, an average of 57% of participants responded consistently between scenarios. As the same group of participants was tested twice within the study, once on each scenario, the concordance of responses between scenarios suggests the study possesses a reasonable level of reliability.

Validity. Whereas Investigation 1 seemingly possessed a reasonable level of ecological validity, but weaker population validity, the opposite seemed to be true for Investigation 2. With regard to ecological validity, the hypothetical and written natures of each of the study's scenarios only approximate what might happen in life. In *The Undoing Project*, his book about Daniel Kahneman and Amos Tversky, Michael Lewis (2016) puts it this way: "People did not choose between things. They chose between descriptions of things" (p. 278). In attempting to sample average adults, however, population validity may be relatively strong.

Overall, internal validity also seems relatively strong. Unlike in Investigation 1, correct and incorrect responses suggest reasonable construct and content validities; like in Investigation 1, criterion validity seems reasonably strong as there is no reason

to suspect the underlying predictive and concurrent validities have been compromised.

These findings, too, will be discussed further in the General Discussion and Conclusion section that follows.

2.5 General Discussion and Conclusion

Summary of results. Consistent with its primary hypothesis, this chapter's first study fails to support the idea that participants will construct an argument when given a natural, unprompted opportunity. This finding echoes those of Kuhn (1991) and Perkins (1985) that suggest people often fail to construct arguments. In this case, *The Guardian* and *The New York Times* commentators made claims but usually offered no evidence in support of them.

In addition, a relationship between topic relevance and argument prevalence is unclear; the study's secondary hypothesis predicted a positive association between the two. While such a relationship existed with regard to the 2016 US presidential election, one did not surface with regard to Brexit. Perhaps online *New York Times* readers are more global than those of *The Guardian*, reflecting *The New York Times's* greater international presence. As a consequence, the average *New York Times* reader may be more likely to take an interest in Brexit than the average *Guardian* reader is to take an interest in the US presidential election.

In this chapter's second study, counter to its primary hypothesis, participants identified genuine evidence in support of a causal claim at better than chance odds with large effect size in its two scenarios. However, participants failed to identify genuine evidence 63% of the time in Scenario 1, 50% of the time in Scenario 2 and 56% of the time on average, or more than half of the time.

In addition, Bayes factors suggest evidence in support of the secondary, null hypotheses of no association between participants' ability to identify genuine evidence and either age or years of undergraduate education. These findings are consistent with those of others that suggest neither age nor post-primary education matter to everyday problem representation and solution (e.g., Perkins, 1985).

Reliability and validity. With a Cohen's kappa of .94 in Investigation 1 and test-retest reliability of 57% in Investigation 2, reasonable levels of reliability were established.

In moving from the naturalistic conditions of Investigation 1 to the more controlled conditions of Investigation 2, external validities were complemented, and internal validities were enhanced. With regard to external validities, in examining a natural occurrence, Investigation 1's findings possess ecological validity; having sought average adult participants, Investigation 2, on the other hand, possesses greater population validity.

With regard to internal validity, correct and incorrect answers suggest relatively strong construct and content validities for Investigation 2. As commentators received no direction in Investigation 1, however, construct and content validities are improbable. Yet both investigations possess criterion validity as we have no reason to assume that the responses in either fail to reflect participants' current understandings of argumentation and evidence and are not predictive.

Implications. As long as our goal is to solve problems and progress, truth will matter, for by definition, no problem can be sufficiently represented and therefore solved without it. In response to *Time*'s (2017) question, truth is not dead; we may, however, have problems arriving at and discerning it, and when those abilities suffer, our abilities to represent and solve problems are impaired. In the words of St.

Augustine, “When regard for truth has been broken down or even slightly weakened, all things will remain doubtful” (as cited in Bok, 1978, p. xv). Like Toulmin’s (1958) data, “truths” become subjective rather than objectively verifiable and are more likely to be falsehoods.

Echoing one of Kuhn’s 1991 findings, Mercier and Sperber (2011) state, “Skilled arguers…are not after the truth, but after arguments supporting their views” (p. 57). Such “arguments,” however, are not arguments at all in a problem representation and solving context.

Boris Johnson, for example, a former Mayor of London and Foreign Secretary and current Member of Parliament, somewhat famously wrote two newspaper articles regarding Brexit. One article attempted to persuade readers that it is in their best interest that the UK leave the EU (2016a), and the other attempted to persuade them that it is in their best interest that the UK remain (2016b). The former article represented his “official” view and was published before the 2016 EU Referendum; the latter was published after the Referendum.

Mr. Johnson seldom argued his claims, however. According to Mr. Johnson, for instance, in his “leave” article, “The fundamental problem remains: that they [the EU] have an ideal that we do not share. They want to create a truly federal union…when most British people do not.” Yet Mr. Johnson provided no genuine evidence in support of these claims. In not arguing his claims and thereby ignoring accuracy, Mr. Johnson merely shared his opinions and failed to make a credible case for leaving or remaining as a solution to any of the UK’s problems.

Unfortunately, perhaps many people, if not most, may not have realised that Mr. Johnson’s views were just opinions and that he was not offering solutions. Just as any misrepresented problem will go unsolved, however, Brexit will not solve any of

the UK's problems to the extent those problems have been misrepresented. Like liars, "skilled arguers" uninterested in truth are shortsighted, for their inevitable failure to solve the problems they misrepresent will eventually expose them.

The better-than-chance performance exhibited in Investigation 2 suggests that people were engaged, and in many matters of everyday problem representation and solution, namely those of relative importance, we seek to outperform chance. It would be inappropriate, for instance, to decide how to raise a child, whom to marry, what job to take or whether the UK should leave the EU, for that matter, with a coin toss. Therefore, Levitt's (2016) suggestion that people may make better decisions regarding important matters by tossing a coin is striking.

In Investigation 2, the identification of genuine evidence at better than chance is encouraging, but not altogether surprising—people were engaged and may have treated the scenarios, albeit hypothetical, as important. The failure to identify genuine evidence 56% of the time, however, should be of concern, for it may mean that we are arguing poorly, neglecting truth and therefore misrepresenting our problems more often than not.

Indeed, the similarities of this chapter's findings to those of Kuhn's 1991 argumentation study are disheartening. It seems little may have changed in the 28 years since her study was published.

For instance, Kuhn observed that her subjects produced genuine evidence less than half (42%) of the time, and that is after, through questioning (see below), being encouraged to provide it. In Investigation 1 herein, 87% of the participants failed to provide any evidence whatsoever, and in Investigation 2, participants correctly cited genuine evidence only 44% of the time. Kuhn's study also suggested that undergraduate education was of little benefit to argumentation; relatedly,

Investigation 2 suggested it had no bearing on one's ability to discern genuine evidence.

While generally supporting Kuhn's conclusions, the studies in this chapter themselves differ from Kuhn's. Kuhn employed discourse analysis to identify claims and evidence in participants' responses to a series of questions regarding three topics. For example, participants were asked, "What causes unemployment?" After a participant offered their theory, they were asked to justify it with evidence, being asked, "How do you know that this is the cause?" Participants were then probed, including being asked specifically for evidence that they would provide to persuade someone that their theory was right.

Investigation 1, while similarly analysed, was utterly naturalistic. Participation was unprompted, readers having voluntarily submitted comments in response to the articles in question. While that approach is advantageous from a generalisability perspective, Investigation 1's internal validity may be weaker as commentators were not asked to make claims or provide evidence in support of them. Investigation 2, on the other hand, was less open-ended. In Investigation 2, participants were asked to identify genuine evidence from a list with regard to a causal claim, and there was only one correct response. As a consequence, Investigation 2's internal validity may be stronger.

The relatively open-ended nature of Kuhn's study as well as its design in general also may explain the prevalence of pseudo-evidence amongst her participants. Whereas Kuhn's participants cited pseudo-evidence in support of their claims more than half of the time, Investigation 2's participants cited pseudo-evidence at least once in each response only 15% of the time. In asking for evidence, as described above, Kuhn was in effect soliciting at least one evidence item, or pseudo-evidence.

This chapter argues that just as problem representation is central to problem solution, truth is central to problem representation. Truths are ascertained through arguments of quality, and that quality is determined by a robust, iterative process of claims and counterclaims, each supported by genuine evidence.

The findings of this chapter's studies suggest that argumentation, let alone argumentation of quality, happens infrequently, and when not ignored, evidence is often misunderstood. In Investigation 1, most participants failed to invoke evidence to move their claims beyond opinion; the findings of Investigation 2 suggest that should participants try, more than half rely on something other than genuine evidence to do so and end up no better off.

Given problem representation's dependency on truth and truth's dependency on quality argumentation, the findings herein suggest that everyday problems may indeed be misrepresented with significant frequency. With an understanding of truth's central importance to problem representation and solution, however, an incentive is created to pursue it through our own arguments and demand it from those of others.

3. The Importance of Correlations to Problem Representation

While poor argumentation is an indicator of problem misrepresentation, in one sense it is secondary. As discussed in the preceding chapters, the quality of an argument is determined in part by the quality of the evidence upon which it relies. A piece of evidence's quality, in turn, is a function of the strength of the correlation between the antecedent and outcome underlying it. Therefore, correlational relationships are primary to problem representations, and errors in correlational inference should correlate with problem misrepresentation. As such, this dissertation investigated the frequency of participants' correlation misjudgements.

3.1 Correlational Relationships

As mentioned in Chapter 1, in addition to determining the quality of evidence, correlational and causal relationships enable goal-directed behaviour, and the objective in problem solving is always the achievement of a goal. As Buehner and Cheng (2005) explain, "Causation, and only causation, licenses the prediction of consequences of actions" (p. 144). In life, however, causation is often hard to prove; as a consequence, correlations become meaningful substitutes. In general, in the words of Crocker (1981):

> Knowing whether events are related, and how strongly they are related, enables individuals to explain the past, control the present, and predict the future. Consequently, people's ability to judge covariations between events is of central importance to a number of psychological theories. (p. 272)

With an understanding of these relationships, we are able to manage evidence, argue truths, and represent and solve problems; without it, we are not.

Despite its importance, however, as well as some views to the contrary (e.g., Mercier & Sperber, 2011), a seemingly overwhelming body of research suggests that our understanding of correlational and causal relationships is poor.

For example, in the Kuhn, Phelps, and Walters (1985) correlation study mentioned in Chapter 1, subjects received frequency data related to one of two structurally similar problems. One problem concerned a product that promised to make a car's engine run better; the other, a product that promised to cause a dishwasher to produce cleaner dishes. Each problem was shared verbally, half of the participants receiving one problem and half receiving the other.

Then participants were tested under one of two conditions: With the antecedent-outcome frequency data presented on cards, half of the subjects were given the data serially, and half were given it all at once. Of those receiving the data serially, participants were given cards in the following order: 1) instances in which the antecedent and outcome were present; 2) instances in which only the antecedent was present; and 3) at the same time, instances in which only the outcome was present and in which neither the antecedent nor outcome was present.

After each of the three tranches of data was received, subjects were asked 1) if there was a relationship between the antecedent and outcome and 2) for the reason for their response. The serial condition was meant to simulate what typically happens in life, where relationships must be inferred from incomplete frequency data. Those subjects who received the data all at once were asked the same two questions, but only once, of course.

Regardless of the condition and based on minimal data, even undergraduate subjects, the oldest of the groups tested, often claimed that correlational relationships between the antecedent and outcome or no relationships between them at all were

causal in nature. In the case of the serial condition, participants often made that claim based on the first frequency data received, or the tranche in which both the antecedent and outcome were present, a bias previously suggested by Schustack and Sternberg (1981). In addition, none of the 129 subjects, recruited from the 4th, 7th and 10th grades (give or take) and undergraduates virtually equally, indicated a need for critical information that was missing.

These findings corroborate those of others (e.g., Nisbett & Ross, 1980) that suggest adults regularly make incorrect causal, rather than correlational, inferences, even under the optimal, unnatural experimental condition in which complete and organised frequency data are provided.

In their reasoning, Schustack and Sternberg's (1981) participants showed themselves to be susceptible to a number of common biases, including base-rate neglect. Accurate base rates are difficult to establish when it comes to many, if not most, everyday problems. Instead, we rely on beliefs regarding possible outcomes based on direct or related experiences, both our own and others', and accurate base rates are neglected in the presence of such specific "evidence" (Bar-Hillel, 1983; Kahneman, Slovic, & Tversky, 1982). In short, prior probabilities are not fully considered, and participants assess the probabilities of possible outcomes incorrectly (A. Tversky & Kahneman, 1974).

In life, of course, rarely is complete frequency data available and organised, so we may be poorer at discerning causal and correlational relationships than some studies indicate. As such, in an attempt to be naturalistic, the first study that follows relied on incomplete and unorganised frequency data. The second study provided participants with complete and organised frequency data.

In sum, the research presented in this chapter investigated the frequency of participant error in assessing correlational relationships to further determine whether the problem of problem misrepresentation was one worthy of our attention and additional research, both within this dissertation and elsewhere.

3.2 General Method

3.2.1 Participants

As in the last study, participation criteria for both studies reported in this chapter were minimised in the hope of capturing a sample representative of the average adult. In fact, the only criterion for participation was that a participant be at least 17 years of age.

Participants were recruited online from the Birkbeck, University of London, participant database; through Hanover College's *Psychological Research on the Net* website; and on an ad hoc basis. Certain Birkbeck undergraduate participants received one credit toward a study participation requirement; otherwise, no compensation was given.

3.2.2 Materials and Procedure

Each of the investigations that follows is a cross-sectional study designed to examine participants' interpretation of antecedent-outcome frequency data in assessing correlational relationships. While no time limit was enforced, participants committed to approximately 10 minutes of testing time. Data were collected online at a time and place of each participant's choosing.

In brief, in each study, each participant received the same written descriptions of hypothetical everyday scenarios involving antecedents (e.g., diet) and outcomes (e.g., weight-loss). In addition, they received either a series of quotes from individuals concerning their experiences with the antecedent and outcome (Investigation 3) or a

summary of groups of individuals' experiences with them (Investigation 4). Then participants were asked whether there was a relationship between the antecedent and outcome. In addition, as each participant was asked to provide a reason for their response, response soundness was examined.

3.3 Investigation 3

This dissertation first investigated the frequency with which the average adult correctly assessed the correlation between an antecedent and outcome under the typical real-life condition of incomplete and unorganised frequency data. The primary hypotheses were that given their assumed poor understanding of correlational relationships, participants will correctly assess the correlations in question less than half of the time and no better than chance. Secondarily, it was hypothesised that there would be no correlation between correctness of response and either age or years of undergraduate education, null hypotheses.

3.3.1 Method

3.3.1.1 Participants

Data for 59 participants were collected between 5 May 2016 and 4 June 2016. Due to nine participants' failure to provide required data, data from only 50 participants were used (28 women, 21 men, one other). As summarised in Table 3.1, M_{age} = 37.54 years, SD_{age} = 14.30 years, age range: 18 – 67, $M_{undergrad\ years}$ = 3.04 years, $SD_{undergrad\ years}$ = 1.26 years, and undergraduate education range: 0 – 4 years:

Table 3.1

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 3

Variable	*N*	*M*	*SD*	Range		Skew	Kurtosis
				Potential	Actual		
Age	50	37.54	14.30	17-NM	18-67	0.25	-1.30
Undergrad	50	3.04	1.26	0-NM	0-4	-1.16	0.19

Note. NM = not meaningful.

In this sample, using the Kolmogorov-Smirnov test, participants' ages were not normally distributed, $D(50) = 0.14$, $p = .014$, and the participants' years of undergraduate education also were not normally distributed, $D(50) = 0.30$, $p < .001$. Whereas participants' ages have a relatively flat distribution, their years of undergraduate education skew high.

Twenty-three of the 50 participants resided in the US, 18 resided in the UK, thee resided in Australia, three resided in France, and one each resided in Spain, Lithuania and Canada. Forty-two of the 50 participants cited English as their primary language.

3.3.1.2 Materials and Procedure

Each participant received the same two scenarios containing descriptions of hypothetical everyday problems in sequential order. The scenarios can be found in Appendix D. In order to minimise systematic variation, the order in which the scenarios were presented was counterbalanced: Of the 50 participants who provided usable data, 24 received Scenario 1 first, and 26 received Scenario 2 first.

Scenario 1 described a situation in which a friend purchases a new car, and the auto dealer recommends the friend also purchase a specific rust-prevention treatment. The friend tells the dealer she will think about it. She then talks to 10 people

regarding their experience with the recommended treatment. All of the people whom she asked had purchased the treatment. The cars of three of those people rusted; the cars of seven of those people did not. In an attempt to make the scenario naturalistic, each of the 10 responses is in the form of a quote. The 10 responses were randomised between participants.

Scenario 2 described a situation in which a colleague is desperate to stop smoking. He has seen an ad for a particular brand of nicotine patch and having tried virtually everything else to stop smoking without success, he is tempted to try it. However, the patches are quite expensive. Before purchasing the patches, he talks to five people regarding their experience with the patch. All of the people whom he asked had used it. Three of those people stopped smoking; two of those people did not. As with Scenario 1, in an attempt to make the scenario naturalistic, each of the five responses is in the form of a quote, and the five responses were randomised between participants.

After each scenario, participants were asked if there is a relationship between the antecedent and outcome and answered with one of three responses: “Yes,” “No” or “Can’t tell.” If they chose “Yes” or “No,” they were asked to give a reason. If they chose “Can’t tell,” they were asked to list the information they would need in order to determine “Yes” or “No.”

As described, each of the 10 quotes provided in Scenario 1 and each of the five quotes provided in Scenario 2 concerned an outcome following a treatment. In other words, only cases in which the antecedent was present were presented, so the frequency data is intentionally incomplete. Cases in which the antecedent was absent would also be needed in order to determine the base rate and hence whether a

correlation existed between the antecedent and outcome. The correct response in each scenario, therefore, is "Can't tell," and the answers of "Yes" and "No" are incorrect.

Soundness of response. Correct answers were further categorised as sound or unsound, and correct but unsound responses were subsequently reclassified as incorrect.

Sound responses acknowledged the need for the frequency data that was missing, or cases in which the antecedent was absent. In other words, participants with sound responses cited that you needed data "from people who didn't buy the [rust-prevention] treatment."

Naturally, unsound responses failed to acknowledge the need for this missing frequency data. Many of them also cited the need for information regarding factors that were initially extraneous. Here, a typical response read that we "need to know the circumstances under which it [the rust-prevention treatment] worked for some cars and not for others—length of ownership, storage conditions, etc."

While storage conditions can have a bearing on the occurrence of rust, participants are wrong to consider their effect *before* studying whether cars that have been treated rusted any less than cars that have not. It is only by looking at the occurrence of rust in cars with *and without* the treatment, or cases in which the antecedent is present *and* cases in which it is absent, that we can tell whether the treatment is likely to work. Storage conditions, in this case, are a secondary factor.

3.3.2 Results

3.3.2.1 Descriptive Statistics

Correctness of response. Table 3.2 summarises the frequency of correct and incorrect responses by scenario:

Table 3.2

Frequency of Correct and Incorrect Responses by Scenario – Investigation 3

Scenario	Response	*n*
1	Can't tell (Correct)	27
	Yes (Incorrect)	8
	No (Incorrect)	15
2	Can't tell (Correct)	22
	Yes (Incorrect)	22
	No (Incorrect)	6

In Scenario 1, 54% of the participants responded correctly; in Scenario 2, 44% responded correctly. In both scenarios, however, a majority of participants chose the correct response of "Can't tell" for an unsound reason. With regard to Scenario 1, 21 of the 27 correct responses were unsound; with regard to Scenario 2, 20 of the 22 correct responses were unsound (see Table 3.3). The majority of correct but unsound responses, 13 of 21 in Scenario 1 and 12 of 20 in Scenario 2, cited the need for information regarding a variety of secondary factors (e.g., the weather, the dealer, geographic location, etc.) in order to determine a relationship.

With regard to the initially incorrect responses, 20 of the 23 incorrect responses in Scenario 1 and 17 of the 28 in Scenario 2 cited the sheer outcome frequencies as the reason for selecting "No" or "Yes," saying, for instance, with regard to Scenario 2, "Most say it didn't help."

3.3.2.2 Inferential Statistics

Correctness of response—frequency. To analyse how often participants failed to assess the correlations in question versus chance, a one-sample Pearson's chi-square test was performed. As the data were independent in nature, and the

expected frequency of each outcome was greater than 5 (see Table 3.3), the assumptions necessary for performing the analysis were met prior to the reclassification of correct but unsound responses as incorrect.

Correctness of response is a categorical variable with three possible outcomes: "Yes," "No," and "Can't tell." Therefore, assuming each response will occur with equal probability, there is a one-third chance of a participant giving a correct response (i.e., "Can't tell") and a two-thirds chance of a participant giving an incorrect response (i.e., "Yes" or "No"). The observed versus expected outcomes both before and after the correct but unsound responses were reclassified as incorrect are summarised in Table 3.3:

Table 3.3

Observed Versus Expected Outcomes by Scenario, Before and After Correct but Unsound Responses Reclassified as Incorrect – Investigation 3

		n	
Scenario	Response	Observed	Expected
1	Correct	27	16.67
	Incorrect	23	33.33
1 – Reclassified	Correct and Sound	6	NM
	Incorrect	44	NM
2	Correct	22	16.67
	Incorrect	28	33.33
2 – Reclassified	Correct and Sound	2	NM
	Incorrect	48	NM

Note. NM = not meaningful.

Analysis of the response frequencies prior to reclassification for Scenario 1 indicates participants chose the correct response of "Can't tell" at greater than chance odds with a medium-to-large effect size ($X^2(1) = 9.65$, $p = .003$; $r = .44$). While a similar qualitative effect was present in the data from Scenario 2, 44% of responses being "Can't tell," the distribution of responses did not differ significantly from chance ($X^2(1) = 2.58$, $p = .132$).

The reclassified data does not permit Pearson's chi-square testing as expected values for correct and sound frequencies cannot be calculated.

Correctness of response versus age and years of undergraduate education—correlations. This study secondarily hypothesised no association between correctness of response and either age or years of undergraduate education, a null hypothesis. Therefore, the results were analysed using Bayesian statistics in order to quantify the extent to which they increase our confidence in the null hypothesis. The results are summarised in Table 3.4:

Table 3.4

Bayesian Pearson Correlations: Scenario Correctness of Response by Age and Years of Undergraduate Education – Investigation 3

		Correctness of Response	
		Scenario 1	Scenario 2
Age	BF_{10}	0.21	0.26
Undergraduate Years	BF_{10}	0.18	0.27

A Bayesian Pearson correlation between age and Scenario 1 correctness of response yielded a Bayes factor (BF_{10}) of 0.21, suggesting evidence in moderate support of no correlation between the variables. A Bayes factor of 0.26 also suggested evidence in

moderate support of the null hypothesis regarding age and Scenario 2 response correctness.

In addition, evidence was found to be in moderate support of the null hypothesis of no correlation between years of undergraduate education and correctness of response for both scenarios. A Bayes factor of 0.18 was calculated regarding Scenario 1, and a Bayes factor of 0.27 was calculated regarding Scenario 2.

3.3.3 Discussion

The primary hypotheses predicted that participants would correctly assess the correlations in question less than half of the time and no better than chance before the reclassification of correct but unsound responses as incorrect. Whereas Scenario 2's results support this hypothesis, Scenario 1's results do not as participants responded correctly more than half of the time and greater than chance. In short, the pre-reclassification results are mixed.

When the soundness of the reasons for participants' choices is considered, however, a clearer picture is painted. Of the 27 who chose the correct response in Scenario 1, only six did so for a sound reason; regarding Scenario 2, only two of the 22 who answered correctly did so. As described above, the majority of correct but unsound responses cited the need for information regarding essentially secondary factors in order to determine correlation; the majority of initially incorrect responses cited the sheer outcome frequencies.

With regard to the secondary hypotheses, for both scenarios evidence was found to be in moderate support of the null hypothesis of no association between correctness of response and either age or years of undergraduate education.

Reliability. As in Investigation 2, the hypothetical scenarios used in Investigation 3 were created for the study, and reliability was assessed by identifying

the percentage of participants who responded consistently between them (i.e., who responded correctly to both scenarios and incorrectly to both scenarios). Fifteen of the 27 participants (56%) who responded to the first scenario correctly responded to the second scenario correctly; of the 23 who responded to the first scenario incorrectly, 16 (70%) responded incorrectly to the second. As the same participants were tested twice within the investigation, once on each scenario, the response concordance suggests that the investigation possesses a reasonable level of test-retest reliability.

Validity. As with the previous study, correct and incorrect responses suggest construct and content validities, and nothing suggests poor criterion validity. Indicators of external validity, however, are mixed. While the naturalistic nature of incomplete and unorganized frequency data suggests some ecological validity, each situation's hypothetical and written nature undermines that validity. In attempting to sample average adults, population validity, on the other hand, seems relatively strong.

These findings will be discussed further in the General Discussion and Conclusion section.

3.4 Investigation 4

Whereas the last study examined participants' ability to discern correlational relationships having been given unorganised and incomplete frequency data, this study investigated their ability to do so under the advantageous circumstance of having complete and summarised frequency data. In a sense, by considering performance under these rather ideal circumstances, this second study sought to investigate the gravity of the problem of problem misrepresentation.

Despite being given the information necessary to evaluate it, however, the primary hypotheses remained that participants will correctly evaluate the association between an antecedent and outcome less than half of the time and no better than

chance. Once again, the secondary hypotheses were that correctness of response will not correlate with either age or years of undergraduate education, null hypotheses.

3.4.1 Method

3.4.1.1 Participants

Data for 114 participants were collected between 26 January 2016 and 26 February 2016. Due to participants' failure to provide required data and incorrect responses to a comprehension question (see below), however, Scenario 1 data from only 69 participants were usable (42 women, 26 men, one other), and the Scenario 2 data from only 67 participants were usable (40 women, 26 men, one other).

As summarised in Table 3.5, in Scenario 1, M_{age} = 35.32 years, SD_{age} = 14.84 years, age range: 17 – 75 years, $M_{undergrad\ years}$ = 2.61 years, $SD_{undergrad\ years}$ = 1.82 years, and undergraduate education range: 0 – 7 years. In Scenario 2, M_{age} = 35.82 years, SD_{age} = 14.80 years, age range: 17 – 75 years, $M_{undergrad\ years}$ = 2.68 years, $SD_{undergrad\ years}$ = 1.79 years, and undergraduate education range: 0 – 7 years:

Table 3.5

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 4

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Scenario 1							
Age	69	35.32	14.84	17-NM	17-75	0.83	-0.31
Undergrad	69	2.61	1.82	0-NM	0-7	0.01	-1.02
Scenario 2							
Age	67	35.82	14.80	17-NM	17-75	0.76	-0.06
Undergrad	67	2.68	1.79	0-NM	0-7	-0.39	-0.96

Note. NM = not meaningful.

Using the Kolmogorov-Smirnov test, regarding Scenario 1, the ages of participants were not normally distributed, $D(69) = 0.16$, $p < .001$, and the participants' years of undergraduate education also were not normally distributed, $D(69) = 0.18$, $p < .001$. Regarding Scenario 2, the ages of participants were not normally distributed, $D(67) = 0.14$, $p = .002$, and the participants' years of undergraduate education also were not normally distributed, $D(67) = 0.20$, $p < .001$. In both scenarios, there were more young participants than old, and years of undergraduate education were relatively unclustered around the mean.

With regard to Scenario 1, 29 of the 69 participants resided in the UK, 24 resided in the US, and 16 resided elsewhere; 62 of the 69 participants cited English as their primary language. With regard to Scenario 2, 29 of the 67 participants resided in the UK, 25 resided in the US, and 13 resided elsewhere; 56 of the 67 participants cited English as their primary language.

3.4.1.2 Materials and Procedure

Each participant received the same two scenarios containing descriptions of hypothetical everyday problems in sequential order. The scenarios can be found in Appendix E. In order to minimise systematic variation, the order in which the scenarios were presented was counterbalanced: Of the Scenario 1 usable participants, 36 received Scenario 1 first, and 33 received Scenario 2 first; of the Scenario 2 usable participants, 35 received Scenario 1 first, and 32 received Scenario 2 first.

Scenario 1 described a situation in which a friend went to the doctor for a routine physical examination, and the doctor recommended that the friend lose 20 pounds. The friend had heard of one diet in particular, having heard people talk about it at the office and having seen it mentioned on television, in magazines and in newspapers. He talked to 12 people regarding their experience with the diet, only

some of whom had tried it. The responses of the 12 people surveyed were given to the participants in the following summary form:

> Six people with whom he talked said they were indeed on the diet, and they all said they were losing weight. Two people said that they were on the diet, but they were not losing weight. Three people said they were not on that diet, but they were losing weight. One person said they were not on the diet, and they were not losing weight.

The order of the sentences was randomised between participants.

In Scenario 2, a friend started losing his hair at an alarming rate, especially for someone so young, and he had seen an ad for an over-the-counter hair-loss treatment. The ad said that the treatment not only stopped additional hair loss but led to hair regrowth. Before deciding to use it himself, he talked to 24 people to see what their experience with the treatment had been. The responses of the 24 people surveyed were given to the participants in the following summary form:

> Twelve people with whom he talked said they were indeed using the treatment, and they all said their hair was re-growing. My friend spoke to four people who said that they were using the treatment, but their hair was not re-growing. Six people said they were not using the treatment, but their hair was re-growing. Two people said they were not using the treatment, and their hair was not re-growing.

As with Scenario 1, the sentences above were randomised between participants.

After a scenario, each participant was asked to first confirm the number of people with whom the friend spoke as a means of testing participant attention and comprehension. Then, each participant was asked if the diet or treatment had anything to do with whether these people were losing weight or had stopped losing or were re-

growing hair, having to choose one of three responses: “Yes,” “No” or “Can’t tell.” Finally, participants were asked to give the reason for their response.

In each of the two scenarios, only some of the 12 people surveyed in Scenario 1 had used the diet, and only some of the 24 people surveyed in Scenario 2 had used the hair-loss treatment. In other words, cases in which the antecedent was present and cases in which the antecedent was absent were presented. Similarly, in some cases, the outcome (i.e., weight loss or stoppage of hair loss or hair regrowth) was present, and in other cases, it was absent. Table 3.6 summarises the frequencies of diet use and non-use as well as success and failure regarding weight loss in Scenario 1 and the frequencies of treatment use and non-use as well as success and failure regarding hair loss and regrowth in Scenario 2:

Table 3.6

Frequency of Antecedents and Outcomes by Scenario – Investigation 4

			Outcome	
			Weight Loss	No Weight Loss
Scenario 1	Antecedent	Diet	6	2
		No Diet	3	1
			Outcome	
			Hair Loss	No Hair Loss
Scenario 2	Antecedent	Treatment	12	4
		No Treatment	6	2

As the success ratios under usage and non-usage are equal in each scenario (i.e., in Scenario 1, 6:2 and 3:1, respectively; in Scenario 2, 4:12 and 2:6, respectively), there

is no relationship between each antecedent and outcome. Therefore, the correct response in each scenario is "No," and the answers of "Yes" and "Can't tell" are incorrect.

Soundness of response. As in the previous study, correct answers were then categorised as sound and unsound. Sound responses explained how the frequency data permitted the assessment of covariation. For example, a typical sound response read, "The same proportion that are on the diet and are not on the diet are losing weight.…As no more people are losing weight, this diet can't be said to help people lose weight."

Most of the correct but unsound responses occurred with regard to the second scenario, concerning the effectiveness of a hair loss and regrowth treatment. Here, a typical response read, "The treatment is for re-growing, not losing hair. The loss of hair is not due to the treatment." In this case, several participants seem to have been confused with regard to the treatment's purpose. These responses will be discussed in the Post-Reclassification section in the General Discussion and Conclusion.

3.4.2 Results

3.4.2.1 Descriptive Statistics

Correctness of response. Table 3.7 summarises the frequency of correct and incorrect responses by scenario:

Table 3.7

Frequency of Correct and Incorrect Responses by Scenario – Investigation 4

Scenario	Response	*n*
1	No (Correct)	9
	Yes (Incorrect)	14
	Can't tell (Incorrect)	46
2	No (Correct)	28
	Yes (Incorrect)	12
	Can't tell (Incorrect)	27

With regard to Scenario 1, 13% of the 69 participants responded correctly; with regard to Scenario 2, 42% of 67 participants responded correctly. As in Investigation 3, however, the correct answer often was chosen for an unsound reason. Fourteen of the 28 participants with correct responses in Scenario 2, for instance, cited the aforementioned loss of hair *before* treatment as the reason for there being no relationship between the treatment and the *prevention* of hair loss.

With regard to the incorrect responses, 32 of the 60 responses in Scenario 1 and 16 of the 39 responses in Scenario 2 cited sheer rather than relative frequencies as the reason for their response, a typical response being that the "diet works for some and not others."

3.4.2.2 Inferential Statistics

Correctness of response—frequency. To analyse how often participants failed to assess the correlations in question versus chance, a one-sample Pearson's chi-square test was performed. As the data were independent in nature, and the expected frequency of each outcome was greater than 5 (see Table 3.8), the

assumptions necessary for performing the analysis were met prior to the reclassification of correct but unsound responses as incorrect.

Correctness of response was a categorical variable with three possible outcomes: “Yes,” “No,” and “Can’t tell.” Therefore, assuming each response will occur with equal probability, there was a one-third chance of a participant giving a correct response (i.e., “No”) and a two-thirds chance of a participant giving an incorrect response (i.e., “Yes” or “Can’t tell”). The observed versus expected outcomes both before and after the correct but unsound responses were reclassified as incorrect are summarised in Table 3.8:

Table 3.8

Observed Versus Expected Outcomes by Scenario, Before and After Correct but Unsound Responses Reclassified as Incorrect – Investigation 4

		n	
Scenario	Response	Observed	Expected
1	Correct	9	23.00
	Incorrect	60	46.00
1 – Reclassified	Correct and Sound	5	NM
	Incorrect	64	NM
2	Correct	28	22.33
	Incorrect	39	44.67
2 – Reclassified	Correct and Sound	5	NM
	Incorrect	62	NM

Note. NM = not meaningful.

As in Scenario 1 of Investigation 3, participants responded significantly different than chance with regard to Scenario 1 with medium-to-large effect size

($X^2(1) = 12.78$, $p < .001$; $r = .43$); unlike in Investigation 3, however, Investigation 4 participants responded correctly at significantly *below* chance odds. With regard to Scenario 2, the distribution of responses did not differ significantly from chance ($X^2(1) = 2.16$, $p = .142$).

As with Investigation 3, the reclassified data do not permit Pearson's chi-square testing as expected values for correct and sound frequencies cannot be calculated.

Correctness of response versus age and years of undergraduate education—correlations. This study also secondarily hypothesised no correlation between correctness of response and either age or years of undergraduate education, a null hypothesis. The results are summarised in Table 3.9:

Table 3.9

Bayesian Pearson Correlations: Scenario Correctness of Response by Age and Years of Undergraduate Education – Investigation 4

		Correctness of Response	
		Scenario 1	Scenario 2
Age	BF_{10}	0.16	0.19
Undergraduate Years	BF_{10}	0.16	0.58

A Bayesian Pearson correlation between age and both Scenario 1 and Scenario 2 response correctness yielded Bayes factors of 0.16 and 0.19, respectively, suggesting evidence in moderate support of the null hypothesis of no association between the variables.

Regarding the association between years of undergraduate education and correctness of response, a Bayes factor of 0.16 was calculated regarding Scenario 1, and a Bayes factor of 0.58 was calculated regarding Scenario 2. These results suggest

evidence in moderate support of the null hypothesis of no association between years of undergraduate education and correctness of response for Scenario 1 and weaker, but still anecdotal support for the null hypothesis with regard to Scenario 2.

3.4.3 Discussion

Before the reclassification of correct but unsound responses as incorrect, the Scenario 1 and Scenario 2 Pearson's chi-square and frequency findings support the primary hypotheses that participants will successfully discern correlational relationships less than half of the time and no better than chance despite being given complete and organised frequency data.

After analysing participants' reasons for their choices, Investigation 4's findings more strongly align with those of Investigation 3. In Scenario 1, four out of nine participants chose the correct response for an unsound reason; in Scenario 2, 23 out of 28 participants chose the correct response for an unsound reason.

With regard to the secondary hypotheses of no association between participants' correctness of response and either age or years of undergraduate education, Bayes factors suggest evidence in moderate-to-anecdotal support of those hypotheses.

Reliability. As with the last two investigations, reliability was assessed by identifying the percentage of participants who responded to both scenarios correctly and both scenarios incorrectly. Five of the 9 participants (56%) who answered the first scenario correctly answered the second scenario correctly; of the 60 who answered the first scenario incorrectly, 37 (62%) responded incorrectly to the second scenario. Once again, the relative consistency of responses between the scenarios suggests that the study possesses a reasonable level of test-retest reliability.

Validity. Relative to Investigation 3, external validity was more difficult to establish. Hypothetical and written scenarios undermined the ecological validity of both studies. Investigation 3, however, also derived some ecological validity from its naturalistic presentation of incomplete and unorganized frequency data; Investigation 4, on the other hand, did not, relying on complete, organised frequency data. As both studies attempted to sample average adults, population validity was relatively strong in each. With regard to internal validity, once again, correct and incorrect responses suggest reasonable construct and content validities, and nothing suggests poor criterion validity.

These findings will be discussed further in the General Discussion and Conclusion section.

3.5 General Discussion and Conclusion

The following discussion examines the results of Investigations 3 and 4 before and after reclassifying correct but unsound responses as incorrect.

3.5.1 Pre-Reclassification

Summary of results. Before soundness of response is considered, the findings of Investigations 3 and 4 are largely consistent. With regard to the second scenario of each study and Scenario 1 of Investigation 4, we find support for the shared primary hypotheses that participants will assess the correlations in question correctly less than half of the time and no better than chance.

With regard to their shared secondary hypotheses, evidence from Bayesian analyses suggests support for the null hypotheses of no association between correctness of response and either age and years of undergraduate education in each scenario of each study.

Indeed, the findings differ largely in only one respect: Participants correctly assessed correlation more than half of the time and better than chance in Scenario 1 of Investigation 3, counter to the shared primary hypothesis.

As discussed in Chapter 2, however, absolute performance may be more insightful than performance versus chance as we typically seek to outperform chance in matters of everyday problem representation and solution, particularly the more important the matter. While a poor grasp of correlational relationships is a likely explanation of performance at chance or worse, as the scenarios herein are hypothetical, they may also have been considered unimportant by the participants.

In Investigation 3, 46% of participants responded incorrectly with regard to Scenario 1; with regard to Scenario 2, however, 56% responded incorrectly. In Investigation 4, 87% of participants responded incorrectly with regard to Scenario 1, and 58% responded incorrectly with regard to Scenario 2. In total, participants responded incorrectly an average of 64% of the time, as well as more than half of the time in three of the four scenarios. Whereas participants responded correctly an average of 49% of the time under the naturalistic circumstances of Investigation 3, they responded correctly 27% of the time on average under the ideal circumstances of Investigation 4.

Implications. On one hand, these findings essentially confirm those of others regarding the frequency with which participants will misjudge causal and correlational relationships regardless of age and university education (e.g., Kuhn, 1991; Kuhn et al., 1985; Nisbett & Ross, 1980; Perkins, 1985). There is, however, an important difference between many of those findings and those of Investigation 3. Much of the previous work was conducted under the relatively ideal circumstances in which complete and organised frequency data or guidance was provided. As

mentioned earlier, this work was often conducted under the assumption that people will perform more poorly in the messier real world.

The findings of Investigations 3 and 4, however, fail to support this notion of poorer real-world performance. In fact, these studies suggest the opposite: Participants performed at chance or better and better in absolute numbers under more naturalistic conditions.

In sum, the pre-reclassification findings of Investigations 3 and 4 would be rather unremarkable, merely confirming the findings of others, were it not for the overall better performance under naturalistic conditions. The frequency of error in correlational judgement stands out, however, as nearly two-thirds of participants responded incorrectly across the two investigations.

Once correct but unsound responses were reclassified as incorrect, the findings became starker.

3.5.2 Post-Reclassification

Summary of results. In both investigations, when the reasons for responses were considered, the vast majority of those who responded correctly did so for unsound reasons. In Investigation 3, the number of correct responses dropped from 27 of 50 (54%) to six (12%) post reclassification regarding Scenario 1 and from 22 of 50 (44%) to two (4%) regarding Scenario 2 (as detailed in Table 3.3). In Investigation 4, the number of correct responses dropped from 9 of 69 (13%) to five (7%) post reclassification regarding Scenario 1 and from 28 of 67 (42%) to five (7%) regarding Scenario 2 (as detailed in Table 3.8). In total, 86 of the 236 (36%) responses in both investigations were correct prior to reclassification, but only 18 (8%) were correct after reclassification.

Under both scenarios in Investigation 3 and Scenario 1 of Investigation 4, the majority of incorrect responses seemed to occur for one of two reasons: “red herrings” and an exclusive focus on outcome data, the latter being more common.

The red herrings in question were not present in the scenarios themselves, but in participants’ minds. With regard to the effectiveness of the rust treatment, for example, not knowing the impact of secondary factors, such as storage conditions and climate, were reasons given in incorrect responses. In Investigation 3, some participants may have wanted to consider these factors because they affect their estimates of base rates regarding the effectiveness of the rust-prevention treatment and nicotine patch. And yet, when base rates were provided in Investigation 4, only two of the 27 participants with correct but unsound responses cited the need to consider such factors versus 25 of 41 in Investigation 3.

More frequently, however, participants overemphasised outcomes. Some participants, for instance, concluded that the rust-prevention treatment is effective when there were more instances of no rust than rust amongst those using the treatment. For other participants, a single instance of rust led them to conclude that the treatment is ineffective. In this case, participants seemed to have binary expectations: If the treatment works, it will prevent all rust; if it does not work, it will prevent no rust.

Scenario 2 of Investigation 4, concerning a treatment and the prevention of hair loss and hair regrowth, shared these reasons for incorrect responses, but a unique reason as well. Some participants seemed to be confused by the fact that the treatment had two potential effects: 1) the prevention of *incremental* hair loss and 2) the regrowth of hair. Several participants seemed to think that they were being asked about a relationship between the treatment and hair loss that occurred *before*

treatment. Of course, as there could be no relationship between hair loss before treatment and the treatment, several participants correctly responded "No." Given the frequency of this reason, however, in hindsight, the question and study may have benefitted from different wording.

Implications. Post reclassification, Investigation 4's participants, with complete and organised frequency data, performed essentially on par with those of Investigation 3, with incomplete and unorganised frequency data. Whereas an average of 8% of participants responded with correct and sound answers in Investigation 3, an average of 7% responded correctly and soundly in Investigation 4.

This chapter set out to infer a frequency of ill-defined problem misrepresentation by investigating the prevalence of correlation misjudgements. Causal and correlational relationships are central to problem representations, for they enable the creation of plans to achieve goals. These relationships also determine the quality of evidence and therefore arguments, and arguments of quality are necessary to establish accuracy and therefore to accurate problem representations.

This chapter's findings corroborated those of others that suggest the average adult's understanding of causal and correlational relationships is poor regardless of age and undergraduate education. When soundness of response was ignored, participants assessed correlational relationships incorrectly approximately two-thirds of the time; when soundness was considered, participants assessed them incorrectly over 90% of the time.

While individuals' poor performance in judging causal and correlational relationships may be well established, its implications with regard to everyday problem representation and solution are not. In light of this chapter's findings and others', particularly as university education, age and IQ (e.g., Sternberg et al., 1995;

Wigdor & Garner, 1982) seemingly do not matter, most of us may misrepresent our problems more often than not.

The general reasons underlying the incorrect responses in the two investigations herein suggest possible origins of our misunderstandings with regard to correlations. Therefore, they may also help us better understand what matters to successful problem representation.

4. Possible Common Errors in Problem Representation

Broadly, this dissertation seeks to inform a general theory of ill-defined problem representation. This dissertation began by investigating the frequency with which everyday problem misrepresentation occurs. As representations go unseen, that frequency was inferred from proxies, namely the occurrence of correlation misjudgements, evidence misassessments and poor argumentation in general. The evidence suggests a phenomenon of frequent problem misrepresentation, therefore this dissertation next sought to explain why we may misrepresent problems regularly.

As mentioned, research suggests that neither age nor undergraduate education (e.g., Kuhn et al., 1985; Nisbett & Ross, 1980) nor cognitive intelligence (e.g., Sternberg et al., 1995; Wigdor & Garner, 1982), all perhaps intuitively likely suspects, have a bearing on individuals' ability to discern causal and correlational relationships or accurately represent and solve ill-defined problems. Investigations 2, 3 and 4 herein confirm these findings with regard to age and years of undergraduate education.

To date, factors that have been suggested to correlate with one's ability to represent problems include a problem's context (e.g., Duncker, 1945; Hayes & Simon, 1977; Kotovsky et al., 1985) and the solvers' background knowledge (e.g., Bassok & Holyoak, 1989; Gick & Holyoak, 1983), including domain expertise (e.g., Chase & Simon, 1973). This research, however, often studies performance regarding well-defined or relatively specific problems, of medicine or law, for instance, that are not representative of the problems of average adults or sufficiently generalisable to them.

Therefore, we may be left with a largely unanswered question: What matters to ill-defined problem representation?

An obvious place to begin this next phase of investigation was to study the impact of knowledge of problem representation itself. As problem representation is a concept that is unknown to the average person, it is worth considering whether awareness of it enhances one's ability to represent problems. In light of Chapter 3's findings, this chapter then looked at the impacts of information regarding the problem of reliance on initially irrelevant information and regarding the problem of over- or under-reliance on relevant information in assessing correlational relationships. Recall that the majority of participants in Investigations 3 and 4 who responded incorrectly did so for one of those two reasons.

The structure of these studies also permitted further testing of Chapter 3's findings that participants frequently fail to discern correlational relationships and do so regardless of age and years of undergraduate education.

4.1 General Method

4.1.1 Participants

As in Chapter 3's investigations, participation criteria were minimised in an effort to capture a sample representative of the average adult. The only criterion for participation was that a participant be at least 17 years of age.

Here, too, participants were recruited online from the Birkbeck, University of London, participant database; through Hanover College's *Psychological Research on the Net* website; and on an ad hoc basis. Certain Birkbeck undergraduate participants received one credit toward an experiment participation requirement; otherwise no compensation was given.

4.1.2 Materials and Procedure

In brief, in each experiment, each participant received the same written description of a hypothetical situation involving an antecedent (e.g., diet) and

outcome (e.g., weight loss). Data were collected online at a time and place of each participant's choosing. While no time limit was enforced, participants committed to approximately 10 minutes of testing time.

Then, in an attempt to make the scenario more naturalistic, each participant received a series of quotes from individuals regarding their experiences with the antecedent and outcome. To safeguard against presentation order effects, the order of the quotes was randomised between participants. Participants were asked whether there was a relationship between the antecedent and outcome and the reason for their response.

Next, each participant received two or three paragraphs that constituted an intervention. Afterwards, participants were asked two comprehension questions, each with correct and incorrect answers, regarding those paragraphs. The comprehension questions served as a filter. Only data from participants who answered both comprehension questions correctly were used.

Finally, participants were given an opportunity to change their original response to the antecedent-outcome relationship question. If a participant changed their response, they were asked why. The impact of the interventions, frequency of correct responses and relationships between correctness of response and participants' ages and years of undergraduate education were measured.

4.2 Investigation 5

This study investigated whether there was a relationship between participants' ability to identify correlational relationships and receipt of basic information regarding problem representation. All of the participants received information regarding problem representation's definition, its importance to problem solving and

the consequences of misrepresentation. Every other participant also received information regarding possible errors in judgment that may lead to misrepresentation.

The primary hypothesis was that participants will better identify correlational relationships after receiving information regarding possible errors in judgment regarding such relationships in addition to the notion of problem representation in general. Secondarily, regardless of the interventions, 1) participants will assess the correlation in question correctly less than half of the time and no better than chance, and 2) correctness of response will not correlate with age and years of undergraduate education, a null hypothesis.

4.2.1 Method

4.2.1.1 Participants

Data from 100 participants were collected between 22 September and 27 October 2016. Due to specific participants' failure to provide required data, the data from only 90 participants were complete. Twenty-eight of those 90 participants, however, were excluded from analysis for having answered at least one comprehension question incorrectly. As a consequence, the data from only 62 participants were used (35 women, 26 men, one other).

As summarised in Table 4.1, M_{age} = 35.24 years, SD_{age} = 15.19 years, age range: 17 – 74 years, $M_{\text{undergrad years}}$ = 2.48, $SD_{\text{undergrad years}}$ = 1.81, and undergraduate education range: 0 – 7 years:

Table 4.1

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 5

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Age	62	35.24	15.19	17-NM	17-74	0.82	-0.04
Undergrad	62	2.48	1.81	0-NM	0-7	-0.27	-0.85

Note. NM = not meaningful.

In this sample, using the Kolmogorov-Smirnov test, the ages of participants, $D(62) = 0.17$, $p < .001$, and the participants' years of undergraduate education, $D(62) = 0.23$, $p < .001$, were both significantly non-normal in their distributions. There were more young participants than old, and years of undergraduate education were relatively unclustered around the mean.

Thirty-four of the 62 participants resided in the UK, 19 resided in the US, and 9 resided elsewhere. Fifty-two of the 62 participants cited English as their primary language.

4.2.1.2 Materials and Procedure

Scenario. Each participant received the same description of a hypothetical everyday problem. The scenario can be found in Appendix F. The scenario described a situation in which a friend who has lost his job is contemplating the use of a service that will help him rewrite his resume, or CV; structure his job search; help him network; and coach him through the search, interview and job acceptance processes. He needs a job, and he thinks this service could be a real advantage in his search.

It is an expensive service, however, so before purchasing it, he talks to five people who have used the service and asks each how helpful they thought it was in

their searches for work. Two of the people thought that the service was helpful enough to justify its purchase; three of the people thought that it was not.

At the end of the scenario, participants were asked if there is a relationship between one's use of the service (i.e., the antecedent) and success in finding a job (i.e., the outcome) and to choose one of three responses: "Yes," "No" or "Can't tell." If they chose "Yes" or "No," they were asked to give a reason for their response. If they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No."

Interventions. On the following screen, participants were presented with an intervention consisting of either two or three paragraphs. The first two paragraphs were the same for all participants:

> Problem representations are critical to problem solving. They are where the process of problem solving begins. A problem *representation* is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we want to be once the change has taken place; and the *obstacles* and *constraints* that lie between and may be determined by the problem and goal.

Followed by:

> Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. When we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues. Let's say, for example, that I'm overweight. I want to lose weight, so I exercise more. If my weight problem, however, has more to do with my diet, in ignoring my diet, I have misrepresented the problem, and I will most likely fail to lose weight.

Half of the participants received only these two paragraphs, and they constituted the "general" intervention.

Then, every other participant received a "specific" intervention, one that included this additional paragraph:

> Problems may be misrepresented for a number of reasons. Sometimes we mistakenly over-emphasize things that may matter little or not at all, like someone's opinion regardless of fact or when their experience could be an exception rather than the norm. Or we may fail to emphasize things that do matter enough, such as diet in the example above. Sometimes we'll look at a simple count and assume it tells the whole story when it may tell only part. And sometimes we're just not ready to be honest with ourselves about the nature of our problems, or we don't know what our goals should be.

Of the 62 participants used, 30 received the specific intervention and 32 received the general intervention.

All participants were then asked the same two comprehension questions regarding the paragraphs presented:

> 1) Which one of the following is not a component of a problem representation?
>
> A) An obstacle
> B) A goal
> C) The problem
> D) A friend's opinion
>
> 2) What factor or factors did I fail to consider in representing my weight problem described above?
>
> A) Exercise
> B) Diet
> C) Exercise with diet
> D) A goal

Item D is the correct response to Question 1. In acknowledging diet, Item B and Item C, an interaction effect, were both considered correct responses to Question 2, and nine of the 62 participants used chose both.

Finally, on the next screen, participants were presented with the scenario again and asked if they would like to change their original response by selecting "Yes" or "No." If they chose "No," the study ended. If they chose "Yes" (i.e., they would like to change their response), they were then asked to select their revised response of "Yes," "No" or "Can't tell." If they then chose "Yes" or "No," they were asked to give a reason for their response. If they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No."

Each of the five quotes were from people who had used the job-search service. In other words, only cases in which the antecedent was present were given, and the frequency data presented to participants was deliberately incomplete. Cases in which the antecedent was absent also would be needed in order to determine whether a correlation existed between the antecedent and outcome. Therefore, the correct response is "Can't tell."

As participants were asked to provide the reason for "Yes" or "No" responses, or if they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No," the intention was to evaluate responses for soundness. Due to a programming error, however, this information was not captured.

4.2.2 Results

4.2.2.1 Descriptive Statistics

Correctness of response. Table 4.2 summarises response frequencies before and after the general and specific interventions:

Table 4.2

Correct and Incorrect Response Frequencies Before and After Specific and General Interventions – Investigation 5

With specific intervention	*n* (Before)	*n* (After)
Can't tell (Correct)	8	8
Yes (Incorrect)	9	9
No (Incorrect)	13	13

With general intervention	*n* (Before)	*n* (After)
Can't tell (Correct)	8	6
Yes (Incorrect)	15	15
No (Incorrect)	9	11

Before the interventions, 16 participants (26%) had responded correctly; after the interventions, a total of 14 participants (23%) had responded correctly. Of those participants subjected to the specific intervention, eight (27%) were correct; of those subjected to the general intervention, six (19%) were correct.

Six of the 62 participants (10%) changed their original response. Four of the six changes occurred amongst those receiving the specific intervention, two changing their response from an incorrect "No" to a correct "Can't tell" and two changing their response from a correct "Can't tell" to an incorrect "No." The other two changes occurred amongst those who received the general intervention, both changing their correct response of "Can't tell" to an incorrect response of "No."

4.2.2.2 Inferential Statistics

Correctness of response—frequency. To analyse the frequency with which participants correctly assessed the relationship between use of the job-search service

and success in finding a job versus chance, a one-sample Pearson's chi-square test was performed. As the data were independent in nature, and the expected frequency of each outcome was greater than 5 (see Table 4.3), the assumptions necessary for performing a chi-square analysis were met.

Correctness of response is a categorical variable with three possible outcomes: "Yes," "No," and "Can't tell." Therefore, assuming each response will occur with equal probability, there was a one-third chance of a participant giving a correct response (i.e., "Can't tell") and a two-thirds chance of a participant giving an incorrect response (i.e., "Yes" or "No"). The observed versus expected responses before and after the general and specific interventions are summarised in Table 4.3:

Table 4.3

Observed Versus Expected Responses Before and After Specific and General Interventions – Investigation 5

		n	
Intervention	Response	Observed	Expected
Before	Correct	16	20.67
	Incorrect	46	41.33
After - Specific	Correct	8	10.00
	Incorrect	22	20.00
After - General	Correct	6	10.67
	Incorrect	26	21.33

Analysis of responses indicated that participants responded correctly no better than chance either before ($X^2(1) = 1.58$, $p = .208$) or after either the specific ($X^2(1) = 0.59$, $p = .441$) or general ($X^2(1) = 3.05$, $p = .081$) interventions.

Correctness of response—changes. McNemar's test was performed to analyse the frequency of changes in response from correct to incorrect and incorrect to correct with regard to the interventions. As discussed, only six of the 62 participants changed their responses, four from correct to incorrect and two from incorrect to correct. The two participants who changed their responses from incorrect to correct were subject to the specific intervention. McNemar's test results suggest that neither the specific intervention, $p = 1.000$, nor the general intervention, $p = .500$, had an effect on response.

Correctness of response versus age and years of undergraduate education—correlations. In addition, no correlations were hypothesised between correctness of response, both before and after the interventions, and either age or years of undergraduate education. Once again, because these hypotheses are null hypotheses, the results were analysed using Bayesian statistics. The results are summarised in Table 4.4:

Table 4.4

Bayesian Pearson Correlations: Correctness of Response by Age and Years of Undergraduate Education Before and After Specific and General Interventions – Investigation 5

		Correctness of Response		
		Before	Specific	General
		$N = 62$	$n = 30$	$n = 32$
Age	BF_{10}	0.17	0.24	0.26
Undergraduate years	BF_{10}	0.23	0.34	0.25

Before intervention, a Bayesian Pearson correlation between age and response correctness yielded a Bayes factor of 0.17, suggesting evidence in moderate support of no association between the variables. Evidence also suggests moderate support for no relationship between years of undergraduate education and correctness of response ($BF_{10} = 0.23$).

For those participants receiving the specific intervention, a Bayes factor of 0.24 suggests evidence in moderate support of no association between age and correctness of response. A Bayes factor of 0.34 suggests evidence in moderate-to-anecdotal support of no association between years of undergraduate education and correctness of response.

Lastly, for those participants subjected to the general intervention, Bayes factors of 0.26 and 0.25 suggest evidence in moderate support of no correlation between age and years of undergraduate education, respectively, and correctness of response.

4.2.3 Discussion

McNemar's test results fail to support this investigation's primary hypothesis that awareness of problem representation will favourably impact participants' discernment of correlational relationships. Findings do support the secondary hypotheses, however, that correct responses will occur less than half of the time and no better than chance, both before and after the specific and general interventions. In addition, Bayesian Pearson correlations suggest evidence in moderate-to-anecdotal support of the null hypotheses of no correlation between correctness of response and either age or years of undergraduate education.

While these results align with the overall findings of the studies presented in Chapter 3, they are nevertheless disappointing. Participants assessed the correlation in

question far less than half of the time and no better than chance despite having received information regarding problem representation that was intended to help them. Before the interventions, 16 participants, or 26%, had responded correctly. Therefore, 46 participants, or 74%, stood to benefit from changing their incorrect responses. After the interventions, however, only six of the 62 participants (10%) changed their original response, and only two of them corrected themselves.

Finally, only participants who answered the comprehension questions correctly were included in the analysis, their correct answers suggesting engagement in the task. Yet in an ironic twist, the participants correctly assessed the correlational relationships in question only at chance to perhaps suggest the opposite.

These findings as well as reliability and validity will be discussed further in the General Discussion and Conclusion section.

4.3 Investigation 6

The study just presented investigated the impact of general information regarding problem representation on participants' ability to discern correlational relationships and therefore represent everyday problems. Study 6 investigated participants' ability to discern correlational relationships having been given specific information regarding the problem of emphasis on irrelevant factors, the second most commonly cited reason for incorrect responses in Investigations 3 and 4.

In this study, the primary hypothesis was that participants will better discern correlational relationships having received basic information on how an emphasis on irrelevant factors in assessing such relationships can undermine that discernment. Secondarily, once again regardless of the interventions, 1) participants will assess the correlation in question correctly less than half of the time and no better than chance,

and 2) correctness of response will not correlate with either age or years of undergraduate education, a null hypothesis.

4.3.1 Method

4.3.1.1 Participants

Data for 113 participants were collected between 7 November 2016 and 16 January 2017. Due to specific participants' failure to provide required data, the data from only 94 participants were complete. Forty-seven of those 94 participants, however, were excluded from analysis due to having answered at least one comprehension question incorrectly. As a consequence, the data from only 47 participants were usable (31 women, 16 men). As summarised in Table 4.5, M_{age} = 35.15 years, SD_{age} = 13.41 years, age range: 18 – 68 years, $M_{undergrad\ years}$ = 2.71, $SD_{undergrad\ years}$ = 1.45, and undergraduate education range: 0 – 6 years:

Table 4.5

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 6

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Age	47	35.15	13.41	17-NM	18-68	0.66	-0.24
Undergrad	47	2.71	1.45	0-NM	0-6	-0.63	-0.87

Note. NM = not meaningful.

Using the Kolmogorov-Smirnov test, the ages of participants, $D(47) = 0.15$, $p = .015$, and the participants' years of undergraduate education, $D(47) = 0.24$, $p < .001$, were both significantly non-normal in their distributions. As in the last study, there were more young participants than old, and years of undergraduate education were relatively unclustered around the mean.

Thirty of the 47 participants resided in the UK, 13 resided in the US, and four resided elsewhere; 40 of the 47 participants cited English as their primary language.

4.3.1.2 Materials and Procedure

Scenario. Each participant received the same description of a hypothetical problem. The scenario can be found in Appendix G. The scenario described a situation in which the daughter of a friend has been accepted into two universities. She thinks she knows what she wants to study, and the programs for her subject seem equally good between the universities. Attendance at both universities will cost roughly the same. One university, however, is considered more prestigious than the other.

So, my friend's daughter is not sure which university to attend. While she preferred the people she met at the less prestigious university, not to mention the milder winters of its climate, she cannot help but think attending the more prestigious university will be more beneficial when it comes to finding a job. Having a job at graduation will be very important given her expected level of student debt.

Over the next few weeks, my friend's daughter asked five recent alumni of the less prestigious university what their job search experience had been. Two of those people had jobs by graduation; three of them did not.

After each scenario, participants were asked if there is a relationship between university prestige (i.e., the antecedent) and one's having a job by graduation (i.e., the outcome). They answered by choosing one of three responses: "Yes," "No" or "Can't tell." If they chose "Yes" or "No," they were asked to give a reason for their response. If they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No."

Interventions. On the following screen, all participants were presented with two paragraphs. The first paragraph was the same for all participants:

> Life is full of problems, and part of being successful is being good at solving those problems. Of course, there are many reasons why people fail to solve problems. Sometimes, for instance, they ignore one or more important aspects of a problem. Let's say that I'm overweight, and since I want to lose weight, I exercise more. But what if my weight problem has more to do with my diet than how much I exercise? If that's the case, I will most likely fail to lose weight, for I will have failed to take the role of my diet in weight loss into account.

The second paragraph varied between participants. Half of the participants received a "general" second paragraph that defined *problem representation* and discussed its impact on problem solving:

> In general terms, problem solving begins with *problem representations*, and they are critical to the process. A representation is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we'd rather be; and the *obstacles* and *constraints* that lie between them. Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. And when we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues.

Alternate participants received a "specific" second paragraph that focused on the problem of emphasis on factors that are irrelevant, at least at first:

> In addition to failing to emphasize things that do matter, sometimes we mistakenly emphasize factors that don't, at least not at first. For example,

let's say I'm considering the purchase of a rust prevention treatment for my car. While climate, let's say, can impact the occurrence of rust, I'd be wrong to consider it *before* looking at whether cars that have been treated rusted any less than cars that haven't. It's only by looking at the occurrence of rust in cars with *and without* the treatment that we can tell whether the treatment is effective and worth purchasing. In short, in deciding whether to purchase the treatment, climate doesn't matter.

Then each participant was asked two comprehension questions. One of the comprehension questions was the same for all participants:

> What factor or factors did I fail to consider in representing my weight problem described above?
>
> A) Exercise
> B) Diet
> C) My being overweight
> D) A goal

The correct answer is Item B. The other comprehension question differed between those receiving the specific second paragraph and those receiving the general second paragraph. Those participants receiving the specific second paragraph were asked:

> In the rust prevention treatment example above, which factor or factors below matter in the treatment purchase decision?
>
> A) The occurrence of rust on cars that have had the treatment
> B) The climate
> C) The occurrence of rust on cars that have not had the treatment
> D) The car dealer's reputation

Items A and C are the correct choices. The participants receiving the general second paragraph were asked:

> Which one of the following is not a component of a problem representation?

A) An obstacle
B) A goal
C) The problem
D) A friend's opinion

Item D is the correct choice. Fourteen of the 47 usable participants received the specific second paragraph; 33 of the 47 participants received the general second paragraph.

Finally, on the next screen, participants were presented with the scenario again and asked if they would like to change their original response by selecting a response of "Yes" or "No." If they chose "No," the investigation ended. If they chose "Yes" (i.e., they would like to change their response), they were then asked to select their revised response of "Yes," "No" or "Can't tell." If they chose "Yes" or "No," they were asked to give a reason for their response. If they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No."

Each of the five people surveyed had attended the less prestigious university. In other words, once again only cases in which the antecedent was absent were presented, and once again, the frequency data presented to participants was deliberately incomplete. Cases in which university prestige was present also would be needed in order to determine whether a correlation existed between university prestige and a student's having a job by graduation. Therefore, the correct response is "Can't tell."

As participants were asked to provide the reason for "Yes" or "No" responses, or if they chose "Can't tell," they were asked to list the information they would need to determine "Yes" or "No," the intention was to evaluate responses for soundness. Due to a programming error, however, this information was not captured for 36 of the 47 participants. Regardless, one of the participants providing a sound reason said, "I'd

need to know the percentage of people finding a job in both universities." An unsound reason, on the other hand, read, "Some actual statistics to prove anything. Really, I doubt the so-called 'prestige' of a place really has any effect: Most employers just want you to have the degree and don't really care where you got it from."

4.3.2 Results

4.3.2.1 Descriptive Statistics

Correctness of response. Table 4.6 summarises response frequencies before and after the general and specific interventions:

Table 4.6

Correct and Incorrect Response Frequencies Before and After Specific and General Interventions – Investigation 6

With specific intervention	*n* (Before)	*n* (After)
Can't tell (Correct)	7	6
Yes (Incorrect)	6	6
No (Incorrect)	1	2
With general intervention	***n* (Before)**	***n* (After)**
Can't tell (Correct)	20	21
Yes (Incorrect)	4	3
No (Incorrect)	9	9

Before and after the interventions, 27 participants (57%) were correct. Of those participants subjected to the specific intervention, six (43%) were correct; of those subjected to the general intervention, 21 (64%) were correct.

In total, three of the 47 participants (6%) changed their original response. Two of those participants were subjected to the specific intervention, one changing their

response from a correct “Can’t tell” to an incorrect “Yes,” and one from an incorrect “Yes” to an incorrect “No.” One participant subjected to the general intervention changed their response from an incorrect “Yes” to a correct “Can’t tell.”

Of those 11 participants for whom response reasons or lists of necessary but missing data were captured, seven responded correctly and four responded incorrectly. One of these 11 subjects who responded correctly and one who responded incorrectly were subjected to the specific intervention. None of the 11 participants, however, changed their response post intervention—the nature of the intervention did not matter. Of the seven of these 11 participants who responded correctly, only four provided sound reasons for their response.

4.3.2.2 Inferential Statistics

Correctness of response—frequency. To analyse the frequency with which participants correctly assessed the correlation between university prestige and a student’s having a job by graduation versus chance, a one-sample Pearson’s chi-square test was performed. The data were independent in nature, and the expected frequency of each outcome was greater than five with the exception of the correct responses after the specific intervention, where the expected frequency equals five when rounded (see Table 4.7).

As in the previous investigation, correctness of response is a categorical variable with three possible outcomes: “Yes,” “No,” and “Can’t tell.” Assuming each response will occur with equal probability, there is a one-third chance of a participant giving a correct response (i.e., “Can’t tell”) and a two-thirds chance of a participant giving an incorrect response (i.e., “Yes” or “No”). Observed versus expected responses before and after the interventions are summarised in Table 4.7:

Table 4.7

Observed Versus Expected Responses Before and After Specific and General Interventions – Investigation 6

Intervention	Response	Observed	Expected
Before	Correct	27	15.67
	Incorrect	20	31.33
After - Specific	Correct	6	4.70
	Incorrect	8	9.30
After - General	Correct	21	11.00
	Incorrect	12	22.00

Analysis of responses indicated that participants responded correctly more often than chance before either intervention with large effect size ($X^2(1) = 12.34$, $p < .001$; $r = .51$). After the specific intervention, participants responded correctly at chance ($X^2(1) = 0.58$, $p = .448$). For those participants subjected to the general intervention, however, participants continued to respond correctly more often than chance with large effect size ($X^2(1) = 13.67$, $p < .001$; $r = .64$).

Correctness of response—changes. McNemar's test was performed to analyse the frequency of changes in response from both correct to incorrect and incorrect to correct with regard to the interventions. Of the 47 participants, three changed their response. One participant subjected to the general intervention changed their response from incorrect to correct, and two participants subjected to the specific intervention changed their responses, one from correct to incorrect and one from an incorrect "Yes" to an incorrect "No." McNemar's test results suggest that neither the specific intervention, $p = 1.000$, nor the general intervention, $p = 1.000$, had an effect on response.

Correctness of response versus age and years of undergraduate education—correlations. As in the previous study, it was secondarily hypothesised that there would be no correlation between correctness of response, before and after the interventions, and either age or years of undergraduate education, a null hypothesis. As a consequence, the data were analysed using Bayesian statistics. The results are summarised in Table 4.8:

Table 4.8

Bayesian Pearson Correlations: Correctness of Response by Age and Years of Undergraduate Education Before and After Specific and General Interventions – Investigation 6

		Correctness of Response		
		Before	Specific	General
		$N = 47$	$n = 14$	$n = 33$
Age	BF_{10}	0.21	0.38	0.22
Undergraduate years	BF_{10}	0.28	0.33	0.33

Before intervention, a Bayesian Pearson correlation between age and response correctness yielded a Bayes factor of 0.21, suggesting evidence in moderate support of the null hypothesis of no correlation between the variables. Evidence also suggested moderate support for no correlation between years of undergraduate education and correctness of response ($BF_{10} = 0.28$).

For those participants receiving the specific intervention, a Bayes factor of 0.38 suggested evidence in anecdotal support of the null hypothesis of no correlation between age and correctness of response. A Bayes factor of 0.33 suggested evidence

in moderate-to-anecdotal support for no correlation between years of undergraduate education and correctness of response.

For those participants subjected to the general intervention, a Bayes factor of 0.22 suggested evidence in moderate support of the null hypothesis of no correlation between age and correctness of response. A Bayes factor of 0.33 suggested evidence in moderate-to-anecdotal support for no correlation between years of undergraduate education and correctness of response.

4.3.3 Discussion

McNemar's test results fail to support this investigation's primary hypothesis that knowledge of the problem of emphasis on irrelevant factors in problem representation will positively impact participants' ability to discern correlational relationships. Regardless of the specific or general intervention, few participants changed their response.

In addition, counter to one of the secondary hypotheses, before the interventions and for those subjected to the general intervention, a majority of participants responded correctly and at better than chance. In total, 27 of the 47 participants (57%) responded correctly before and after the interventions, and 21 of the 33 participants (64%) subjected to the general intervention responded correctly.

Consistent with that secondary hypothesis, however, less than half of the participants subjected to the specific intervention responded correctly and at chance. Six of the 14 participants (43%) subjected to the specific intervention responded correctly. Yet the relatively few participants subjected to the specific intervention may explain these observations.

Bayesian Pearson correlations suggest evidence in moderate-to-anecdotal support of the null hypotheses of no correlation between correctness of response and either age or years of undergraduate education.

Although the interventions seem to have had no effect, with more than half of the participants responding correctly and doing so more often than chance before the interventions and after the general intervention, the findings sound promising on balance in that they largely run counter to the findings of the previous studies herein.

As virtually all of the participants in this study had some undergraduate education, it may be tempting to attribute the number of correct responses to background knowledge or "expertise," a topic that will be discussed in the next chapter, as the hypothetical situation concerns making a preference decision between two universities. This study's findings suggest, however, that there is no relationship between years of undergraduate education and correctness of response.

In addition, only four of the seven participants who responded correctly and for whom response reasons were captured provided sound rationales for their response. While that sample of seven is admittedly small, when these findings are considered in light of those of Investigations 3 and 4, in which the vast majority of those who responded correctly did so for unsound reasons, the response bias observed in Investigation 6 may lessen.

As nearly half of the original 112 participants had been subjected to the specific intervention, the number of participants ultimately used that had been subjected to it was relatively small at 14. Perhaps the comprehension question unique to the specific intervention was relatively difficult compared to that of the general intervention.

Of the 112 original participants, 57 were subjected to the specific intervention. Nine of those 57 were disqualified for having failed to respond to the initial correlational relationship question. An additional 19 were disqualified for responding to *both* comprehension questions incorrectly. In other words, they answered the comprehension question shared with the general intervention incorrectly. In contrast, only one participant having received the general intervention was disqualified for answering both comprehension questions incorrectly.

Of the remaining 29 participants that had been subjected to the specific intervention, 15 more were disqualified for answering one of the two comprehension questions incorrectly, and two of those answered the comprehension questions shared with the general intervention incorrectly. In short, 13 of the original 57 participants (23%) subjected to the specific intervention were disqualified for answering solely the comprehension question unique to their intervention incorrectly; only two of the original 55 participants (4%) subjected to the general intervention were similarly disqualified.

These findings as well as reliability and validity will be discussed further in the General Discussion and Conclusion section.

4.4 Investigation 7

Whereas Investigation 6 studied the impact of information regarding the emphasis on irrelevant factors on the understanding of correlational relationships, this study investigated participants' ability to discern correlational relationships in light of information regarding the impact of the problem of over- or under-emphasis on relevant factors in problem representation.

In this seventh study, the primary hypothesis was that participants will better discern correlational relationships after receiving basic information regarding the

problem of over- and under-emphasis on relevant factors. As in this chapter's previous two studies in this chapter, it was secondarily hypothesised that 1) participants will assess the correlation in question correctly less than half of the time and no better than chance, and 2) correctness of response will not correlate with either age or years of undergraduate education, a null hypothesis.

4.4.1 Method

4.4.1.1 Participants

Data for 92 participants were collected between 28 November 2016 and 11 January 2017. Due to specific participants' failure to provide required data, the data from only 74 participants were complete. Thirty-nine of those 74 participants, however, were excluded from analysis due to at least one incorrect comprehension question answer. As a consequence, the data from only 35 participants were usable (26 women, 9 men). As summarised in Table 4.9, M_{age} = 35.97 years, SD_{age} = 14.44 years, age range: 17 – 64 years, $M_{undergrad\ years}$ = 2.86, $SD_{undergrad\ years}$ = 1.40, and undergraduate education range: 0 – 6 years:

Table 4.9

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 7

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Age	35	35.97	14.44	17-NM	17-64	0.38	-1.21
Undergrad	35	2.86	1.40	0-NM	0-6	-0.14	-0.15

Note. NM = not meaningful.

Using the Kolmogorov-Smirnov test, the ages of participants were not normally distributed, $D(35) = 0.16$, $p = .025$, and the participants' years of undergraduate

education were also not normally distributed, $D(35) = 0.17$, $p = .012$. Participants' ages were unclustered around the mean and had a relatively flat distribution.

Eighteen of the 35 participants resided in the UK, 12 resided in the US, and five resided elsewhere; 27 of the 35 participants cited English as their primary language.

4.4.1.2 Materials and Procedure

Scenario. Each participant received the same description of a hypothetical everyday problem. The scenario can be found in Appendix H. The scenario described a situation involving a friend of mine who collects art, and she prefers to buy her art from one dealer in particular. She likes this dealer for two reasons: She seems to have access to the "best" works by the artists my friend likes, and she seems to price the works she sells more reasonably than other dealers. However, another dealer now has a painting that my friend covets, and while my friend loves art, she primarily views art as an investment, so a work's "price-to-value" is of utmost importance to her.

As the art world is relatively small, my friend decided to ask five people she knows who have dealt with this other dealer whether the dealer offers good value for money. Three of the people thought that this other dealer offered good value for money; two did not.

After the scenario, participants were asked if there is a relationship between this other dealer (i.e., the antecedent) and good value for money (i.e., the outcome). They answered by choosing one of three responses: "Yes," "No" or "Can't tell." If they chose "Yes" or "No," they were asked to give a reason for their response. If they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No."

Interventions. On the following screen, participants were presented with two paragraphs. The first paragraph was the same for all participants:

> Life is full of problems, and part of being successful is being good at solving those problems. Of course, there are many reasons why people fail to solve problems. Sometimes, for instance, they ignore one or more important aspects of a problem. Let's say that I'm overweight, and since I want to lose weight, I exercise more. But what if my weight problem has more to do with my diet than how much I exercise? If that's the case, I will most likely fail to lose weight, for I will have failed to take the role of my diet in weight loss into account.

The second paragraph varied between participants. Half of the participants received a "general" second paragraph that defined *problem representation* and discussed its impact on problem solving:

> In general terms, problem solving begins with *problem representations*, and they are critical to the process. A representation is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we'd rather be; and the *obstacles* and *constraints* that lie between them. Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. And when we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues.

Alternate participants received a "specific" second paragraph that elaborated on the problem of over- and under-emphasis on relevant factors in problem representation specifically:

In addition to being a case of neglecting to emphasize something that matters, this example is also a case of overemphasis. Let's say, for instance, that I'm considering a specific diet I've seen advertised, so I ask people who have been on the diet what their experience has been. And let's say most of them lost weight. Given that majority, I conclude that the diet is worth trying. Conversely, given that the diet wasn't successful for *everyone*, I could also conclude it *doesn't* work. Yet I'd be incorrect in drawing either of those conclusions. In the end, it's only by looking at the incidence of weight loss with *and without* the diet that we can tell whether the diet is effective and worth purchasing.

Then each participant was asked two comprehension questions. One of the comprehension questions was the same for all participants:

> What factor did I fail to consider in representing my weight problem described above?
>
> A) Exercise
> B) Diet
> C) My being overweight
> D) A goal

The correct answer is Item B. The other comprehension question differed between those receiving the specific second paragraph and those receiving the general second paragraph. Those participants receiving the specific second paragraph were asked:

> In the diet example above, which factor or factors below matter in the decision of whether or not to try the diet?
>
> A) The occurrence of weight loss in people that have used the diet
> B) The number of ads I've seen for the diet, each of which showcases a handful of "success stories"
> C) The occurrence of weight loss in people that have not used the diet
> D) The fact that my best friend used it and failed to lose weight

Items A and C are the correct choices. The participants receiving the general second paragraph were asked:

> Which one of the following is not a component of a problem representation?
>
> A) An obstacle
> B) A goal
> C) The problem
> D) A friend's opinion

Item D is the correct choice. Eight of the 35 usable participants received the specific second paragraph; 27 of the 35 participants received the general second paragraph.

Finally, on the next screen, participants were presented with the scenario again and asked if they would like to change their original response regarding a relationship between this other art dealer and her offering value for money by selecting a response of "Yes" or "No." If they chose "No," the investigation ended. If they chose "Yes" (i.e., they would like to change their response), they were then asked to select their revised response of "Yes," "No" or "Can't tell." If they chose "Yes" or "No," they were asked to give a reason for their response. If they chose "Can't tell," they were asked to list the information they would need in order to determine "Yes" or "No."

Each of the five people surveyed in the scenario had purchased art from the other dealer. In other words, only cases in which the antecedent was present were provided. Again, the frequency data was deliberately incomplete. Cases in which the antecedent was absent would also be needed in order to determine whether a correlation existed between the antecedent and event. The correct response was "Can't tell."

As participants were asked to provide the reason for "Yes" or "No" responses, or if they responded, "Can't tell," they were asked to list the information they would need to determine "Yes" or "No," response soundness was also examined. A sound

reason acknowledged the need for data regarding other dealers, such as, “How does she compare to other dealers?” An unsound reason, on the other hand, read, “Details of the prices people have paid and what they'd sold them on for or current values.” In this case, the participant did not reference the need for data regarding the other dealer.

4.4.2 Results

4.4.2.1 Descriptive Statistics

Correctness of response. Table 4.10 summarises response frequencies before and after the general and specific interventions:

Table 4.10

Correct and Incorrect Response Frequencies Before and After Specific and General Interventions – Investigation 7

With specific intervention	*n* (Before)	*n* (After)
Can’t tell (Correct)	2	4
Yes (Incorrect)	5	3
No (Incorrect)	1	1

With general intervention	*n* (Before)	*n* (After)
Can’t tell (Correct)	10	10
Yes (Incorrect)	15	15
No (Incorrect)	2	2

Before intervention, 12 participants (34%) were correct; after the interventions, 14 (40%) were correct. Of those participants who were subjected to the specific intervention, four (50%) responded correctly; of those subjected to the general intervention, 10 (37%) responded correctly.

Two of the 35 participants (6%) changed their original response. Both of those participants had been subjected to the specific intervention, and both of them changed incorrect responses of "Yes" to correct responses of "Can't tell."

Of the 12 participants who responded correctly before the intervention, however, only three provided a sound reason for their response. Of the 14 who responded correctly after the intervention, five provided a sound reason for their response, meaning the two participants who changed their response did so soundly.

4.4.2.2 Inferential Statistics

Correctness of response—frequency. To analyse the frequency with which participants correctly assessed the correlational relationship between the other art dealer and value for money versus chance, a one-sample Pearson's chi-square test was performed. The data are independent in nature, and the expected frequency of each outcome is greater than five with the exception of correct responses after the specific intervention, where the expected frequency was below 5 (see Table 4.11).

Correctness of response is a categorical variable with three possible outcomes: "Yes," "No," and "Can't tell." Assuming each response will occur with equal probability, there is a one-third chance of a participant giving a correct response (i.e., "Can't tell") and a two-thirds chance of a participant giving an incorrect response (i.e., "Yes" or "No"). The observed versus expected scenario responses both before and after the general and specific interventions are summarised in Table 4.11:

Table 4.11

Observed Versus Expected Responses Before and After Specific and General Interventions – Investigation 7

Intervention	Response	Observed	Expected
Before	Correct	12	11.67
	Incorrect	23	23.33
After - Specific	Correct	4	2.70
	Incorrect	4	5.30
After - General	Correct	10	9.00
	Incorrect	17	18.00

Analysis of the response frequencies indicates that participants responded correctly in line with chance before the interventions ($X^2(1) = 0.02$, $p = .902$) and after the specific ($X^2(1) = 1.01$, $p = .316$) and general ($X^2(1) = 0.17$, $p = .680$) interventions.

Correctness of response—changes. McNemar's test was performed to analyse the frequency of changes in response from both correct to incorrect and incorrect to correct with regard to the interventions. Of the 35 participants, only two changed their responses, both of them being subjected to the specific intervention and both changing their responses from incorrect to correct. McNemar's test suggests that neither the specific, $p = .500$, nor general, $p = 1.000$, interventions had an effect on response.

Correctness of response versus age and years of undergraduate education—correlations. As in the previous study, it was secondarily hypothesised that there would be no correlation between correctness of response, before and after interventions, and either age or years of undergraduate education, a null hypothesis. The results are summarised in Table 4.12:

Table 4.12

Bayesian Pearson Correlations: Correctness of Response by Age and Years of Undergraduate Education Before and After Specific and General Interventions – Investigation 7

		Correctness of Response		
		Before	Specific	General
		$N = 35$	$n = 8$	$n = 27$
Age	BF_{10}	0.63	0.54	0.40
Undergraduate years	BF_{10}	0.26	1.39	0.27

Before intervention, a Bayesian Pearson correlation between age and response correctness yielded a Bayes factor of 0.63, suggesting evidence in anecdotal support of no correlation between the variables. A Bayes factor of 0.26 suggested evidence in moderate support of the null hypothesis regarding years of undergraduate education and correctness of response.

For those participants receiving the specific intervention, a Bayes factor of 0.54 suggested evidence in anecdotal support of no correlation between age and correctness of response. A Bayes factor of 1.39 regarding years of undergraduate education and correctness of response, however, suggests evidence in anecdotal support of the *alternative* hypothesis, or support for the existence of a correlation between those two variables.

Regarding those participants subjected to the general intervention, a Bayes factor of 0.40 suggested evidence in anecdotal support of the null hypothesis of no correlation between age and correctness of response. A Bayes factor of 0.27

suggested evidence in moderate support of no correlation between years of undergraduate education and correctness of response.

4.4.3 Discussion

McNemar's test results do not support the primary hypothesis that participants will better discern correlational relationships after receiving basic information regarding the problem of over- and under-emphasis on relevant factors. In fact, regardless of the specific or general interventions, few participants changed their responses.

In support of the secondary hypotheses, however, Pearson's chi-square findings suggest that correct responses will occur at chance before as well as after either intervention. In addition, before and after the interventions, on average fewer than half of the participants answered correctly: 34% before either intervention, 50% amongst those subjected to the specific intervention, and 37% amongst those subjected to the general intervention.

However, the majority of correct responses were correct for unsound reasons. When these responses were reclassified as incorrect, the number of participants that answered correctly fell further: 9% before the interventions, 38% amongst those subjected to the specific intervention, and 7% amongst those subjected to the general intervention.

While not of statistical significance, it still may be worth noting that the only two participants to change their response after intervention had been subjected to the specific intervention and that they had changed their response from incorrect to correct and for a sound reason.

With regard to the secondary hypotheses predicting no relationship between correctness of response and either participants' age or years of undergraduate

education, once again no support for these correlations was in evidence with the exception of between years of undergraduate education and correctness of response amongst those subjected to the specific intervention, where support for the alternative hypothesis was anecdotal.

With the exception of Scenario 1 in Investigation 3, these results essentially confirmed the findings of the studies presented in Chapter 3, as disappointing as they are: Participants chose the correct response no more often than chance and less than half of the time on average when soundness of response was ignored. When soundness of response was considered, and correct but unsound responses were reclassified as incorrect, the number of incorrect responses increased, including in Scenario 1 of Investigation 3, and in every case, participants responded correctly less than half of the time.

With only eight subjects, the sample subjected to the specific intervention in Investigation 7 was unintentionally small. To lend perspective, of the 92 original participants, 46 received the specific intervention and 46 received the general intervention.

Of the 46 receiving the specific intervention, 6 did not respond to the scenario questions, 18 answered *both* comprehension questions incorrectly, and 14 responded to one comprehension question incorrectly to leave eight eligible participants. Of the 14 participants who responded to one comprehension question incorrectly, eight of those answered the question unique to the specific intervention incorrectly.

With regard to the 46 participants receiving the general intervention, 12 did not respond to the scenario questions, only two answered *both* comprehension questions incorrectly, and only five responded to one comprehension question incorrectly to leave 27 eligible participants. Of the five participants who responded to

one comprehension question incorrectly, none of them answered the question unique to the general intervention incorrectly.

In short and as in Investigation 6, the comprehension question unique to the specific intervention may have been more difficult. Eight of the original 46 participants (17%) subjected to the specific intervention were disqualified for answering solely the comprehension question unique to their intervention incorrectly whereas none of the original 46 participants subjected to the general intervention were similarly disqualified.

These findings as well as reliability and validity will be discussed further in the General Discussion and Conclusion section.

4.5 General Discussion and Conclusion

Summary of results. Their similar structures make Investigations 5, 6 and 7 ripe for comparison, and their comprehension questions permitted analysis focused on those participants more likely to have been engaged. The number of participants who responded to the scenario questions and answered at least one of the two comprehension incorrectly in each study was noteworthy: 28, or 31% in Study 5; 47, or 50%, in Study 6; and 39, or 53% in Study 7. Those rates are troubling as they suggest a notable lack of engagement in general, but at the same time, they underscore the importance of focusing on those who are relatively engaged.

Even participants with correct comprehension question responses, however, may have failed to engage consistently across the studies. While the number of correct responses versus chance (with large effect size) suggest participants were engaging before the interventions and after the general intervention in Investigation 6, in Investigations 5 and 7, both before and regardless of the interventions, as well as

after the specific intervention in Investigation 6, participants responded at chance to make engagement suspect.

In no study in this chapter did the interventions have a bearing on participants' discernment regarding the correlational relationships in question. Of the 144 participants between the studies, only 11 (8%) changed their original assessment regarding the relationship between antecedents and outcomes, and the net change was not in favour of correctness. Five participants changed their response from incorrect to correct, but six changed their response from correct to incorrect. Eight of the 11 participants who changed their response had been subjected to specific interventions, four changing from an incorrect response to a correct response. Again, while not of statistical significance, it may be interesting to note that four of the five participants who changed their response from incorrect to correct had been subjected to the specific interventions.

This finding aligns with that of Kuhn et al. (1985), in which the provision of additional and necessary antecedent-outcome frequency data did not affect subjects' initial inferences with regard to correlational and causal relationships.

In total across Investigations 5, 6 and 7, the majority of participants (89 out of 144, or 62%) responded incorrectly both before and after the interventions regardless of response soundness. On a study-by-study basis, too, the majority of participants in Investigations 5 and 7 responded incorrectly—74% and 66%, respectively, before the interventions, changing to 77% and 60%, respectively, post interventions. These findings, too, align with the work of others regarding adults' frequency of error in assessing causal and correlational relationships (e.g., Kuhn, 1991; Kuhn et al., 1985; Nisbett & Ross, 1980).

Investigation 6's correct response bias, however, is an anomaly. The scenario in Investigation 6 concerned an individual's selection of a university in light of post-graduation employment prospects. As virtually all participants were attending or had attended university and therefore had gone through the university selection process, one might assume that perhaps these findings speak to the power of experience and expertise. In problem representation, experts' representations tend to emphasize solution-relevant features whereas novices' representations tend to focus on irrelevant ones (e.g., Chase & Simon, 1973; Duncker, 1945).

Should experience and expertise have mattered in Investigation 6, given the unusually large number of correct responses, there was less opportunity for performance improvement, or in changing responses from incorrect to correct, the desired effect. Given the large number of incorrect responses in Investigations 5 and 7, on the other hand, the opposite is true: There was more opportunity for performance improvement. Indeed, whereas two participants changed their responses from incorrect to correct in Investigations 5 and 7, only one participant did so in Investigation 6.

Coincidentally, however, each of this chapter's investigations analysed the relationship between years of undergraduate education and correctness of response, and no relationship was revealed between the two. In Investigation 6—in Investigations 3, 4, 5 and 7, too, with the exception of after the specific intervention in Investigation 7—evidence was found to be in moderate-to-anecdotal support of the null hypothesis of no correlation between correctness of response and years of undergraduate education. In other words, it is difficult to argue the importance of undergraduate experience or expertise in explaining the Investigation 6 findings.

Reliability. As with Investigations 2, 3 and 4, the hypothetical scenarios used in Investigations 5, 6 and 7 were created specifically for the studies, so there were no pre-existing reliability assessments of the tools.

Once again, however, the same group of participants was tested twice within each experiment, once before and once after intervention. The relative consistency of responses, as evidenced by the 92% of participants who did not change their original response across the studies, suggests a high level of reliability. While that test-retest reliability may have been compromised by the short interval between the scenarios and the prompting of participants regarding changes in their responses, the sheer number of participants that remained steadfast in their responses likely still suggests reasonable reliability. Regardless, in hindsight, simply re-presenting each scenario and questions without first asking whether a participant would like to change their response would have been a stronger approach.

Validity. Once again, correct and incorrect responses suggest criterion and content validities, and as with previous studies, nothing suggests poor criterion validity. External validity, however, was again more difficult to ascertain. A case for ecological validity is perhaps stronger given the studies' naturalistic bent, but that case, in turn, is weakened by their hypothetical and written natures. The case for population validity, on the other hand, is somewhat stronger as participants are meant to represent average adults.

Implications. So far, this dissertation's findings seem relatively clear: Participants discern correlational relationships no better than chance and less than half of the time, and age and years of undergraduate education do not matter to this ability. This news, however, is perhaps old news. What is new is that this failure suggests a problem with regard to problem representation.

A weakness in the ability to identify correlations implies an inability to evaluate evidence in the arguments necessary to truth or accuracy, and problem representations must be accurate if a problem is to be solved. This weakness also suggests an inability to construct the plans that allow us to solve them.

This chapter's objective was to further explore what matters to one's ability to represent problems. Claims have been made that a problem's context (e.g., Duncker, 1945; Hayes & Simon, 1977) and the solvers' background knowledge (e.g., Chase & Simon, 1973; Duncker, 1945; Gick & Holyoak, 1980) matter to ill-defined problem representation (for an overview, see Novick & Bassok, 2005). Yet as mentioned, these studies often rely on well-defined problems or ideal circumstances in which complete and organised frequency data are available, rather than the more naturalistic conditions of incomplete and unorganised frequency data.

As knowledge of well-defined problems has failed to transfer to our understanding of ill-defined problems (Galotti, 1989), however, these claims may not be sound, particularly due to reliability. Take context, for instance. Hayes and Simon (1977) suggest that story content and problem phrasing may affect how the solver represents a problem. While that may make intuitive sense, others may argue the same with regard to education and IQ, or cognitive intelligence levels, only the evidence suggests otherwise.

Hayes and Simon base their claim on a study that used variants of the Tower of Hanoi problem, only their study concerned monsters and globes. In addition to the Tower of Hanoi being a well-defined problem, the use of monsters and globes further begs questions of generalisability. Similarly, while expertise and experience may matter to the representation of ill-defined problems, Kuhn (1991) suggests that one's

expertise and experience may also make it more difficult for one to recognise and consider opposing views.

After identifying seemingly common errors in interpreting correlational relationships in the studies of Chapter 3, this chapter further explored their role in participants' understanding of such relationships. Yet this chapter's findings suggest that they may not matter to problem representation at all.

Kuhn claims that the most revealing finding of her 1991 study is the high level of certainty amongst her participants in their explanations of phenomena about which they are asked. And in the studies of this chapter, only five, or 3%, of the participants deemed engaged changed their response from incorrect to correct, and the interventions did not seem to matter. Of the 144 participants in this chapter's studies, 89 (62%) failed to change an incorrect response. Participants will not consider re-representing their problems, however, if they never perceive their representations as possibly being incorrect.

Therefore, perhaps in their overconfidence, participants commit so many errors in their correlational assessments and therefore everyday problem representations that those errors dominate any correlational analyses in which they are a factor. As a consequence, any relationships regarding other factors that may matter, including those that are the subject of this chapter, fail to come to light. As such, this dissertation now explores the frequency and consequences of overconfidence.

5. Overconfidence and Problem Representation

Often in life, one's perception of their competence is as important as their actual competence. It is imperative when driving a car or caring for someone (including oneself), for instance, that one possess the skills and knowledge they think they possess, for their overestimation could jeopardize people's safety and wellbeing (Bjork, 1999).

Similarly, it is imperative that one possess the skills and knowledge necessary to sufficiently represent a problem, for their overestimation will prevent them from solving it. When one overestimates their competence, they will fail to see and fill the knowledge and skill gaps that may cause them to misrepresent their problem.

Numerous studies, however, suggest that individuals not only frequently miss-assess their competence (e.g., Klayman, Soll, González-Vallejo, & Barlas, 1999), but also "tend to be blissfully unaware of their incompetence" (Dunning, Johnson, Ehrlinger, & Kruger, 2003, p. 83). These miss-assessments of our competence typically take the form of overconfidence (e.g., Dunning, Heath, & Suls, 2004; Hoch, 1985; Lichtenstein & Fischhoff, 1977).

5.1 Expertise and Overconfidence

As discussed in Chapter 1, representing an initially ill-defined problem involves determining what information is relevant to its solution (Galotti, 1989). When one includes irrelevant factors, one risks misrepresenting their problem.

Expertise, in the form of background knowledge and prior experience, can affect whether and to what extent one focuses on what is relevant. Indeed, across numerous domains, from law and medicine to figure skating and chess, research suggests that expertise matters in problem representation (e.g., Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Deakin & Allard, 1991). Experts' representations

tend to highlight solution-relevant structural features, such as relationships between problem objects; novices' representations, on the other hand, tend to highlight solution-irrelevant features, such as the objects themselves (Duncker, 1945; Novick, 1988).

Expertise, however, is double-edged. While expertise can help one determine what is relevant to a problem's representation and solution, the overfamiliarity that expertise engenders can constrain representations and undermine solution too. Kuhn (1991) suggests that expertise can make it more difficult for an expert to recognise another view, as Tolstoy (1943) illustrates:

> Doctors came to see her singly and in consultation; talked much in French, German and Latin; blamed one another and prescribed a great variety of medicines for all the diseases known to them, but the simple idea never occurred to any of them that they could not know the disease Natasha was suffering from. (p. 721)

Ottati, Price, Wilson, and Sumaktoyo (2015) suggest social norms encourage experts to assume relatively closed-minded, dogmatic orientations. While experts consistently hold themselves in high regard, however, that view correlates poorly with their performance (Burgman et al., 2011; Kahneman & Klein, 2009). In other words, experts and their expertise should not be taken at face value but be considered.

Regardless, one's knowledge and experience are limited, of course, and with regard to everyday problems, often one does not get to pick them. Therefore, one will not be an expert regarding most of life's problems, which makes the ability to discern those who can help important. When someone is overconfident, however, they will not seek additional information or experience, expert or otherwise, when they should, for they will never see the need.

5.2 The Greater Problem of Overconfidence

While sometimes viewed as advantageous, with claims of increased morale and resolve (Johnson & Fowler, 2011), overconfidence is typically considered problematic. Researchers have suggested, for instance, that overconfidence may help explain wars (e.g., Howard, 1983; Johnson, 2004) and stock market bubbles (e.g., Odean, 1999). Despite its potential importance, however, overconfidence has been inconsistently defined and studied, which muddles rather than clarifies our understanding of it.

In their review of empirical studies on overconfidence, Moore and Healy (2008) propose that overconfidence has been interpreted in largely three ways: overestimation, overprecision, and overplacement. *Overestimation* refers to one's overestimation of their ability, performance, control, or odds of success. *Overprecision* refers to one's excessive certainty regarding the accuracy of their beliefs. Less common, *Overplacement* refers to one's belief that they are better than others.

Regardless of its subcategories, Dunning, Heath, and Suls (2004) sum up the breadth of the problem of overconfidence:

> On average, people say that they are "above average" in skill (a conclusion that defies statistical possibility), over-estimate the likelihood that they will engage in desirable behaviours and achieve favourable outcomes…and reach judgments with too much confidence. (p. 69)

Heath and Tversky (1990), in turn, sum up the issue's importance: Confidence controls behaviour.

Griffin and Tversky (1992) propose that confidence is a function of the balance of arguments for and against competing hypotheses without sufficient regard

for evidence. Their suggestion seems particularly relevant to problem representation in light of its dependency on genuine evidence and sound argumentation.

Overconfidence, then, occurs when arguments are weak, either in their balance of claims and counterclaims or in the absence of genuine evidence and sometimes both, and the arguer fails to see them as such. As this dissertation asserts that weak arguments to some extent explain the phenomenon of problem misrepresentation, overconfidence and problem misrepresentation may go hand in hand.

While an overconfidence bias appears to be systematic, overconfidence seems more pronounced when questions are open-ended (Klayman et al., 1999). Therefore, as ill-defined problems are relatively open-ended in nature, they may be inherently susceptible to overconfidence. Research also suggests that most people are extremely overconfident in their answers to general knowledge questions (Arkes, Christensen, Lai, & Blumer, 1987; Fischhoff, Slovic, & Lichtenstein, 1977), a category into which many everyday problems fall.

Perhaps relatedly, in Studies 3, 4 and 5 herein, participants were presented with everyday scenarios after which they were asked if there was a relationship between each scenario's antecedent and outcome, such as an auto rust-prevention treatment and the occurrence of rust, respectively. After responding, participants were given information that was intended to help them assess the relationship correctly. After receiving that information, participants were given an opportunity to change their responses, yet virtually no participant did.

Unsurprisingly, overconfidence correlates positively with underachievement (Dunlosky & Rawson, 2012). Overconfident entrepreneurs may generate a higher number of start-ups than others, for instance, but that overconfidence correlates

negatively with the survival rates of those businesses (Koellinger, Minniti, & Schade, 2007).

In sum, an honest assessment of one's knowledge and skills is essential to problem representation. When overconfident, people never question whether their knowledge and skills are adequate and whether they may have misrepresented their problem. This dissertation therefore investigated the frequency of overconfidence or, to be more specific, overprecision with regard to evidence in arguments.

5.3 Investigation 8

As previously discussed, in their lack of clarity, everyday problems lend themselves to misrepresentation, yet only sufficiently accurate problem representations permit solution. In general, accuracy, or truth, is arrived at through argumentation, an iterative process of claims and counterclaims, each supported by genuine evidence. One is more likely to develop argumentation skills and engage in argument, of course, if one appreciates their value. In overconfidence, however, one is unlikely to do either—when one believes they are correct from the start, there is no need (Kuhn, 1991).

Therefore, this study examined participants' confidence in evidence in the context of argumentation. The primary hypothesis was that participants will be overconfident in pseudo-evidence or non-evidence with significant frequency. In line with previous findings, the secondary hypotheses were that 1) participants will assess the strength of the evidence correctly less than half of the time and no better than chance, and 2) the correctness of those judgements will not correlate with either age or years of undergraduate education, a null hypothesis.

5.3.1 Method

5.3.1.1 Participants

As in this dissertation's previous studies, participation criteria were minimised in the hope of capturing a sample representative of the average adult: The only criterion for participation was that a participant be at least 17 years of age.

Data for 56 participants were collected between 5 December 2017 and 26 January 2018. Participants were recruited online from the Birkbeck, University of London, participant database; through Hanover College's *Psychological Research on the Net* website; and on an ad hoc basis.

Due to specific participants' failure to provide required data, however, data from only 42 participants were usable (27 women, 14 men, one other). As summarised in Table 5.1, M_{age} = 41.48 years, SD_{age} = 14.84 years, age range: 18 – 76 years, $M_{undergrad\ years}$ = 3.11 years, $SD_{undergrad\ years}$ = 1.30 years, and undergraduate education range: 0 – 4 years:

Table 5.1

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 8

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Age	42	41.48	14.84	17-NM	18-76	-0.02	-1.11
Undergrad	42	3.11	1.30	0-NM	0-4	-0.78	-0.44

Note. NM = not meaningful.

In this sample, using the Kolmogorov-Smirnov test, the ages of participants were normally distributed, $D(42) = 0.10$, $p = .200$; the participants' years of undergraduate education, however, were not normally distributed, $D(42) = 0.35$, $p < .001$. Participants had a relatively high number of years of undergraduate education.

Nineteen of the 42 participants resided in the US, 17 resided in the UK, and six resided elsewhere; 37 of the 42 participants cited English as their primary language. Certain Birkbeck undergraduate participants received one credit toward an experiment participation requirement; otherwise no compensation was given.

5.3.1.2 Materials and Procedure

While no time limit was enforced, participants committed to approximately five minutes of testing time. Data were collected online at a time and place of each participant's choosing.

Each participant received the same two scenarios in sequential order. The scenarios can be found in Appendix I. In an effort to minimise systematic variation, half of the participants received Scenario 1 first, and alternate participants received Scenario 2 first.

Each scenario described a hypothetical everyday problem and made a claim regarding its cause. Scenario 1 described a situation in which in some of the world's most developed countries, 15-year-olds perform no better than average on international standardized assessment tests despite higher-than-average annual education spending per student. A group of researchers claim that developed-country wealth causes more money to be invested less thoughtfully. Scenario 2 described a situation in which a friend of mine is purchasing a new computer, and the salesperson has recommended that she buy an extended warranty that will cover repairs for the two-year period after the manufacturer's first-year warranty expires.

A list of five items was then provided, each item being possible evidence in support of the causal claim. Participants were asked to rate the strength of each item as evidence in support of the causal claim on a scale of 1 to 5, 1 being strongest and 5

being weakest or non-evidence. In an effort to further minimise systematic variation, the order of the list items was randomised between participants.

In each of the two scenarios, only one list item was genuine evidence; all other items were either pseudo-evidence or non-evidence. As discussed in Chapter 2, genuine evidence is primarily co-variation evidence that suggests at least a correlational relationship and at best a causal relationship between an antecedent and outcome. Pseudo-evidence is typically a single case or generalised summary that depicts how the phenomenon might occur (i.e., evidence chosen to illustrate a claim, rather than selected at chance) and does not provide a sufficient basis for inferring a relationship between an antecedent and outcome. Non-evidence is typically unconnected to the causal claim or cites the outcome as evidence of the antecedent.

In Scenario 1, the following item constituted genuine evidence due to the suggested negative co-variation between annual spend increases and test performance: "In most developed countries, 15-year-olds' performance on international standardized assessment tests does not improve with annual education-spending-per-student increases." In Scenario 2, the following item constituted genuine evidence due to the suggested positive co-variation between the computer model in question and problems after the first year of ownership: "A survey of consumers compared a number of computers, including my friend's. Of 100 people who purchased hers, 64 had problems in the second or third year of ownership."

In each scenario, two of the remaining four list items constituted pseudo-evidence and the other two constituted non-evidence. For example, in Scenario 2, one of the pseudo-evidence items read, "One of my friend's colleagues bought the same computer and the warranty, and he was glad he did. The extended warranty covered the problems he had two years after purchase." In Scenario 1, on the other hand, a

non-evidence item read, “15-year-olds in some of the most developed countries perform no better than average on international standardized assessment tests because the tests measure the wrong things.” This item has no bearing on the relationship between developed-country wealth and thoughtful investment that is claimed by the researchers.

As each scenario’s potential evidence list contained only one genuine evidence item, only that item would be correctly rated a 1; the other items would be correctly rated a 5.

5.3.2 Results

5.3.2.1 Descriptive Statistics

Ratings by evidence item. Table 5.2 summarises the frequency of participants’ ratings by evidence item:

Table 5.2

Frequency of Ratings by Evidence Item – Investigation 8

	Scenario 1 Item					Scenario 2 Item				
	1, n	2, *g*	3, p	4, n	5, p	1, p	2, n	3, p	4, g	5, n
1 - Strongest	0	19	4	2	4	11	3	4	26	3
2	4	7	4	6	4	12	3	3	11	7
3	6	4	7	4	8	10	6	6	2	6
4	6	6	9	11	8	3	10	8	2	4
5 - Weakest	26	6	18	19	18	6	20	21	1	22

Notes. $N = 42$; g = genuine evidence, p = pseudo-evidence, n = non-evidence.

For Scenario 1, the mode for each evidence item was the correct rating. The median, on the other hand, was the correct rating for Item 1 only. For Item 2, while the correct rating was 1, the median was 2; for Items 3, 4 and 5, the correct rating was 5, while

the median was 4. For Scenario 2, with the exception of Item 1, the mode and median rating for each evidence item was the correct rating; for Item 1, the median and mode were 2, but the correct rating was 1.

Correct ratings by evidence item. The frequency of correct ratings (i.e., genuine evidence items rated 1 and pseudo- and non-evidence items rated 5) is presented in Table 5.3:

Table 5.3

Frequency of Correct Ratings by Evidence Item – Investigation 8

Scenario	Item	% Correct
1	1, n	62
	2, g	45
	3, p	43
	4, n	45
	5, p	43
2	1, p	14
	2, n	48
	3, p	50
	4, g	62
	5, n	52

Note. $N = 42$; g = genuine evidence, p = pseudo-evidence, n = non-evidence.

Pseudo- or non-evidence item ratings less than 5 suggest varying degrees of overconfidence in the item; genuine evidence item ratings greater than 1 suggest varying degrees of under-confidence in the item.

Between the scenarios, for seven of the ten evidence items, half or more of the participants rated the evidence incorrectly. With regard to the genuine evidence items, 45% of the participants in Scenario 1 and 62% of the participants in Scenario 2 rated

the evidence correctly. With regard to each of the four pseudo-evidence items, however, in each case half or more of the participants rated the evidence incorrectly. For two of the non-evidence items, more than half of the participants rated the item correctly; for the other two, more than half did not.

Correct ratings by evidence type. Given that there were 42 participants and that each participant rated 10 items, there were a total of 420 ratings. Four of the 10 items were non-evidence (168 ratings), four were pseudo-evidence (168 ratings), and two were genuine evidence (84 ratings). Rating correctness by evidence type is summarised in Table 5.4:

Table 5.4

Frequency of Correct Ratings by Evidence Type – Investigation 8

Evidence Type	Correct	% of Type
Non (*n* = 168)	87	52
Pseudo (*n* = 168)	63	38
Genuine (*n* = 84)	45	54

Note. N = 420

The strength of the pseudo-evidence items was most often judged incorrectly, at 62% of the time, whereas that of the non-evidence and genuine evidence items was misjudged slightly less than half of the time.

5.3.2.2 Inferential Statistics

Ratings by evidence type—associations. To assess whether there was an association between evidence type and participants' ratings, a Pearson's chi-square test was performed. The data was independent in nature, and the expected frequency of each outcome was greater than 5, so the assumptions necessary for performing a chi-square analysis were met.

The strength rating was a categorical variable with five sub-categories as there were five possible ratings for each potential evidence item: 1, 2, 3, 4 and 5. Evidence type was a categorical variable with three sub-categories: genuine, pseudo-evidence and non-evidence.

Results suggest a significant association between evidence type and ratings ($X^2(8) = 116.82$, $p < .001$). Table 5.5 summarises the underlying *z*-scores:

Table 5.5

Overall z-scores, Strength Ratings by Evidence Type – Investigation 8

	Rating				
Type	1(Strongest)	2	3	4	5(Weakest)
Genuine	7.6	1.7	-1.7	-1.5	-4.4
Pseudo	-1.3	-0.3	1.5	0.2	0.0
Non	-4.1	-0.9	-0.3	0.8	3.1

Analysis of *z*-scores suggest that the overall association between evidence type and ratings is driven by the genuine evidence ratings of 1 (i.e., the correct rating) and 5 together with non-evidence ratings of 1 and 5 (i.e., the correct rating), all significant at $p < .001$ with the exception of non-evidence with a rating of 5, which is significant at $p < .01$. In other words, participants are rating genuine evidence and non-evidence items correctly significantly more often than expected under the null hypothesis and rating them incorrectly at the extremes of 5 and 1, respectively, significantly less often than expected under the null hypothesis.

When Scenarios 1 and 2 were analysed separately, the association between evidence type and ratings remained significant ($X^2(8) = 53.20$, $p < .001$, and $X^2(8) = 68.37$, $p < .001$, respectively). Table 5.6 summarises the underlying *z*-scores by scenario:

Table 5.6

Scenario z-scores, Ratings by Evidence Type – Investigation 8

Scenario	Type	Rating				
		1(Strongest)	2	3	4	5(Weakest)
1	Genuine	5.5	0.9	-0.7	-0.7	-2.7
	Pseudo	-1.1	-0.6	1.0	0.3	0.2
	Non	-2.8	0.0	-0.5	0.3	1.7
2	Genuine	5.4	1.4	-1.6	-1.5	-3.5
	Pseudo	-0.9	0.2	1.2	0.1	-0.2
	Non	-3.0	-1.2	0.0	1.0	2.6

The scenario findings, both chi-square and *z*-scores, mirror those of the overall findings with the exception of non-evidence ratings of 5 in Scenario 1, where they were no longer significant. Otherwise, participants continued to rate the genuine evidence items correctly significantly more often than expected under the null hypothesis and rated them and the non-evidence items incorrectly at the extremes of 5 and 1, respectively, significantly less often than expected under the null hypothesis.

More specifically, in Scenario 1, the associations between evidence type and ratings remained driven by the genuine evidence ratings of 1 ($p < .001$) and 5 ($p < .01$) together with a non-evidence rating of 1 ($p < .01$). In Scenario 2, the associations between evidence type and ratings remained driven by the genuine evidence ratings of 1 ($p < .001$) and 5 ($p < .001$) together with non-evidence ratings of 1 ($p < .01$) and 5 ($p < .01$).

Correct ratings by evidence item—frequency. To analyse how often participants rated the evidence in support of these causal claims correctly versus chance, a one-sample Pearson's chi-square test was performed. As the data were

independent in nature, and the expected frequency of each outcome was greater than 5, the assumptions necessary for performing a chi-square analysis were met.

Correctness of rating was a categorical variable with five sub-categories as there were five possible ratings for each potential evidence item. Assigning a rating of 1 to a genuine evidence item or a rating of 5 to a pseudo- or non-evidence item constituted a correct response; all other ratings were considered incorrect. Therefore, assuming each rating will occur with equal probability, there was a one-out-of-five chance of a participant giving a correct response and a four-fifths chance of a participant giving an incorrect response. The observed versus expected outcomes are summarised in Table 5.7:

Table 5.7

Observed Versus Expected Correct Ratings, X^2 and p by Evidence Item – Investigation 8

Scenario	Item	Correct			
		Observed	Expected	$X^2(4)$	p
1	1, n	26	8.4	48.95	.000
	2, g	19	8.4	17.29	.002
	3, p	18	8.4	15.86	.003
	4, n	19	8.4	22.05	.000
	5, p	18	8.4	15.62	.004
2	1, p	6	8.4	6.81	.146
	2, n	20	8.4	23.95	.000
	3, p	21	8.4	25.38	.000
	4, g	26	8.4	53.95	.000
	5, n	22	8.4	28.71	.000

Note. g = genuine evidence; p = pseudo-evidence; n = non-evidence.

With the exception of Scenario 2, Item 1, there was a significant tendency for participants to rate items correctly; participants rated Item 1 in Scenario 2 correctly at chance.

Correct ratings versus age and years of undergraduate education—correlations. It was also secondarily hypothesised that there would be no relationship between correct ratings and either age or years of undergraduate education, a null hypothesis. Once again, Bayesian statistics were used to analyse the extent to which the data increased our confidence in this null hypothesis. The results are summarised in Table 5.8:

Table 5.8

Bayesian Linear Regression: Correct Ratings Versus Age and Years of Undergraduate Education – Investigation 8

	Total	Scenario 1	Scenario 2
	BF_{10}	BF_{10}	BF_{10}
Age (A)	0.56	1.00	0.31
Undergraduate (U)	0.66	0.57	0.41
A x U	0.32	0.49	0.18

With the exception of age in Scenario 1, Bayesian linear regression yields Bayes factors that suggest evidence in moderate-to-anecdotal support for the null hypotheses overall and in each scenario. A Bayes factor of 1.00 suggests the evidence supports neither the null or alternative hypothesis regarding age and correctness of rating in Scenario 1.

5.4 Discussion and Conclusion

Summary of results. The findings herein fail to support this study's primary hypothesis of participant overconfidence in pseudo- and non-evidence. Pearson's chi-

square findings suggest that there are significant associations between ratings and evidence types overall and in each scenario. Overall, those associations are the result of genuine evidence and non-evidence items being rated correctly as the strongest and weakest forms of evidence, respectively, more often than expected under the null hypothesis and incorrectly as the weakest and strongest, respectively, less often than expected under the null hypothesis. Overall, all other associations between ratings and evidence types were in line with expectations under the null hypothesis.

With the exception of the correct rating for non-evidence items in Scenario 1, those findings held for each scenario, the correct rating for non-evidence items in Scenario 1 no longer being significant.

With regard to the secondary hypotheses, however, in total participants judged evidence strength incorrectly more than half of the time (54%) due to misjudgements regarding pseudo-evidence items. Participants overestimated the strength of pseudo-evidence 62% of the time, a finding akin to Kuhn's (1991), and overestimated the strength of non-evidence nearly half of the time (48%). Regardless, with the exception of Item 1 in Scenario 2, a pseudo-evidence item, Pearson's chi-square findings suggest that participants will rate the strength of evidence correctly better than chance.

Bayesian linear regression suggests evidence in moderate-to-anecdotal support of the null hypotheses of no relationships between the correctness of evidence strength judgements and either age or years of undergraduate education.

Reliability. In Investigation 8, of the 19 participants that rated the genuine evidence item correctly (i.e., a rating of "1") in Scenario 1, 14 (74%) rated the genuine evidence item correctly in Scenario 2. Of the 15 participants that rated both non-evidence items correctly (i.e., a rating of "5") in Scenario 1, 11 (73%) rated both

non-evidence items correctly in Scenario 2. Lastly, of the 14 participants that rated both pseudo-evidence items correctly (i.e., a rating of "5") in Scenario 1, 4 (29%) rated both pseudo-evidence items correctly in Scenario 2.

Similarly, of the 23 participants that rated the genuine evidence item incorrectly (i.e., a rating of "5") in Scenario 1, 11 (48%) rated the genuine evidence item incorrectly in Scenario 2. Of the 27 participants that rated both non-evidence items incorrectly (i.e., a rating of "1") in Scenario 1, 22 (81%) rated both non-evidence items incorrectly in Scenario 2. Lastly, of the 28 participants that rated both pseudo-evidence items incorrectly (i.e., a rating of "1") in Scenario 1, 27 (96%) rated both pseudo-evidence items incorrectly in Scenario 2.

In other words, 60% of participants consistently rated the genuine evidence items a 1 or 5 in both scenarios, 79% consistently rated the non-evidence items a 1 or 5 in both scenarios, and 74% of participants rated the pseudo-evidence items a 1 or 5 in both scenarios. As once again the same group of participants was tested twice within the study, once on each scenario, concordance of participants' responses between the scenarios suggests that the study possesses a reasonable level of reliability.

The variance between the correct and incorrect response concordances with regard to pseudo-evidence further evidences and highlights participants' struggles with evidence of this nature—whereas only 29% of participants rated both pseudo-evidence items correctly in each scenario, 96% rated both incorrectly.

Validity. As in this dissertation's previous studies, the correct and incorrect responses of Study 8 suggest an internal validity, particularly with regard to construct and content validities. External validity, however, is more difficult to ascertain. Ecological validity may or may not be present—on one hand, the scenarios are

hypothetical; on the other, the nature of the evidence items may be realistic enough. The sample of average adults helps in establishing population validity.

Implications. In problem representation, the goal is to minimise error to increase the odds of sufficient representation and therefore solution. Needless to say, overconfidence undermines those minimisation efforts. Chance aside, the incidence of overconfidence in pseudo-evidence and non-evidence make overconfidence in problem representation a phenomenon worthy not just of our attention, but concern: 62% of participants over-weighted the strength of pseudo-evidence, and 48% over-weighted the strength of non-evidence.

In addition, there is the problem of under-confidence in genuine evidence, and while not a focus of this study and less studied than overconfidence, it is nevertheless a threat to problem representation. Participants under-weighted the strength of genuine evidence also nearly half (46%) of the time. In other words, participants may fail to consider genuine evidence enough in their representations almost half of the time.

The findings in previous chapters suggest that people often misjudge or neglect evidence in establishing the correlational and causal relationships necessary to arrive at accuracies, or truths, and construct problem solutions. In this context, a prevalence of overconfidence is perhaps less surprising—that mishandling or disregard for evidence is perhaps the same insufficient regard for evidence that underlies overconfidence. In overconfidence, by definition, one fails to reflect reality, or the truth.

Overconfidence seems much more likely when participants are faced with pseudo-evidence versus non-evidence. These findings may align with Levine, Park and McCornack's (1999) "veracity effect," in which participants detected truth

significantly greater than chance, but their detection of lies was often significantly below chance. In Investigation 8, the strength of genuine evidence and non-evidence items, perhaps in their relative blatancy, was better understood than that of pseudo-evidence, underscoring pseudo-evidence's potential to mislead, confuse or challenge people.

This chapter argues that overconfidence plays a role in the problem of problem misrepresentation, and its frequency suggests that that role may be significant. The problem of overconfidence seems particularly acute in the face of pseudo-evidence.

Pallier et al. (2002) suggest that personality traits and cognitive ability appear to play only a small role in determining the accuracy of self-assessment and that overconfidence has multiple causes. Indeed, that finding is consistent with the evidence that suggests cognitive intelligence fails to correlate with everyday problem representation and solving abilities. The role of personality traits in everyday problem representation is ripe for investigation, however, and the chapter that follows investigates what it argues is two of them.

6. Empathy, Self-Awareness and Problem Representation

A number of researchers have suggested the importance of emotion in everyday decision-making (e.g., Damasio, 1994; DeSousa, 1987). Indeed, emotional intelligence refers in part to one's ability to solve problems having recognised the potential meaning of emotional patterns (Mayer & Salovey, 1997; Salovey & Mayer, 1990). Instances in which emotions could impact one's ability to solve a problem are common. Love, for instance, could cause a solver to distort the facts upon which a problem's representation is dependent, the solver not wanting to acknowledge a loved one's flaws.

More specifically, common sense suggests self-awareness and empathy, often cited as dimensions of emotional intelligence, could play a role in solving everyday problems. *Self-awareness*, or our understanding of ourselves and how we feel, and *empathy*, the ability to feel what others may feel, could impact our problem representations of problems involving the self and others, respectively. Therefore, this dissertation next examined self-awareness and empathy with regard to problem representation.

6.1 Emotional Intelligence

As discussed in Chapter 1, it is largely Salovey and Mayer (1990) who are credited with proposing the first theory of emotional intelligence. Again, Mayer and Salovey (1997) define *emotional intelligence* as "the ability to perceive and express emotion, assimilate emotion in thought, understand and reason with emotion, and regulate emotion in the self and others" (p. 11). This ability is to be evaluated with regard to how successfully an individual navigates their world (Salovey & Mayer, 1990), or rather how well they solve their everyday problems.

Daniel Goleman's (1995) popular book, *Emotional Intelligence*, a lay work based on Salovey and Mayer's (1990) theory, introduced emotional intelligence to the masses. At the time of the book's publication, however, little emotional intelligence research had been conducted (Goleman, 2005), and no means of measuring it existed (Salovey, 2011). Unfortunately, the heuristic value of emotional intelligence may have undermined its proper study (Barrett & Salovey, 2002).

Since 1995, a great deal of emotional intelligence research has been conducted, and several tools that profess to measure it have been developed. Yet important disagreements regarding the definition of emotional intelligence remain, and if it is unclear as to what emotional intelligence is, those measurements will lack validity. As a consequence, emotional intelligence's role in everyday problem solving is unclear, despite intentions and claims to the contrary.

Ability, trait and mixed models. In the absence of clarity regarding its definition, emotional intelligence has been conceptualised and measured in various ways (Roberts, Zeidner, & Matthews, 2001). Petrides and Furnham (2000) suggest that two predominant emotional intelligence models have evolved, *ability* and *trait*. Ability models (e.g., Mayer & Salovey, 1997) emphasise the differences amongst people in cognitively processing emotional information. Trait models (e.g., Petrides & Furnham, 2003), on the other hand, incorporate non-intelligence factors.

Trait emotional intelligence theory conceptualises emotional intelligence as a personality trait, rather than an intelligence (Petrides, Pita, & Kokkinaki, 2007). Indeed, Petrides (2009a) asks us to consider renaming trait emotional intelligence as *trait emotional self-efficacy* to sidestep a murky and perhaps ultimately irrelevant intelligence debate. As a personality trait, trait emotional intelligence does not overlap with cognitive ability (Carroll, 1993). Galotti (1989) finds that IQ correlates only

slightly with trait emotional intelligence measurements; others (e.g., Amelang & Steinmayr, 2006; Warwick & Nettelbeck, 2004) report no or negative correlations between them.

Mayer, Salovey, Caruso, and Cherkasskiy (2011) distinguish between *mixed* and ability models. Mixed models are effectively a mix of trait and ability models (e.g., Bar-On, 1997; Goleman, 1995). Incorporating personality traits, mixed models rely on self-report questionnaires to claim valid assessments of mental ability. Mental abilities, however, cannot be assessed by self-reports; self-reports assess only self-perceptions (Field, 2009).

By incorporating non-intelligence factors, trait models have come under criticism (e.g., Mayer et al., 2008, 2011), mainly due to their divorce from intelligence. The intelligence to which the criticism refers is seemingly cognitive in nature.

Whereas emotional intelligence is intended to facilitate everyday problem solving (Salovey & Mayer, 1990), however, cognitive intelligence does not (e.g., Sternberg et al., 1995). To reconcile this misalignment, emotional intelligence either has to incorporate non-cognitive intelligences or be something other than an intelligence.

In a non-argument, Mayer, Salovey, Caruso, and Cherkasskiy (2011) are reluctant to have ability emotional intelligence do either of those things due to 1) their claim that emotional intelligence is indeed an intelligence and 2) the 100-year legacy of intelligence as a relatively unmalleable construct. Trait emotional intelligence, on the other hand, makes no such claim and feels no such indebtedness.

In sum, a construct of emotional intelligence pinned to cognitive intelligence seems at odds with an objective of better everyday problem representation, and an

unmooring of emotional intelligence from cognitive intelligence is necessary. Therefore, the criticism of trait models seems unwarranted. Perhaps it is the definition of intelligence that needs to be expanded rather than the trait models discounted, and it is the ability and mixed models, with their narrower definition of intelligence, that fall short.

Measurement. As mentioned, numerous tools profess to measure emotional intelligence. The ability and mixed scales tend to be classified as either *specific-ability* or *integrative-model* measures (Mayer et al., 2011). Whereas specific-ability tests concentrate on singular dimensions of emotional intelligence, integrative-model tests focus on several dimensions simultaneously. Essentially, specific-ability tests are of two types: 1) the assessment of emotions in faces, postures and voice; and 2) the understanding and management of emotions in emotional situations. Integrative-model tests are similar in nature to specific-ability tests, they just seek to evaluate emotional intelligence more broadly.

In contrast, trait models rely on self-report questionnaires to capture personality traits. Rather than in their broader theoretical conceptualisations, Petrides (2009a) argues that ability and trait models differ primarily in their measurement approach. Those championing ability and mixed emotional intelligences contend that they, like all cognitive intelligences, employ mental abilities to solve problems with right and wrong answers (Mayer et al., 2011).

Trait emotional intelligence advocates, however, contend that the subjective nature of emotions does not make emotional intelligence amenable to IQ-type scoring and legitimises a self-report methodology (Petrides, 2009a). To ask if self-perceptions are accurate runs counter to trait theory—there can be no right or wrong rating of one's emotional levels, so there can be no over- or under-estimation of them. If one's

own ratings correlate with observer ratings, however, concordance can be interpreted as an indication of reliability. In sum, whereas ability models stress maximal performance, trait models emphasise typical performance (Petrides, 2009a).

With the definition of emotional intelligence unclear, and therefore without agreement, there is little value in comparing one tool's measurement of emotional intelligence to another's, for they may be measuring different constructs. There is value, however, in comparing the relative strengths and weaknesses of the tools. A test that is relatively reliable and valid is preferred to one that is not. Whereas trait assessments often demonstrate significant levels of reliability and validity, ability and mixed models often do not (Keele & Bell, 2008).

Faux intelligences. As described in Chapter 1, Gardner's (1983) theory of multiple intelligences suggests there are at least eight intelligences. Chapters in *The Cambridge Handbook of Intelligence* (Sternberg & Kaufman, 2011) suggest there may be several others: successful, practical, social, cultural and mating, for instance. The sheer number of potential intelligences suggests there may be something opaque and loose about intelligence's definition.

The tendency to classify almost any behaviour as an intelligence is longstanding and well-documented (Eysenck, 1998). These faux intelligences are intuitively appealing in that they reintroduce personality traits as cognitive abilities amenable to IQ-type testing, yet they are not (Furnham, 2006; Waterhouse, 2006). With their cognitive underpinnings and reliability and validity issues, ability and mixed emotional intelligences may indeed be faux intelligences.

In sum, an understanding of the impact of emotions on problem representation (and therefore solution) naturally depends upon the definition and measurement of those emotions. Research suggests that cognitive intelligence has little-to-no bearing

on one's ability to represent and solve everyday problems; unsurprisingly, it also suggests that measurements of emotion that are dependent on cognitive intelligence are relatively unreliable and invalid. Therefore, the relationship between emotions and problem representation may be better understood by investigating emotions when they are defined and measured, with significant reliability and validity, as personality traits.

6.2 Investigation 9

This study set out to explore the relationship between both self-awareness and empathy and everyday problem representation. For the reasons described, the investigation sought measures of these emotions that were independent of cognitive intelligence. Therefore, this study relied on measures of empathy and self-awareness as personality traits. Inaccurate judgements regarding evidence strength were a proxy for everyday problem misrepresentations.

Specifically, participants were asked to identify genuine evidence with regard to causal claims made in three self-focused and three others-focused everyday problem scenarios and complete standardised measures of self-awareness and empathy. It was hypothesised that the correct identification of genuine evidence regarding the self-focused scenarios will correlate positively with self-awareness, while correctness of response regarding the others-focused scenarios will correlate positively with empathy.

As in previous studies herein, it was secondarily hypothesised that 1) participants will assess the strength of the evidence correctly less than half of the time and no better than chance, and 2) correctness of response will not correlate with either age or years of undergraduate education, a null hypothesis.

6.2.1 Method

6.2.1.1 Participants

As in most of this dissertation's previous studies, participation criteria were minimised in order to capture a sample representative of the average adult: The only criterion for participation was that a participant be at least 17 years of age.

Data for 61 participants were collected between 30 April 2018 and 24 May 2018. Participants were recruited exclusively from the Birkbeck, University of London, participant database. Due to two participants' failure to provide required data, however, data from only 59 participants were usable (32 women, 27 men). As summarised in Table 6.1, M_{age} = 28.63 years, SD_{age} = 9.38 years, age range: 19 – 57 years, $M_{undergrad\ years}$ = 3.03 years, $SD_{undergrad\ years}$ = 1.36 years, and undergraduate education range: 0 – 8 years:

Table 6.1

Age and Years of Undergraduate Education Descriptive Statistics – Investigation 9

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
Age	59	28.63	9.38	17-NM	19-57	1.52	1.67
Undergrad	59	3.03	1.36	0-NM	0-8	0.65	2.86

Note. NM = not meaningful.

In this sample, using the Kolmogorov-Smirnov test, the ages of participants, $D(59) = 0.22$, $p < .001$, and their years of undergraduate education, $D(59) = 0.29$, $p < .001$, were both significantly non-normal. Participants tended to be young, and their ages and years of undergraduate education clustered around the mean.

Forty-five of the 59 participants cited English as their primary language. Participants received £7.50 for taking part and were offered a bottle of water during the study.

6.2.1.2 Materials and Procedure

Participants scheduled appointments and committed to one hour of testing time, although all participants finished with time to spare. In their testing, participants completed: 1) questions regarding evidence in support of causal claims made in six hypothetical everyday scenarios, 2) the Trait Emotional Intelligence Questionnaire Version 1.50 (Petrides, 2009b) and 3) the Empathy Quotient Questionnaire (Baron-Cohen & Wheelwright, 2004), all of which are described in detail below. The order in which participants completed these three tasks was counterbalanced. Testing occurred in designated testing rooms at Birkbeck, and data were collected on paper.

Everyday scenarios. Each participant received the same six scenarios. The scenarios can be found in Appendix J. In three of the scenarios, the participant was the focus (i.e., Scenarios 3, 4 and 5); in three of the scenarios, others were the focus (i.e., Scenarios 1, 2 and 6).

In a further effort to minimise systematic variation, the order in which participants received the scenarios was counterbalanced, so no two participants experienced the scenarios in the same order.

Each scenario described an ill-defined problem and made a claim regarding its cause. Scenario 1, for example, stated that in 2014, 13% of adults worldwide were obese. In one of the world's most developed countries, however, the obesity rate amongst adults is currently at 40%. A group of scientists in that country claim that the primary reason for this much-higher-than-average obesity rate is lack of exercise.

Participants then had to select the genuine, or standalone, evidence item or items from a list of five potential evidence items. If they believed none of the five items were genuine evidence, they could choose "None of the above," the sixth item on each list.

In Scenario 1, for example, the genuine evidence item read, "In this country, 60% of adults do not exercise. A global study shows that in countries with higher-than-average adult obesity rates, more than 50% of adults do not exercise." One of the two pseudo-evidence items read, "In yet another developed country, the adult obesity rate is 20%, and 70% of its adults do not exercise regularly." As is always the case, one instance is an insufficient basis for generalising a correlational relationship. One of the two non-evidence items read, "The news tells us that scientific reports often aren't reliable, so while this group of scientists claim lack of exercise, others will likely claim something else." In this case, the statement does not speak to the relationship between obesity rates and exercise.

Of the five potential evidence items in Scenarios 1, 2, 3, 4 and 6, one was genuine evidence, two were pseudo-evidence and two were non-evidence; in Scenario 5, all five items were pseudo-evidence. Therefore, each scenario had one correct response: the genuine evidence item in Scenarios 1, 2, 3, 4 and 6 and "None of the above" in Scenario 5. As participants were able to choose more than one item in response to each scenario, those participants who chose only the correct item were classified as correct. All other participants, meaning those who chose a single incorrect item or more than one item, were classified as incorrect.

Trait Emotional Intelligence Questionnaire (TEIQue). As discussed, with their cognitive intelligence underpinnings seemingly at odds with the objective of better ill-defined problem representation, ability emotional intelligence models were

deemed inappropriate to the research question. Given its relatively strong reliability and validity (discussed below) and evidence of its independence from cognitive intelligence, the Trait Emotional Intelligence Questionnaire Version 1.50 (Petrides, 2009b) was used to assess participant levels of trait emotional intelligence and self-awareness specifically.

A copy of the TEIQue can be found in Appendix K. The TEIQue comprises 153 items that map to 15 facets. Each TEIQue item is a Likert item. Seven responses to each item, a statement, are possible, each response being a whole number. For instance, Item 12 reads, "Others admire me for being relaxed." Each participant then chooses one of those seven responses. The response set ranges from 1, which means the participant completely disagrees with the statement, to 7, which means the participant completely agrees with the statement. Table 6.2 summarises the facets' definitions:

Table 6.2

Summary of TEIQue Facet Definitions

Facet	Definition
Adaptability	Flexible and willing to adapt to new conditions
Assertiveness	Forthright, frank, stand up for their rights
Emotion expression	Capable of communicating their feelings
Emotion management (others)	Capable of influencing other people's feelings
Emotion perception (self and others)	Clear about their own and other's feelings
Emotion regulation	Capable of controlling their emotions
Impulsiveness (low)	Reflective and less likely to succumb to urges
Relationships	Capable of maintaining fulfilling relationships
Self-esteem	Successful and self-confident
Self-motivation	Driven and unlikely to give up
Social awareness	Accomplished networker, superior social skills
Stress management	Ability to withstand pressure, regulate stress
Trait empathy	Ability to take someone else's perspective
Trait happiness	Cheerful and satisfied with their lives
Trait optimism	Confident and with a positive disposition

Note. Adapted from the Technical Manual for the Trait Emotional Intelligence Questionnaires (TEIQue), by K. V. Petrides, 2009, p. 14. Copyright 2009 by K. V. Petrides.

The TEIQue is hierarchical in nature, and the 15 component facets map onto four factors that in turn map onto a global trait emotional intelligence construct. Table 6.3 summarises this mapping:

Table 6.3

TEIQue Factor-Facet Mapping

Factor	Facet
Emotionality	Emotion perception (self and others) Trait empathy Emotion expression Relationships
Self-control	Emotion regulation Stress management Impulsiveness (low) Adaptability Self-motivation
Sociability	Assertiveness Emotion management (others) Social awareness
Well-being	Self-esteem Trait happiness Trait optimism

Note. Adapted from the Technical Manual for the Trait Emotional Intelligence Questionnaires (TEIQue), by K. V. Petrides, 2009, p. 23. Copyright 2009 by K. V. Petrides.

With the exception of self-esteem, which is outside of the interests of this study, the facets have high loadings on their respective factors exclusively (Petrides, 2009a), and the factor structure is clear and replicable (Freudenthaler, Neubauer, Gabler, Scherl, & Rindermann, 2008; Mikolajczak, Luminet, Leroy, & Roy, 2007; Petrides, 2009a). In support of the hierarchical structure of trait emotional intelligence, the four factors are also inter-correlated, implying that while they may be mutually exclusive, the factors can co-exist within an individual (Petrides, 2009a).

The TEIQue is scored at the facet, factor and global trait emotional intelligence levels. Given the scope of this study, ultimately only two measures of

trait emotional intelligence were considered: trait empathy and emotion perception. Further, as empathy is the exclusive focus of Baron-Cohen and Wheelwright's (2004) Empathy Quotient Questionnaire, presented below, emotion perception was of primary interest, it being the facet most closely aligned to self-awareness, albeit not exclusively.

Empathy Quotient Questionnaire (EQQ). The Empathy Quotient Questionnaire – Adult Version (Baron-Cohen & Wheelwright, 2004) is also a self-report measure rather than an ability measure. In its development, Baron-Cohen and Wheelwright (2004) reviewed other methods of measuring empathy and found them wanting for a variety of reasons, the possible exception being the Interpersonal Reactivity Index (Davis, 1980). Most, however, were rightly criticised for attempting to measure more than empathy (Muncer & Ling, 2006).

As a consequence, the EQQ was used to measure participants' empathy levels specifically. A copy of the EQQ can be found in Appendix L. The EQQ consists of 40 statements, and participants chose one of four responses to each: "Strongly Agree," "Slightly Agree," "Slightly Disagree," or "Strongly Disagree." Statement 5, for instance, reads, "People often tell me that I went too far in driving my point home in a discussion."

Scoring direction was counterbalanced. Participants received two points for a response of "Strongly Agree" and one point for a response of "Slightly Agree" for each of the following statements: 1, 3, 11, 13, 14, 15, 21, 22, 24, 26, 27, 28, 29, 34, 35, 36, 37, 38, 39 and 40. Participants received two points for a response of "Strongly Disagree" and one point for a response of "Slightly Disagree" for each of the remaining statements.

The sum of the points received for each response is the total EQQ score. An EQQ score of 0-32 is considered low, 33-52 is considered average, 53-63 is considered above average, and 64-80 is considered very high, 80 being the maximum possible score.

6.2.2 Results

6.2.2.1 Descriptive Statistics

Correct responses by scenario. Table 6.4 summarises the frequency of correct and incorrect responses by scenario:

Table 6.4

Frequency of Correct Responses by Scenario – Investigation 9

Scenario	Correct	% Correct
1, o	14	24
2, o	11	19
3, s	23	39
4, s	19	32
5, s	7	12
6, o	13	22

Note. $N = 59$; s = self-focused, o = other-focused.

Across the six scenarios, an average of 25% of participants responded correctly, Scenario 5 having the minimum number of correct responses (12%) and Scenario 3 having the maximum number (39%). Both of those scenarios were self-focused scenarios.

In sum, both in total and by scenario, the vast majority of participants responded incorrectly. There was no notable difference in scenarios in which the self was the focus versus others: In scenarios in which oneself was the focus, participants

were correct 28% of the time; in scenarios in which others were the focus, participants were correct 22% of the time. Across the scenarios, there were 15 pseudo-evidence items versus 10 non-evidence items; pseudo-evidence was chosen 246 times, however, more than twice the number of times non-evidence was chosen, 113 times.

TEIQue scores. As shown in Table 6.5, the study mean scores for global trait emotional intelligence, emotionality, emotion perception and trait empathy were lower than Petrides' (2009b) standardisation sample's means, but not markedly so:

Table 6.5

TEIQue Study Sample Versus Standardisation Sample Descriptive Statistics – Investigation 9

					Range			
Level	*N*	*M*	*SD*	α	Potential	Actual	Skew	Kurtosis
GTEI								
Study	59	4.58	0.65	NA	1-7	3.2-6.2	0.16	0.08
Standard	1,712	4.90	0.59	.90	1-7	NA	-0.11	-0.17
Emotionality								
Study	59	4.73	0.74	NA	1-7	2.7-6.4	0.25	0.32
Standard	1,712	5.05	0.71	.78	1-7	NA	-0.27	-0.16
Perception								
Study	59	4.64	0.87	NA	1-7	2.3-6.6	-0.26	0.21
Standard	1,712	4.84	0.81	.73	1-7	NA	-0.37	0.26
Empathy								
Study	59	4.83	0.92	NA	1-7	1.0-6.8	-0.85	4.29
Standard	1,712	5.12	0.77	.87	1-7	NA	-0.33	0.01

Note. NA = not available. Adapted from the Technical Manual for the Trait Emotional Intelligence Questionnaires (TEIQue), by K. V. Petrides, 2009, p. 19. Copyright 2009 by K. V. Petrides.

In the standardisation sample, the emotionality factor, trait empathy facet, emotion perception facet and global trait emotional intelligence scores are normally distributed (Petrides, 2009b). Using the Kolmogorov-Smirnov test, the TEIQue scores of relevance to this study were also normally distributed: emotion perception facet scores, $D(59) = 0.09$, $p > .05$; trait empathy facet scores, $D(59) = 0.10$, $p > .05$. Unfortunately, however, these scores are not parametric in nature, as will be discussed below.

EQQ scores. The descriptive statistics for the EQQ scores are summarised in Table 6.6:

Table 6.6

Empathy Quotient Questionnaire Score Descriptive Statistics – Investigation 9

				Range			
Variable	*N*	*M*	*SD*	Potential	Actual	Skew	Kurtosis
EQQ Score	59	40.14	12.10	0-80	11-63	-0.08	-0.59

In this sample, using the Kolmogorov-Smirnov test, EQQ scores were normally distributed, $D(59) = 0.08$, $p > .05$. As mentioned earlier, an EQQ score of 0-32 is considered low, 33-52 is considered average, 53-63 is considered above average, and 64-80 is considered very high, 80 being the maximum possible score (Baron-Cohen & Wheelwright, 2004). When this sample's scores are categorised, while no participant scored very high, 10 scored above average, 30 had average scores, and 19, or nearly a third, had low scores.

While normally distributed, these scores, like the TEIQue scores, are not parametric in nature, as will be discussed below.

6.2.2.2 Inferential Statistics

While the TEIQue and EQQ generate scores for each participant, the intervals on their measurement scales do not represent equal differences in the property being measured. In the TEIQue, for instance, a 3 to one participant on the scale from 1, "Completely Disagree," to 7, "Completely Agree," could be a 4 to another participant, as the scale's degrees of agreement are not well defined. Similarly, in the EQQ, "Strongly Agree" to one participant could be "Slightly Agree" to another. In other words, while scores in general are often treated as continuous in nature, the TEIQue and EQQ scores are not continuous, but ordinal. Therefore, while the TEIQue and EQQ scores are normally distributed in this study, they do not lend themselves to parametric testing.

In any case, the predictive validity shortcomings inherent in any trait emotional intelligence model recommend correlation over regression. TEIQue and EQQ assess self-perceptions. While one's feelings may have some predictive ability with regard to their own future, the uniqueness of those feelings and their context undermine greater predictive powers. Therefore, investigating a predictive relationship between TEIQue and EQQ scores for a number of people and their ability to identify genuine evidence in support of causal claims, the sort of relationship that regression seeks to study, becomes meaningless. Correlation, on the other hand, simply investigates the co-existence of variables; no predictive relationship between them is implied.

As a consequence, participants' TEIQue scores were rounded to the nearest whole number. Scores of 1 and 2 were then classified as below average; 3, 4 and 5 as average; and 6 and 7 as above average. EQQ scores were also categorised as below

average (i.e., scores of 0-32), average (i.e., scores of 33-52) or above average (i.e., scores of 53-63).

The scenario correctness of response variable is categorical by nature, responses being either correct or incorrect.

Correctness of response and EQQ scores—association. Fisher's exact test was performed with regard to correctness of response and EQQ scores for Scenarios 1, 2 and 6, those in which others were the focus. Again, of the 59 participants, 19 had below-average scores (i.e., scores of 0-32), 30 had average scores (i.e., scores of 33-52) and 10 had above-average scores (i.e., scores of 53-63). Analysis of response correctness by EQQ categories suggested no relationship between the variables in Scenarios 1, 2 and 6, $p = .386$, $p = .196$ and $p = .335$, respectively.

If the EQQ score had been treated as a continuous variable, logistic regression also suggested that there was no correlation between correctness of response and EQQ scores. The findings for each scenario have been summarised in Table 6.7:

Table 6.7

Logistic Regression: Scenario 1, 2 and 6 Response Correctness and EQQ Scores – Investigation 9

				95% CI for Odds Ratio		
	B (SE)	Wald (1)	*p*	Lower	Odds	Upper
Scenario 1						
Constant	-0.76 (1.05)					
EQQ	-0.01 (0.03)	0.16	.685, *ns*	0.94	0.99	1.04
Scenario 2						
Constant	-1.90 (1.20)					
EQQ	0.01 (0.03)	0.14	.707, *ns*	0.96	1.01	1.07
Scenario 6						
Constant	-2.31 (1.18)					
EQQ	0.03 (0.03)	0.89	.346, *ns*	0.97	1.03	1.08

Note. ns =not significant.

Using the TEIQue's trait empathy scores instead of the EQQ merely confirms the absence of correlation between empathy and response correctness in the others-focused scenarios.

Correctness of response and TEIQue emotion perception scores—association. In light of low expected frequencies, Fisher's exact test was performed with regard to correctness of response and TEIQue emotion perception (i.e., self-awareness) scores for Scenarios 3, 4 and 5, those in which the self was the focus. Of the 59 participants, 1 had a below-average score (i.e., a score of 1 or 2), 49 had average scores (i.e., scores of 3, 4 or 5), and 9 had above-average scores (i.e., scores of 6 or 7). Analysis of response by emotion perception category frequencies for

Scenarios 3, 4 and 5 suggest no association between the variables, $p = .561$, $p = 1.000$, and $p = .379$, respectively.

If the TEIQue emotion perception score had been treated as a continuous variable, logistic regression also suggested that there was no correlation between correctness of response and TEIQue emotion perception scores in all three scenarios. The findings for each scenario have been summarised in Table 6.8:

Table 6.8

Logistic Regression: Scenario 3, 4 and 5 Response Correctness and TEIQue Emotion Perception Scores – Investigation 9

				95% CI for Odds Ratio		
	B (SE)	Wald (1)	*p*	Lower	Odds	Upper
Scenario 3						
Constant	-2.65 (1.58)					
Emotion Perception	0.47 (0.33)	2.04	.153, *ns*	0.84	1.60	3.06
Scenario 4						
Constant	-0.98 (1.53)					
Emotion Perception	0.05 (0.32)	0.03	.874, *ns*	0.56	1.05	1.98
Scenario 5						
Constant	-4.18 (2.44)					
Emotion Perception	0.46 (0.49)	0.85	.356, *ns*	0.60	1.58	4.14

Note. ns =not significant.

Correctness of response—frequency. To assess how participants responded to the scenarios correctly versus chance, a one-sample Pearson's chi-square test was performed. As the data were independent in nature, and the expected frequency of

each outcome was greater than 5 (see Table 6.9), the assumptions necessary for performing the analysis were met.

Correctness of response was a categorical variable with six possible outcomes as there were six list items in each scenario. A selection of the one item from each list that was correct (i.e., the genuine evidence item in Scenarios 1, 2, 3, 4 and 6; and "None of the above" in Scenario 5) constituted a correct response; all other responses are incorrect. As in Investigation 2, as participants were able to choose more than one list item, the single correct item in effect prescribed the odds: If there is a one-out-of-six chance of a participant choosing the correct response, there must be a five-sixths chance of a participant giving an incorrect response. The observed versus expected outcomes are summarised in Table 6.9:

Table 6.9

Observed Versus Expected Correctness of Response by Scenario – Investigation 9

	Correct			
Scenario	*n* Observed	*n* Expected	$X^2(1)$	*p*
1, o	14	9.83	2.10	.148
2, o	11	9.83	0.16	.689
3, s	23	9.83	21.06	< .001
4, s	19	9.83	10.19	.001
5, s	7	9.83	0.99	.319
6, o	13	9.83	1.21	.272

Note. s = self-focused, o = others-focused.

For Scenarios 3 and 4, both self-focused scenarios, there was a significant tendency for participants to rate items correctly and with large effect sizes: $r = .96$ and $r = .73$, respectively.

Correctness of response versus age and years of undergraduate education—correlations. It was secondarily hypothesised that there would be no correlation between correctness of response and either age or years of undergraduate education, a null hypothesis. Once again, Bayesian statistics were used to analyse the extent to which the data increased our confidence in this null hypothesis. The results are summarised in Table 6.10:

Table 6.10

Bayesian Pearson Correlations: Scenario Correctness of Response by Age and Years of Undergraduate Education – Investigation 9

		Scenario					
		1	2	3	4	5	6
Age	BF_{10}	0.31	0.46	0.42	0.66	0.27	0.17
Undergrad	BF_{10}	0.65	0.38	0.43	1.49	0.19	0.17

As in the previous studies herein, Bayesian Pearson correlations between correctness of response and both age and years of undergraduate education largely suggest evidence in moderate-to-anecdotal support of no correlation between the variables, the exception being between correctness of response and years of undergraduate education in Scenario 4, where evidence is in only anecdotal support of the alternative hypothesis.

6.3 Discussion and Conclusion

Summary of results. Counter to its primary hypothesis, this investigation's Fisher's exact test findings suggest that participants' levels of empathy and self-awareness do not correlate with their ability to identify genuine evidence in support of causal claims in situations regarding others and the self, respectively.

In line with this investigation's secondary hypotheses, however, participants assessed the strength of the evidence correctly only 25% of the time overall, less than half of the time in each scenario and, with the exception of Scenarios 3 and 4, no better than chance. While Scenarios 3 and 4 are self-focused, the study's primary findings suggest that their better-than-chance results are unrelated to participants' self-awareness, at least as it is measured by the TEIQue's emotion perception scores, which perhaps speaks to one of the study's limitations, as discussed below.

In addition, Bayesian Pearson correlations largely support the null hypotheses of no relationships between scenario correctness of response and either participants' ages or years of undergraduate education. Only with regard to Scenario 4, a self-focused scenario, was evidence suggested to be in anecdotal support of a relationship between correctness of response and participants' years of undergraduate education. As the evidence was merely anecdotal, however, and the scenario concerned whether or not you should pay for hypnosis in an attempt to stop smoking, a subject unlikely to correlate with years of undergraduate education, the relationship may simply be the result of a Type I error.

Regardless, the TEIQue's, EQQ's and scenario's reliabilities and certain of their validities (discussed below) suggest that these findings have value. All three tools seem to possess reasonable levels of reliability, and while their ecological and criterion, both predictive and concurrent, validities are harder to argue, reasonable cases can be made for their content and construct validities.

Reliability and validity—scenarios. Of the 59 participants, 55 (93%) responded incorrectly and 27 (46%) responded correctly to three or more, or at least half, of the scenarios. In other words, an average of 70% of participants responded consistently to three or more scenarios, although far more responded incorrectly than

correctly. As it was the same group of participants being tested across all six scenarios, once with regard to each scenario, the relative consistency of participant responses between scenarios suggests the scenarios possessed a reasonable level of reliability.

As in many of the previous studies comprising this dissertation, the scenario's correct and incorrect responses suggest they possessed reasonable construct and content validities. Again, however, external validity, is more difficult to ascertain. The hypothetical nature of the scenarios made ecological validity questionable; the average adult sample, on the other hand, helps the case for population validity.

Reliability and validity—TEIQue. TEIQue's reliability was assessed by internal consistency. Cronbach's alpha is the most common measure of internal consistency (Field, 2009). In general, alphas above .70 to .80 are acceptable, and markedly lower values suggest an unreliable scale. Kline (1999) goes on to suggest that while .80 is appropriate for cognitive tests, such as intelligence tests, .70 is more appropriate for tests of psychological constructs given their diversity.

Based on a sample comprised of 1,721 individuals (see Table 6.5—912 female, 764 male, 61 unreported; M_{age} = 29.65 years, SD_{age} = 11.94 years, range 16 – 77 years) of primarily White UK and White European origins (58% and 19%, respectively), the alphas for the trait empathy and emotion perception scores are .77 and .81, respectively, suggesting that they possess acceptable levels of internal consistency and therefore are reliable (Petrides, 2009b).

In addition, test-retest reliability for emotion perception seems significantly stable. The stability coefficient for emotion perception is .66, $p < .01$ (Petrides, 2009b). Personality traits in normal adults have been shown to be relatively stable, ranging from .6 to .8 (Terracciano, Costa, & McCrae, 2006). As trait emotional

intelligence has been construed as such a trait, similar stability is expected. Whereas the trait empathy stability coefficient is .19, trait empathy was not of primary consideration in this study.

Trait emotional intelligence as measured by TEIQue appears to be reasonably well defined with meaningful degrees of construct, incremental and discriminatory validity over other measures (Freudenthaler et al., 2008; Mikolajczak et al., 2007; Petrides, 2009b). Criterion validity, however, is more difficult to substantiate. While a case for concurrent validity is seemingly supported (for a summary see Petrides, 2009b), the inherently subjective nature of emotions and the self-report nature of TEIQue limit its predictive validity. One's emotional state is unique as is one's own perception of it. Therefore, while the TEIQue may possess some predictive validity with regard to the individual who has taken the questionnaire, it is more difficult to generalise their findings to others. In addition, TEIQue's questionnaire-based nature undermines its ecological validity.

Reliability and validity—EQQ. Baron-Cohen and Wheelwright (2004) established that the EQQ possessed reasonable levels of reliability and validity during its development. Their initial subjects were asked to fill in the EQQ for a second time 12 months after first taking the questionnaire as a test of its retest reliability. Test-retest reliability for the EQQ was $r = 0.97$ and highly significant ($p < .001$) (Baron-Cohen & Wheelwright, 2004). As a measure of internal consistency, Cronbach's alpha was also calculated to be 0.92, which is high (Baron-Cohen & Wheelwright, 2004).

In the design of the EQQ itself and with an objective of internal validity, namely construct validity, Baron-Cohen and Wheelwright (2004) sought to ensure that their assessment of the 40 EQQ items as good tests of empathy was not

subjective. To do so, they provided a definition of empathy to a panel of six judges, all of whom were experimental psychologists working in this field and asked them to rate whether each of the 40 items related to the definition of empathy. The definition read, "Empathy is the drive or ability to attribute mental states to another person/animal and entails an appropriate affective response in the observer to the other person's mental state." Results showed that all 40 items were rated as being related to empathy (and all 20 filler items were correctly identified as being unrelated to empathy) by at least five out of six judges. The probability of obtaining such agreement on each item by chance is $p < .003$.

Muncer and Ling (2006) further tested construct validity by using confirmatory factor analysis to evaluate EQQ fit versus a number of competing models of empathy. As discussed previously, most other models arguably define empathy too broadly, therefore, Muncer and Ling's (2006) fit findings support not only superior construct validity, but incremental and discriminant validities as well. The authors claim that the fit is perhaps particularly impressive considering the inherent difficulties in both factor analysing ordinal level data (Gorsuch, 1974) and confirming the structure of ordinal level data (Green, Akey, Fleming, Hershberger, & Marquis, 1997).

As discussed regarding the TEIQue, however, the use of self-report methods introduces certain limitations. Empathy can vary, for instance, with a person's current state—your own emotional state may inhibit your ability to see another person's perspective in an argument, so your view will dominate (Baron-Cohen & Wheelwright, 2004). The EQQ only assesses an individual's beliefs about their own empathy or how they might like to be seen or think about themselves, but this may

differ from how empathic they truly are. As a consequence, it becomes harder to argue in defence of the EQQ's criterion validity, both predictive and concurrent.

Limitations. Common sense and experience suggest that emotion, self-awareness and perhaps other emotions should matter to everyday problem representation and solution. Therefore, the perhaps unexpected finding that levels of empathy and self-awareness do not correlate with the ability to identify genuine evidence in support of claims regarding others and the self, respectively, may in part be due to measurement issues. Ostensibly, empathy and self-awareness are being measured by the EQQ and the TEIQue's emotion perception facet scores, respectively. With regard to self-awareness, however, emotion perception is only the TEIQue facet most closely aligned with self-awareness, and its definition is actually broader. Perhaps trait emotion perception is not defined specifically enough for it to be an effective proxy for self-awareness.

In addition, perhaps the identification of genuine evidence, while essential to problem representation, is not a function of empathy and self-awareness. Maybe self-awareness and empathy do matter to problem representation, just not in the identification of evidence in support of causal and correlational claims. A relationship between emotions and the identification of evidence is not necessarily intuitive, particularly when the matter in question is not emotionally charged. As mentioned earlier, love may cause a solver to neglect seeing a loved one's flaw, but the same sort of reaction might be hard to conjure regarding a hypothetical situation.

On a related note, while the Scenarios 3, 4 and 5, have been considered self-focused by suggesting that the problems in question is the participants', needless to say, it is a distinction in name only. In Scenario 3, for example, the use of the pronoun *you* implies that it is the participant that is trying to lose weight and is considering a

specific diet. The situation, however, remains hypothetical, and it is not the participant that is trying to lose weight and considering a specific diet, of course. As a consequence, participants' level of self-awareness likely will not matter to this problem or its representation.

Implications. While certain measurement and validity issues must be considered, perhaps it is the sheer number of incorrect assessments of evidence in support of causal claims that is most important in this study. In Investigation 9, participants assessed the evidence strength incorrectly 75% of the time, and their accuracy differed little between the scenarios focused on the self and others.

Emotion perception scores, as well as trait empathy scores, under TEIQue and EQQ scores were normally distributed. When correctly categorised as ordinal variables, however, they thankfully still reflected a "normal" sample, or one in which most participants had average or above average levels of empathy and self-awareness. Only six participants had a low level of self-awareness, and 53 had an average or above level. With regard to empathy, 19 participants had a low level, and 40 had an average level or above.

Yet as participants were incorrect in their assessments of evidence strength most of the time, no relationships between response correctness and these aspects of trait emotional intelligence, if you will, could be revealed.

In addition, consistent with the findings of Investigation 8 and Kuhn's 1991 findings, those incorrect responses cited pseudo-evidence more than twice as often as non-evidence. Participants' confusing pseudo-evidence for genuine evidence here is likely further testament to participants' misunderstandings of the correlational and causal relationships underlying them.

In perhaps an act of foreshadowing, the frequency of incorrect responses regarding evidence in support of ultimately correlational relationships has been largely consistent throughout this dissertation. What that frequency is foreshadowing will be discussed in the following, and last, chapter.

7. CONCLUSION

This dissertation began by suggesting that we often study problem solving, but we neglect to study the problems being solved. We either assume that our everyday problem representations are correct or, perhaps more likely, do not think about their representations much, if at all. Either way, we overlook that which is most central to a problem's solution: its representation.

Indeed, the research presented herein, corroborated by the work of others, suggests that we misrepresent the problems that require us to engage, our System 2 problems (Kahneman, 2011), if you will, more often than not. This suggestion runs counter to Nisbett and Ross's (1980) unsubstantiated claim that people manage their lives relatively well. As discussed below, Levitt's (2016) claim that people may make better decisions regarding important matters by tossing a coin becomes less surprising too.

Just as we have overlooked representations—at times, perhaps in eagerness or desperation for solution, but probably just not knowing any better—we have overlooked that which may be most central to them. In our search for factors that explain the phenomenon of problem misrepresentation, we failed to consider correlation and argumentation themselves.

7.1 Reliability and Validity

As suggested in Chapter 2, the value of any research is first and foremost a function of its reliability and validity (Dunbar, 2005). Again, while its value increases in proportion to both properties, research must be reliable for validity to matter (Field, 2009).

Each of the studies comprising this dissertation seems to possess a reasonable level of reliability, namely test-retest reliability with agreements ranging from 56% to

97%. In Study 1, however, inter-rater reliability was measured instead with Cohen's kappa at .94, and in Study 9, internal consistency was measured for the TEIQue with Cronbach's alpha at .81.

With reasonable reliabilities, validities became relevant, and the strengths and weaknesses by type were largely consistent across the studies. On one hand, external validities were mixed. In seeking average adult participants, population validity was strengthened, but that strength was undermined by a heavy reliance on people with undergraduate educations. Ecological validity was mixed too—while weakened by the use of hypothetical and written scenarios as well as questionnaires, the incomplete and unorganised natures of much of the data presented simulated two important aspects of most everyday problems.

Internal validities, on the other hand, seemed stronger. While the problems described were ill-defined, which tend to be relatively weak from a validity perspective, participants were not asked to solve the problems. Instead, they were asked to assess the correlational relationships in question and give correct or incorrect answers. As a consequence, construct and content validities were relatively strong. Indeed, construct validity (Cronbach & Meehl, 1955) is central to validity as a test cannot measure what it intends to measure if it does not know what it is intending to measure. Criterion validity seemed relatively strong across the studies, too, with nothing seeming to jeopardise the predictive and concurrent validities comprising it.

In short, with reasonable reliabilities, relatively strong internal validities and mixed external validities, this dissertation's studies seem to be of value. As such, we can now consider its overall findings.

7.2 Summary of Findings

To recap, Investigation 1, a naturalistic study, found an overwhelming absence of argument in *The New York Times* and *The Guardian* readers' comments. Claims were often made, but they were infrequently supported by evidence. While genuine evidence is necessary to transform claims and counterclaims from opinion to argument, evidence quality was not considered in this first study.

In light of argumentation's importance to the ascertainment of truth, or accuracy, and its importance, in turn, to problem representation, that absence of evidence and therefore argument suggested frequent everyday problem misrepresentation. Investigation 1's participants were not asked to argue, however, so its findings are rather tenuous and served primarily as a catalyst for more rigorous investigation.

Given argumentation's dependency on genuine evidence, Investigation 2 examined participants' ability to identify genuine evidence versus pseudo-evidence and non-evidence—in other words, evidence quality mattered. Then, as a piece of evidence's quality is determined by the strength of the correlation between the antecedent and outcome underlying it, Studies 3 and 4 investigated participants' ability to discern correlational relationships. Study 3 relied on incomplete and unorganised frequency data, or how data is typically encountered in life; Study 4 relied on complete and organised frequency data.

Across Studies 2 – 4, while sometimes performing at better than chance on a scenario basis, on average a majority of participants failed to identify genuine evidence (Study 2) and discern correlational relationships (Studies 3 and 4), even when responding correctly for unsound reasons. When those correct responses made

for unsound reasons were reclassified as incorrect, error rates increased dramatically. Regardless, frequent problem misrepresentation was inferred.

The rest of this dissertation's studies then sought to explain the phenomenon of ill-defined problem misrepresentation. Investigation 5 looked at the impact of information regarding the importance of problem representation to solution, a topic not usually taught in school or university, on correlational relationship judgments. Investigations 6 and 7 investigated the reasons most common to Study 3 and 4 participants' correlational relationship misjudgements. Investigation 6 looked at the impact of information regarding the problem of emphasis on irrelevant factors in these judgements; Investigation 7 looked at the impact of information regarding the problem of over- and under-emphasis on relevant factors. Study 8 examined the frequency of overconfidence in seeking to explain correlation misjudgements, and finally, Study 9 investigated the impact of empathy and self-awareness on these misjudgements.

With the exception of Investigation 6, in Investigations 5 – 9, while occasionally performing better than chance on an item or scenario basis, by and large participants assessed evidence strength or discerned correlational relationships correctly no better than chance and less than half of the time. In Investigation 6, participants judged the correlational relationship in question correctly 57% of the time. As discussed, however, when response soundness was considered in Studies 3 and 4, as well as Study 7, the number of incorrect responses increased dramatically, suggesting that Investigation 6's findings may ultimately conform.

In addition, in those studies in which participants were asked to discern genuine evidence from non-genuine evidence, namely Studies 2, 8 and 9, pseudo-evidence was often mistaken for genuine evidence, in line with Kuhn's (1991)

findings. While pseudo-evidence was cited as genuine evidence a minority of the time and less frequently than non-evidence and "None of the above" in Study 2, participants were overconfident in pseudo-evidence 62% of the time in Study 8 and chose it more than twice as often as non-evidence as evidence in support of causal claims in Study 9.

7.3 Limitations

As discussed, as representations are unobservable, their existence must be inferred. Therefore, the studies in this dissertation studied proxies for everyday problem representations rather than representations themselves.

This dissertation relied on the incidence of poor argumentation, as determined by either the absence of counterclaims or genuine evidence in support of claims, and correlational relationship misjudgements, which determine evidence quality, as proxies for everyday problem misrepresentation. Perhaps needless to say, these proxies, however, also differ from problem misrepresentations. With regard to empathy, for instance, it is possible that empathy levels correlate with representation quality, just not with one's ability to discern genuine evidence.

As external validities were mixed, they also imply limitations and therefore opportunity for improvement. Less reliance on undergraduates, for instance, would enhance population validity. Ecological validity, on the other hand, proves more challenging to improve. While the focus on ill-defined problems and largely incomplete and unorganised frequency data lends ecological validity, the hypothetical and written natures of the problems studied undermine it. The studies herein tried to address these challenges by focusing on only aspects of these problems rather than on the problems in their entirety and their solution.

7.4 Implications and Final Words

In its review of previous research, this dissertation cited research that suggests background knowledge, domain expertise and context matter to everyday problem representation (see Novick & Bassok, 2005) and that cognitive intelligence, "emotional intelligence" and post-primary education do not. Studies 5 – 9 herein suggest that knowledge of the importance of problem representation to problem solving, an emphasis on irrelevant factors, the over- and under-emphasis on relevant factors, overconfidence, self-awareness and empathy also do not matter. To draw these or similar conclusions, however, is premature.

We cannot measure covariance between variables when at least one of the variables is not varying. In Investigation 1, for instance, topic relevance cannot vary with instances of argument if relatively few participants are constructing arguments. Similarly, age and years of university education cannot vary with any of the proxies for misrepresentation—failures to argue, misassessments of evidence and incorrect evaluations of correlational relationships—if those proxies are not varying. In other words, we cannot identify factors impacting problem representations if people are misrepresenting their problems more often than not.

7.4.1 The Significance of Chance

While performance significantly different from chance suggests engagement and something other than random behaviour, it is the sheer number of failures to identify correlational relationships, discern genuine evidence and create arguments that should be of primary interest. In addressing many if not most of the everyday problems that require engagement, we typically seek to outperform chance. As stated

earlier, it would be inappropriate to determine whether or not to have a child, for instance, by tossing a coin.

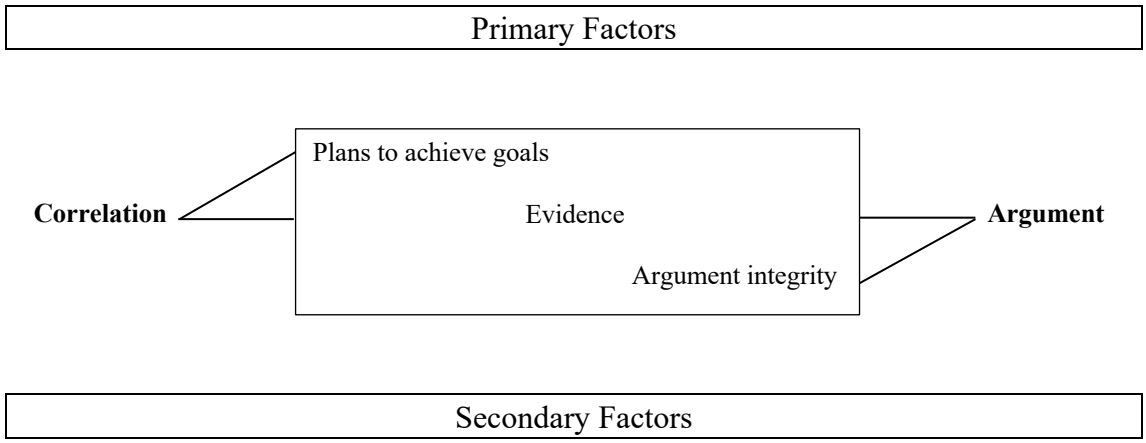
Yet Levitt (2016) suggests that people would be better off doing just that—he suggests chance has become the goal, a ceiling when it should be a floor. In failing to represent our problems correctly, however, solving them at chance becomes the best we should hope to do. Therefore, when we find ourselves representing our problems at chance or worse or simply just more than half of the time, we need to understand why if we are to improve. Progress, after all, is a function of the number of problems solved.

7.4.2 What Matters to Problem Representation—Revisited.

If we cannot identify factors that may impact our problem representations because most problems are being misrepresented, nothing may matter more than people's poor understandings of correlation and argumentation themselves. In other words, understandings of correlation and argumentation are primary problem representation factors—they are not just proxies, but building blocks; other factors, whatever they may be, become secondary. Figure 7.1 summarises this hierarchy:

Figure 7.1

Problem Representation Hierarchy



The frequency of misrepresentation that overwhelmed the studies in this dissertation, however, also calls into question others' claims of what matters to problem representation. If most problems are being misrepresented, how were they able to identify the factors they claim matter? In examining the underlying studies, their reliabilities and validities specifically, the question is justified regardless.

Some of the studies relied on well-defined problems, yet for the most part, as discussed, knowledge of well-defined problems has failed to enhance our understanding of ill-defined problems (Galotti, 1989). Chase and Simon (1973), for instance, in their study of expertise, rely on the well-defined game of chess to test their hypotheses. Similarly, Hayes and Simon (1977) and Kotovsky et al. (1985) rely on the well-defined Tower of Hanoi problem in their investigations of context. Often the issue of reliability (and therefore validity) simply goes unaddressed, yet reliability is not apparent (e.g., Chase & Simon, 1973; Gick & Holyoak, 1980, 1983; Hayes & Simon, 1977).

At present, however, a continued critique of the research regarding what matters to everyday problem representation is without merit as is research that seeks to identify such factors. It is only once the average adult better understands correlational relationships and argumentation to become better at representing their problems that the frequency of misrepresentations may cease to overwhelm such studies, and the roles of secondary factors may surface. It is not that none of the factors investigated herein and elsewhere matter to problem representation—in fact, all or some of them very well may, as intuition suggests—it is just that nothing matters more to our problem representations than an understanding of correlational relationships and argumentation.

7.4.3 Education

As most problems in life are ill defined, the main goal of education should be to prepare students to solve them (Leighton, Rogers, & Maguire, 1999; Perkins, 1986). Of course, in order to become better at solving life's problems, people must first become better at representing them. If at present, however, post-primary education does not matter or matters very little to everyday problem representation and solution, our understanding of correlational relationships and argumentation included (e.g., Kuhn et al., 1985; Perkins, 1985), perhaps general education is ripe for change.

Once that understanding is achieved, its application becomes of equal importance. It is in its application that people identify genuine evidence, argue soundly, ascertain truths and, finally, represent their problems, rather than misrepresent them unwittingly. That understanding also allows them to create plans, or a series of actions with likely consequences, that will allow them to accomplish

their intended goals as well as better evaluate the plans of others to, as Carl Sagan (1997) might have said, detect the baloney.

7.4.4 Final words

So, Einstein and Infeld (1938) were not quite right over 80 years ago when they said, "The formulation of a problem is often more essential than its solution …" (p. 92). Over 30 years later and nearly 50 years ago, Daniel P. Moynihan (1970), a now-deceased United States Senator from New York, wrote an article for *The New York Times* entitled, "To Solve a Problem, One Must First Define It." That definition and Einstein and Infeld's "formulation," of course, refer to a problem's representation, and it *always* will be more essential than its solution.

As mentioned in Chapter 1, our problem of problem misrepresentation, or our problem with problems, is also an ill-defined problem. While the problem's goal of sufficiently accurate problem representations is clear, the problematic situation and therefore the paths to solution have not been. This dissertation suggests that the primary problematic situation is a lack of understanding with regard to correlations and argumentation, and the path to solution includes education to better these understandings.

Better understandings of correlations and argumentation alone, however, may not allow us to sufficiently represent our problem of problem misrepresentation so that we may solve it. Hopefully, though, they will lead to enough improvement in our representations that we can then investigate secondary factors.

The importance of solving this problem of problem misrepresentation cannot be understated. Virtually every problem is an ill-defined problem, from the latest humanitarian crisis to one's job dissatisfaction, from Brexit to a personal relationship in trouble, and in their ill definition, all ill-defined problems are susceptible to

misrepresentation. Therefore, it is only by solving our problem with problems and thereby reduce the frequency of misrepresentation that we enable ourselves to solve so many of our other problems.

A problem's sufficient representation, while a necessary step, is only the first, however—it enables solution, but it does not guarantee it. Solving a problem depends on a variety of essentially secondary factors, including the solver's courage and will to solve it. And neither a problem's representation or solution matter until one first assumes responsibility for their problems.

References

Agosta, S., Pezzoli, P., & Sartori, G. (2013). How to detect deception in everyday life and the reasons underlying it. *Applied Cognitive Psychology*, *27*(2), 256–262. https://doi.org/10.1002/acp.2902

Alloy, L., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112–149.

Amelang, M., & Steinmayr, R. (2006). Is there a validity increment for tests of emotional intelligence in explaining the variance of performance criteria? *Intelligence*, *34*(5), 459–468.

Argument. (2016). *The American heritage dictionary of the English language* (5th ed.). Retrieved from https://ahdictionary.com

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, *39*(1), 133–144. https://doi.org/10.1016/0749-5978(87)90049-5

Bar-Hillel, M. (1983). The base rate fallacy controversy. *Advances in Psychology: Decision Making Under Uncertainty: Cognitive Decision Research Social Interaction Development and Epistemology*, *16*, 39–61. https://doi.org/10.1016/S0166-4115(08)62193-7

Bar-On, R. (1997). *The Emotional Quotient Inventory (EQ-i): Technical manual*. Toronto, Canada: Multi-Health Systems.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Barrett, L. F., & Salovey, P. (2002). Introduction. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 1–8). New York, NY: Guilford Press.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 153–166. https://doi.org/10.1037/0278-7393.15.1.153

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge, England: Cambridge University Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application.* (pp. 435–459). Cambridge, MA: The MIT Press.

Bloom, A. D. (1988). *The closing of the American mind: How higher education has failed democracy and impoverished the souls of today's students*. London, England: Simon & Schuster.

Bond, J., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

British Psychological Society. (2009). Code of ethics and conduct.

British Psychological Society. (2014). Code of human research ethics.

British Psychological Society. (2018). Code of ethics and conduct.

Bruni, F. (2016, November 5). Why this election terrifies me. *The New York Times*. Retrieved from https://www.nytimes.com

Buehner, M. J., & Cheng, P. W. (2005). Causal learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 143–168). New York, NY: Cambridge University Press.

Burgman, M. A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., … Twardy, C. (2011). Expert status and performance. *PLOS ONE*, *6*(7), e22998. https://doi.org/10.1371/journal.pone.0022998

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chen, J. Q., & Gardner, H. (1997). Assessment based on multiple-intelligences theory. In P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (Vol. 2, pp. 77–102). New York, NY: Guilford Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York, NY: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, *90*(2), 272–292.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Putnam's.

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*. Washington, DC: American Psychological Association.

de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: research, method and theory* (pp. 19–50). New York, NY: Wiley.

Deakin, J. M., & Allard, F. (1991). Skilled memory in expert figure skaters. *Memory & Cognition*, *19*(1), 79–86. https://doi.org/10.3758/BF03198498

DePaulo, B. M., Kirkendol, S. E., Kashy, D. A., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979

DeSousa, R. B. (1987). *The rationality of emotion*. Cambridge, MA: MIT Press.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Dunbar, G. (2005). *Evaluating research methods in psychology: A case study approach*. Malden, MA: BPS Blackwell.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5), i–113. https://doi.org/10.1037/h0093599

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, Supplement*, *5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychological Science*, *5*(2), 94–98. https://doi.org/10.1111/j.1467-9280.1994.tb00637.x

Duschl, R. A. (2008). Quality argumentation and epistemic criteria. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 159–179). New York, NY: Springer.

Editorial. (2016, November 7). Editorial: The Guardian view on America's choice: Don't vote for Trump. Elect Clinton [Editorial]. *The Guardian*. Retrieved from http://www.theguardian.com

Editorial. (2017, April 23). The Guardian view on France's election: A win for Macron and hope [Editorial]. *The Guardian*. Retrieved from http://www.theguardian.com

Einstein, A., & Infeld, L. (1938). *The evolution of physics: The growth of ideas from early concepts to relativity and quanta*. Cambridge, England: Cambridge University Press.

Evans, J. S. B. T. (1989). *Bias in reasoning: Causes and consequences*. Hove, England: Erlbaum.

Evans, J. S. B. T. (2005). Deductive reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 169–184). New York: Cambridge University Press.

Eysenck, H. J. (1998). *A new look intelligence*. New Brunswick, NJ: Transaction.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London, England: Sage.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Freudenthaler, H. H., Neubauer, A. C., Gabler, P., Scherl, W. G., & Rindermann, H. (2008). Testing and validating the Trait Emotional Intelligence Questionnaire (TEIQue) in a German-speaking sample. *Personality and Individual Differences*, *45*(7), 673–678.

Fromm, E. (1941). *Escape from freedom*. New York, NY: Holt, Rinehart and Winston.

Furnham, A. (2006). Explaining the popularity of emotional intelligence. In K. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* Mahwah, NJ: Lawrence Erlbaum Associates.

Galotti, K. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, *105*(3), 331–351. https://doi.org/10.1037/0033-2909.105.3.331

Garcia-Mila, M., Gilabert, S., Erduran, S., & Felton, M. (2013). The effect of argumentative task goal on the quality of argumentative discourse. *Science Education*, *97*(4), 497–523.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

Gardner, H. (1999a). *The disciplined mind: What all students should understand*. New York, NY: Simon & Schuster.

Gardner, H. (1999b). Who owns intelligence? *Atlantic Monthly*, *283*(2), 67–76.

Garnham, A., & Oakhill, J. (1994). *Thinking and reasoning*. Oxford, England: Blackwell.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355. https://doi.org/10.1016/0010-0285(80)90013-4

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Gilovich, T., Griffin, D., & Kahneman, D. (2002). Preface. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. xv–xvi). Cambridge, England: Cambridge University Press.

Goel, V., & Grafman, J. (2000). Role of the right prefrontal cortex in ill-structured planning. *Cognitive Neuropsychology*, *17*(5), 415–436.

Goldstone, R. L., & Son, J. Y. (2005). Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 13–36). New York, NY: Cambridge University Press.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*(4), 565–610. https://doi.org/10.1016/S0364-0213(01)00046-5

Goleman, D. (1995). *Emotional intelligence*. New York, NY: Bantam Books.

Goleman, D. (2005). *Emotional intelligence: 10th anniversary edition*. New York, NY: Bantam Dell.

Goodman, P. S. (2016, May 20). ‘Brexit,’ a feel-good vote that could sink Britain’s economy. *The New York Times*. Retrieved from https://www.nytimes.com

Gorsuch, R. (1974). *Factor analysis*. Philadelphia, PA: W. B. Saunders.

Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(2), 108–120. https://doi.org/10.1080/10705519709540064

Greeno, J. G. (1977). Process of understanding in problem solving. In N. Castellan, D. Pisoni, & G. Potts (Eds.), *Cognitive theory* (Vol. 2, pp. 43–83). Hillsdale, NJ: Erlbaum.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*(3), 411–435. https://doi.org/10.1016/0010-0285(92)90013-R

Hahn, U., Bluhm, R., & Zenker, F. (2017). Causal argument. In M. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 475–493). New York, NY: Oxford University Press.

Halevy, R., Shalvi, S., & Verschuere, B. (2014). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research*, *40*(1), 54–72. https://doi.org/10.1111/hcre.12019

Halpern, D. F. (2013). *Thought and knowledge: An introduction to critical thinking* (5th edition.). Hove, England: Psychology Press.

Hambleton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, *5*(1), 1–16.

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. Castellan, D. Pisoni, & G. Potts (Eds.), *Cognitive theory* (Vol. 2, pp. 21–44). Hillsdale, NJ: Erlbaum.

Heath, C., & Tversky, A. (1990). Preference and belief: Ambiguity and competence in choice under uncertainty. In K. Borcherding, O. I. Larichev, & D. M. Messick (Eds.), *Contemporary issues in decision making* (pp. 93–123). Oxford, England: North-Holland.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 719–731. https://doi.org/10.1037/0278-7393.11.1-4.719

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59–87). New York, NY: Academic Press.

Holyoak, K. J., & Morrison, R. G. (2005). Thinking and reasoning: A reader's guide. In Keith J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 1–9). New York, NY: Cambridge University Press.

Howard, M. E. (1983). *The causes of war and other essays*. Cambridge, MA: Harvard University Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98.

Johnson, B. (2016a, March 16). There is only one way to get the change we want--vote to leave the EU. *The Telegraph*. Retrieved from https://www.telegraph.co.uk

Johnson, B. (2016b, October 16). Cripes! I jolly nearly backed Dave on Europe. *The Sunday Times*. Retrieved from https://www.thetimes.co.uk

Johnson, D. D. P. (2004). *Overconfidence and war: The havoc and glory of positive illusions*. Cambridge, MA: Harvard University Press.

Johnson, D. D. P., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, *477*(7364), 317–320. https://doi.org/10.1038/nature10384

Johnson-Laird, P. N. (1982). Ninth Bartlett Memorial Lecture: Thinking as a skill. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *34A*(1), 1–29.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Keele, S. M., & Bell, R. C. (2008). The factorial validity of emotional intelligence: An unresolved issue. *Personality and Individual Differences*, *44*(2), 487–500.

Kelly, G. J., Drucker, S., & Chen, K. (1998). Students' reasoning about electricity: Combining performance assessment with argumentation analysis. *International Journal of Science Education*, *20*, 849–871.

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216–247. https://doi.org/10.1006/obhd.1999.2847

Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). Florence, KY: Taylor & Frances/Routledge.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1534–1555.

Koellinger, P., Minniti, M., & Schade, C. (2007). "I think I can, I think I can": Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology*, *28*(4), 502–527. https://doi.org/10.1016/j.joep.2006.11.002

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248–294. https://doi.org/10.1016/0010-0285(85)90009-X

Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.

Kuhn, D., Phelps, E., & Walters, J. (1985). Correlational reasoning in an everyday context. *Journal of Applied Developmental Psychology*, *6*, 85–97.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press.

Leighton, J. P., Rogers, W. T., & Maguire, T. O. (1999). Assessment of student problem-solving on ill-defined tasks. *The Alberta Journal of Educational Research*, *45*(4), 409–427.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs*, *66*(2), 125–144. https://doi.org/10.1080/03637759909376468

Levine, T. R., Serota, K. B., Carey, F., & Messer, D. (2013). Teenagers lie a lot: A further investigation into the prevalence of lying. *Communication Research Reports*, *30*(3), 211–220. https://doi.org/10.1080/08824096.2013.806254

Levitt, S. D. (2016). *Heads or tails: The impact of a coin toss on major life decisions and subsequent happiness* (Working Paper No. 22487). National Bureau of Economic Research. https://doi.org/10.3386/w22487

Lewis, M. (2016). *The undoing project: A friendship that changed the world*. London, England: Allen Lane.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior & Human Performance*, *20*(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher*, *23*, 4–14.

Lipton, P. (1991). *Inference to the best explanation*. London, England: Routledge.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 176–201. https://doi.org/10.1037/0278-7393.27.1.176

Mackintosh, N. J. (2011). History of theories and measurement of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 3–19). New York, NY: Cambridge University Press.

Maier, N. (1930). Reasoning in humans. I. On direction. *Journal of Comparative Psychology*, *10*, 15–43.

Marcus, G. (2014, July 11). The trouble with brain science. *The New York Times*. Retrieved from https://www.nytimes.com

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507–536.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–34). New York, NY: Basic Books.

Mayer, J. D., Salovey, P., Caruso, D. R., & Cherkasskiy, L. (2011). Emotional intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 528–549). New York, NY: Cambridge University Press.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74. https://doi.org/10.1017/S0140525X10000968

Mikolajczak, M., Luminet, O., Leroy, C., & Roy, E. (2007). Psychometric properties of the Trait Emotional Intelligence Questionnaire: Factor structure, reliability, construct, and incremental validity in a French-Speaking population. *Journal of Personality Assessment*, *88*(3), 338–353.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences*. Princeton, NJ: Princeton University Press.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30. https://doi.org/10.1109/JRPROC.1961.287775

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Moynihan, D. P. (1970, January 12). To solve a problem, one must first define it. *The New York Times*. Retrieved from https://www.nytimes.com

Muncer, S. J., & Ling, J. (2006). Psychometric analysis of the Empathy Quotient (EQ) scale. *Personality and Individual Differences*, *40*(6), 1111–1119. https://doi.org/10.1016/j.paid.2005.09.020

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. London, England: Prentice-Hall.

Nossiter, A. (2017, May 3). Le Pen and Macron clash in vicious presidential debate in France. *The New York Times*. Retrieved from https://www.nytimes.com

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510–520. https://doi.org/10.1037/0278-7393.14.3.510

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 321–349). New York, NY: Cambridge University Press.

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *The Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology*, *56A*(2), 351–382. https://doi.org/10.1080/02724980244000288

Odean, T. (1999). Do investors trade too much? *American Economic Review*, *89*, 1279–1298.

Ohlsson, S. (2012). The problems with problem solving: Reflections on the rise, current status, and possible future of a cognitive research paradigm. *The Journal of Problem Solving*, *5*(1), 101–128.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.

Ottati, V., Price, E. D., Wilson, C., & Sumaktoyo, N. (2015). When self-perceptions of expertise increase closed-minded cognition: The earned dogmatism effect. *Journal of Experimental Social Psychology*, *61*, 131–138. https://doi.org/10.1016/j.jesp.2015.08.003

Pallier, G., Wilkinson, R., Danthir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, *129*(3), 257–299. https://doi.org/10.1080/00221300209602099

Perkins, D. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, *77*(5), 562–571.

Perkins, D. (1986). *Knowledge as design*. Hillsdale, NJ: Erlbaum.

Petrides, K. V. (2009a). Psychometric properties of the Trait Emotional Intelligence Questionnaire (TEIQue). In C. Stough, D. H. Saklofske, & J. D. A. Parker (Eds.), *Assessing emotional intelligence: Theory, research, and applications* (pp. 85–101). New York, NY: Springer.

Petrides, K. V. (2009b). *Technical manual for the Trait Emotional Intelligence Questionnaires (TEIQue)*. London, England: London Psychometric Laboratory.

Petrides, K. V., & Furnham, A. (2000). On the dimensional structure of emotional intelligence. *Personality and Individual Differences*, *29*, 313–320.

Petrides, K. V., & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, *17*(1), 39–57. https://doi.org/10.1002/per.466

Petrides, K. V., Pita, R., & Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. *British Journal of Psychology*, *98*, 273–289.

Quine, W. V. (1977). Natural kinds. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 155–175). Ithaca, NY: Cornell University Press.

Raven, J. C., Court, J. H., & Raven, J. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales--Advanced Progressive Matrices, Sets I & II.* Oxford, England: Oxford Psychologists Press Ltd.

Reitman, W. R. (1965). *Cognition and thought: An information-processing approach*. New York, NY: Wiley.

Renzi, M. (2016, June 22). Don't vote remain for us Europeans – Do it for yourselves. *The Guardian*. Retrieved from http://www.theguardian.com

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion*, *1*(3), 196–231. https://doi.org/10.1037/1528-3542.1.3.196

Rogers, W. T., Maguire, T. O., & Leighton, J. P. (1998, July). *Alternative methods for assessing problem solving in junior high school mathematics*. Paper presented at the annual meeting of the Canadian Psychological Association, Edmonton.

Sagan, C. (1997). The fine art of baloney detection. In *The demon-haunted world: Science as a candle in the dark* (pp. 189–206). London, England: Headline.

Salovey, P. (2011). Emotional intelligence. In R. Arkin (Ed.), *Most underappreciated: 50 prominent social psychologists describe their most unloved work* (pp. 254–258). New York, NY: Oxford University Press.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, *9*(3), 185–211.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, *36*(10), 1128–1137.

Schustack, M., & Sternberg, R. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, *110*, 101–120.

Serota, K. B., & Levine, T. R. (2015). A few prolific liars: Variation in the prevalence of lying. *Journal of Language and Social Psychology*, *34*(2), 138–157. https://doi.org/10.1177/0261927X14528804

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, *36*(1), 2–25. https://doi.org/10.1111/j.1468-2958.2009.01366.x

Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 95–116). New York, NY: Cambridge University Press.

Smith, M. E., Hancock, J. T., Reynolds, L., & Birnholtz, J. (2014). Everyday deception or a few prolific liars? The prevalence of lies in text messaging. *Computers in Human Behavior*, *41*, 220–227. https://doi.org/10.1016/j.chb.2014.05.032

Spearman, C. E. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Sternberg, R. J. (1996). *Successful intelligence: How practical and creative intelligence determine success in life*. New York, NY: Simon & Schuster.

Sternberg, R. J., & Kaufman, S. B. (2011). Preface. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. xv–xix). New York, NY: Cambridge University Press.

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist*, *50*(11), 912–927.

Terracciano, A., Costa, P. T. J., & McCrae, R. R. (2006). Personality Plasticity After Age 30. *Personality and Social Psychology Bulletin*, *32*(8), 999–1009.

Thurstone, L. L. (1934). The vectors of mind. *The Psychological Review*, *41*(1), 1–32.

Time. (1995, October 2). *Time*, *146*(14).

Time. (2017, April 3). *Time*, *189*(12).

Tolstoy, L. (1943). *War and peace*. (L. Maude & A. Maude, Trans.). London, England: Oxford University Press.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge, England: Cambridge University Press.

True; Truth. (2016). *The American heritage dictionary of the English language* (5th ed.). Retrieved from https://ahdictionary.com

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 84–98). Cambridge, England: Cambridge University Press.

Tversky, B. (2005). Visuospatial reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 209–240). New York, NY: Cambridge University Press.

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence*, *34*(5), 487–502.

von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, *45*(1), 101–131.

Warwick, J., & Nettelbeck, T. (2004). Emotional intelligence is...? *Personality and Individual Differences*, *37*, 1091–1100.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: structure and content*. London, England: Batsford.

Waterhouse, L. (2006). Inadequate evidence for multiple intelligences, Mozart effect, and emotional intelligence theories. *Educational Pscyhologist*, *41*, 247–255.

Watson, J. (1973). The growing up of cancer research. *Science Year: The Book World Science Annual*.

Wechsler, D. (1940). Nonintellective factors in general intelligence. *Psychological Bulletin*, *37*, 444–445.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). London, England: Pearson.

Weisberg, Robert W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General*, *110*(2), 169–192. https://doi.org/10.1037/0096-3445.110.2.169

Weisberg, R.W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General*, *110*(2), 169–192. https://doi.org/10.1037/0096-3445.110.2.169

Wertheimer, M. (1961). *Productive thinking*. London, England: Tavistock Publications.

Wigdor, A., & Garner, W. (Eds.). (1982). *Ability testing: Uses, consequences, and controversies*. Washington, DC: National Academy Press.

Zernike, K. (2016, February 29). Testing for joy and grit? Schools nationwide push to measure students' emotional skills. *The New York Times*. Retrieved from http://www.nytimes.com

Appendices

A. Sample Information Sheet

DEPARTMENT OF PSYCHOLOGICAL SCIENCES
BIRKBECK, UNIVERSITY OF LONDON

Interpreting evidence in support of causal claims

As some of you may know, I'm a second-year PhD student in psychology at Birkbeck, University of London, conducting the research for my dissertation, and I'd like you to help.

Before you decide to take part in this study, it is important for you to understand why the research is being done and what it will involve. Please read the following carefully and make your decision whether to participate a thoughtful one.

In my dissertation, I'm investigating the frequency with which we misrepresent everyday problems and why with the aim of becoming better at representing and thereby solving them. Everyday problems are virtually every problem in life. Whereas a well-defined problem is clear in every respect, at least one critical aspect of an everyday problem is unclear. For example, we may be uncertain as to our actual problem, or we may be unsure of our goal.

You have been approached as a potential participant given your assumed willingness to participate in such studies, perhaps as indicated by being in the Birkbeck participant database.

Should you participate, you will do so online starting in late July/early August 2017. In participating, you will be presented with two hypothetical everyday problems as well as suggested causes of those problems. Then, for each problem you will be asked to choose from a list the item or items that serve(s) as standalone evidence in support of the causal claim. Your responses will be confidential and anonymized. There are no risks envisioned in participating, but you are free to withdraw at anytime prior to the aggregation of any data collected. The data will be referred to in my dissertation and may be referred to collectively in conference proceedings and other published works.

This study is one in a series that will inform my dissertation. The project has been approved by Birkbeck's Department of Psychological Sciences Research Ethics Committee.

Many thanks for your consideration, and I sincerely hope you'll participate. By participating, you will help me in my search for answers to some important and currently unanswered questions. Should you want any further information, please feel free to contact one of us:

Primary Investigator
Ronald Balzan
rbalza01@mail.bbk.ac.uk

Supervisor
Professor Richard Cooper
r.cooper@bbk.ac.uk

Best regards,

Ron Balzan

B. Sample Consent Form

DEPARTMENT OF PSYCHOLOGICAL SCIENCES
BIRKBECK UNIVERSITY OF LONDON

Interpreting evidence in support of causal claims

I have had the details of the study explained to me and willingly consent to take part. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I understand that I will remain anonymous and that all the information given will be used for this study only.

I understand that I may withdraw my consent for the study at any time without giving any reason and to decline to answer particular questions. Furthermore I understand that I will be able to withdraw my data up to the point of data aggregation.

I understand that all information given will be kept confidential, and none of my personal details will be collected. Only my study responses will be retained, and they will reside with the service provider according to their usage terms as well as on a secure departmental server. Only the Primary Investigator and his Supervisor will have access to the data.

I understand how the results of the study will be used. The results of this study will become part of the Primary Investigator's dissertation. They may also be referred to in journal articles or other publications as well as in conference presentations. Results will be presented collectively rather than on an individual participant basis.

I confirm that I am at least 17 years of age.

Please tick the appropriate boxes:

□ I do not consent to participate in this study. *(Clicking this box will terminate your participation.)*
□ I consent to participate in this study.
Click here if you would like a copy of this form sent to your e-mail address.

Sex: □ Female □ Male □ Other
Age: ___ *(input number of years)*
Years of undergraduate university education: ___ *(input number of years)*
Country of residence: ___ *(input)*
Primary language: ___ *(input)*
How I heard about the study: ___ *(select from dropdown menu)*

Primary Investigator
Ronald Balzan
rbalza01@mail.bbk.ac.uk

Supervisor
Professor Richard Cooper
r.cooper@bbk.ac.uk

C. Investigation 2 Scenarios

SCENARIO 1

In developed countries, the average percentage of students that fail to complete secondary, or high, school is 10%. However, in the country of Imaginaria, one of these developed countries, the failure rate is 25%. A group of secondary school heads in Imaginaria claim that the reason for this high failure rate is a shortage of good teachers.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the school heads' claim:

1) As administrators, secondary school heads really aren't close enough to the completion-rate problem to understand its true cause or causes.
2) In Imaginaria, teaching jobs attract few applicants. Studies show that countries with many applicants per teaching job have better teachers and higher completion rates.
3) The country of Subservia has the same number of students per good teacher as Imaginaria, and 25% of Subservia's students also fail to complete secondary school.
4) The percentage of students that fail to complete secondary school in Imaginaria is so high because there aren't enough good teachers in the country.
5) The country of Explora has the same number of students per good teacher as Imaginaria, and 28% of Explora's students fail to complete secondary school.
6) None of the above

SCENARIO 2

A friend of mine, Joan, is purchasing a new car, and her car salesperson has suggested she purchase a treatment that the salesperson claims will prevent rust. My friend is on a budget, yet she can still afford the treatment, and she will buy it if it works. She's just not convinced that the treatment is likely to work.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the car salesperson's claim that the treatment will help to prevent the car from rusting:

1) One of Joan's colleagues, Tibur, bought the treatment when he purchased his new car, and his car started to rust after about two years.
2) As Joan lives by the Atlantic Ocean, the car will be continuously exposed to salty moist air, which is commonly known to cause rust.
3) Sabi, Joan's best friend, purchased a new car—a convertible—last year, and he bought the treatment. So far, his car has not rusted.
4) An independent survey found that of 500 people who purchased the treatment, 6 of their cars rusted, and of 500 who did not, 300 rusted.
5) The salesperson that sold Joan her new car does not receive a commission on the sale of the rust prevention treatment.
6) None of the above

D. Investigation 3 Scenarios

SCENARIO 1

Recently, a friend of mine purchased a new car. At the time of purchase, the dealer recommended a rust-prevention treatment, and she recommended one treatment in particular. My friend said she'd think about it.

Over the next few weeks, my friend asked around to see what other people's experience had been with the particular treatment her dealer recommended. Here's what each of those people had to say:

> **Person 1:** "Oh, my dealer recommended it to me too. I've purchased several cars from her over the years, so I trust her. I bought it, and I've never had any rust."
>
> **Person 2:** "Yeah, I've purchased it, and I've never had a rust problem."
>
> **Person 3:** "I purchased it when I got my last car, and it certainly didn't stop my car from rusting!"
>
> **Person 4:** "I got it when I purchased my last car, but it didn't prevent my car from rusting."
>
> **Person 5:** "I get it every time I get a new car, and I've never had a rust problem."
>
> **Person 6:** "I would never purchase another car from my last dealer. He recommended it to me, so I bought it. It wasn't cheap! And I've had a terrible rust problem."
>
> **Person 7:** "Yes, I've purchased it. It seems to work."
>
> **Person 8:** "I purchased it, and I've gotten no rust."
>
> **Person 9:** "It's a no-brainer. I always get it, and I have never had rust. Never."
>
> **Person 10:** "I've started getting it whenever I get a new car—not that I've had many. And I've never had rust."

Please answer the following questions:

Based on the above responses, does the particular rust-prevention treatment recommended by my friend's car dealer prevent rust? Please select one of the following responses: Yes, No or Can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if it prevents rust?

SCENARIO 2

I have a colleague at work who smokes, and he's desperate to stop. He saw an ad recently for a particular brand of nicotine patch that's meant to help you stop smoking, and he's considering trying it. He's actually seen ads for it a lot.

He's tried virtually everything else, but the patches are quite expensive. So, before deciding to make the investment, he asked a number of friends and some of our other colleagues who are or were smokers to see if they've had any experience with it. Here's what each of those people had to say:

> **Person 1:** "Well, they are expensive, but in a sense it's cheap given it's your health at stake. I used them, and I stopped smoking pretty quickly thereafter."
>
> **Person 2:** "I used them. And let me put it this way, 'Do you have a cigarette?' I still smoke."
>
> **Person 3:** "What a waste of money! I used them for months! I just can't seem to stop smoking."
>
> **Person 4:** "I did use it, and I haven't had a cigarette in a long time. No desire whatsoever."
>
> **Person 5:** "You've got to try them! They may sound expensive, but believe me, over time it'll end up being a lot cheaper than cigarettes. I used them, and I haven't wanted a cigarette since."

Please answer the following questions:

Based on the above responses, do the advertised nicotine patches seem to help people stop smoking? Please select one of the following responses: Yes, No or Can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if it helps?

E. Investigation 4 Scenarios

SCENARIO 1

Last year, a friend of mine went to the doctor for a routine physical examination. While my friend was in otherwise seemingly good health, the doctor recommended my friend lose 20 pounds. My friend had recently heard people at his office talk about one diet in particular. Recently it had been featured on several television programs, and he often saw it mentioned in newspapers and magazines. Indeed several of his colleagues were on the diet as were several other people he knew. He decided to ask some of these people about their experience with it.

Six people with whom he talked said they were indeed on the diet, and they all said they were losing weight. My friend also spoke to two people who said that they were on the diet, but they were not losing weight. Three people said they were not on that diet, but they were losing weight. And one person said they were not on the diet, and they were not losing weight.

How many people did my friend speak to?

Does the diet have anything to do with whether or not these people are losing weight?

Please select one of the following responses: Yes, No or Can't tell.

Why did you choose this response?

SCENARIO 2

Last year, a friend of mine started losing his hair at an alarming rate, especially for someone so young. My friend had recently seen an ad for an over-the-counter hair-loss treatment. The ad said that the treatment not only stops additional hair loss, but leads to hair regrowth. He had also seen guys using it at the gym. My friend wanted more hair, but he had seen ads like this one before. Before deciding to use it himself, he decided to ask around to see what other people's experience with the treatment had been.

Twelve people with whom he talked said they were indeed using the treatment, and they all said their hair was re-growing. My friend also spoke to four people who said that they were using the treatment, but their hair was not re-growing. Six people said they were not using the treatment, but their hair was re-growing. And two people said they were not using the treatment, and their hair was not re-growing.

How many people did my friend speak to?

Does the hair-loss treatment have anything to do with whether or not these people are losing or re-growing hair?

Please select one of the following responses: Yes, No or Can't tell.

Why did you choose this response?

F. Investigation 5 Scenario

SCENARIO

Recently, a friend of mine lost his job. It's been awhile since he's been in the job market, and he says it seems that everywhere he turns, there's someone offering some service that's "guaranteed" to make it easier for him to find a new job.

He was approached by one service in particular that says it will help him rewrite his resume, or CV; structure his job search; help him make connections; and coach him through the search, interview and acceptance processes. He needs a job, and he thinks this service could be a real advantage in his search.

That being said, it's an expensive service. So, my friend has decided to ask some people who have used the service about their job search experiences. Here's what each of those people had to say:

> **Person 1:** "They were brilliant! I was nervous about it as they're not cheap, but it was a great investment, without question. They definitely helped me get a job."
>
> **Person 2:** "To be honest, I think it depends on the person at the company, or "coach," with whom they have you work. I actually regret using them because I don't think they helped me all that much."
>
> **Person 3:** "You know—and they tell you this too—you get out of it what you put into it. They did help me, but make no mistake: most of the work still falls on you. Don't expect them to find a job for you. That being said, I think it was a good investment. It helped."
>
> **Person 4:** "I did use them, but I should have trusted my intuition and not done so. When it comes to finding a job, it really is your responsibility. The pointers they gave me were pretty much common sense in nature, and what I couldn't have figured out for myself, I could have figured out with a Google search."
>
> **Person 5:** "D-O-D-G-Y is how I'd describe them. Do not waste your money!"

Please answer the following questions:

Based on the above responses, does the service my friend is considering have any relationship to one's success in finding a job? Please select one of the following responses: yes, no or can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if this service has any relationship to success in finding a job?

PROBLEM REPRESENTATION

INFORMATION SHEET 1

(Text presented to alternate participants)

Problem representations are critical to problem solving. They are where the process of problem solving begins. A problem *representation* is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we want to be once the change has taken place; and the *obstacles* and *constraints* that lie between and may be determined by the problem and goal.

Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. When we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues. Let's say, for example, that I'm overweight. I want to lose weight, so I exercise more. If my weight problem, however, has more to do with my diet, in ignoring my diet, I have misrepresented the problem, and I will most likely fail to lose weight.

Problems may be misrepresented for a number of reasons. Sometimes we mistakenly over-emphasize things that may matter little or not at all, like someone's opinion regardless of fact or when their experience could be an exception rather than the norm. Or we may fail to emphasize things that do matter enough, such as diet in the example above. Sometimes we'll look at a simple count and assume it tells the whole story when it may tell only part. And sometimes we're just not ready to be honest with ourselves about the nature of our problems, or we don't know what our goals should be.

Please answer the following questions:

1) Which one of the following is not a component of a problem representation?

E) An obstacle
F) A goal
G) The problem
H) A friend's opinion

2) What factor or factors did I fail to consider in representing my weight problem described above?

E) Exercise
F) Diet
G) Exercise with diet
H) A goal

PROBLEM REPRESENTATION

INFORMATION SHEET 2

(Text presented to participants that did not receive Information Sheet 1)

Problem representations are critical to problem solving. They are where the process of problem solving begins. A problem *representation* is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we want to be once the change has taken place; and the *obstacles* and *constraints* that lie between and may be determined by the problem and goal.

Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. When we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues. Let's say, for example, that I'm overweight. I want to lose weight, so I exercise more. If my weight problem, however, has more to do with my diet, in ignoring my diet, I have misrepresented the problem, and I will most likely fail to lose weight.

Please answer the following questions:

1) Which one of the following is not a component of a problem representation?

A) An obstacle
B) A goal
C) The problem
D) A friend's opinion

2) What factor or factors did I fail to consider in representing my weight problem described above?

A) Exercise
B) Diet
C) Exercise with diet
D) A goal

SCENARIO RE-PRESENTED

Recently, a friend of mine lost his job. It's been awhile since he's been in the job market, and he says it seems that everywhere he turns, there's someone offering some service that's "guaranteed" to make it easier for him to find a new job.

He was approached by one service in particular that says it will help him rewrite his resume, or CV; structure his job search; help him make connections; and coach him through the search, interview and acceptance processes. He needs a job, and he thinks this service could be a real advantage in his search.

That being said, it's an expensive service. So, my friend has decided to ask some people who have used the service about their job search experiences. Here's what each of those people had to say:

> **Person 1:** "They were brilliant! I was nervous about it as they're not cheap, but it was a great investment, without question. They definitely helped me get a job."
>
> **Person 2:** "To be honest, I think it depends on the person at the company, or "coach," with whom they have you work. I actually regret using them because I don't think they helped me all that much."
>
> **Person 3:** "You know—and they tell you this too—you get out of it what you put into it. They did help me, but make no mistake: most of the work still falls on you. Don't expect them to find a job for you. That being said, I think it was a good investment. It helped."
>
> **Person 4:** "I did use them, but I should have trusted my intuition and not done so. When it comes to finding a job, it really is your responsibility. The pointers they gave me were pretty much common sense in nature, and what I couldn't have figured out for myself, I could have figured out with a Google search."
>
> **Person 5:** "D-O-D-G-Y is how I'd describe them. Do not waste your money!"

Please answer the following questions:

Do you wish to change your original response(s) to this scenario? If yes, please continue to the next question; if no, you will be taken to the final page of the study.

Based on the above responses, does the service my friend is considering have any relationship to one's success in finding a job? Please select one of the following responses: yes, no or can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if this service has any relationship to success in finding a job?

G. Investigation 6 Scenario

SCENARIO

Recently, the daughter of a friend of mine was accepted into two universities. She thinks she knows what she wants to study, and the programs for her subject seem equally good between the universities. Attendance at both universities will cost roughly the same. Yet one university is considered more prestigious than the other.

So, my friend's daughter isn't sure which university to attend. While she liked the people she met at the less prestigious university more, not to mention the milder winters of its climate, she can't help but think attending the more prestigious university will be more beneficial when it comes to finding a job. And having a job at graduation will be very important given her level of student loans.

Over the next few weeks, my friend's daughter asked some recent alumni of the less prestigious university what their job search experience had been. Here's what each of those people had to say:

> **Person 1:** "I had a job at graduation. I went straight from university to work, and it's a good job too."
>
> **Person 2:** "I didn't have a job by graduation. That being said, I had a lot of fun at university, so maybe that's the trade-off."
>
> **Person 3:** "Well, I'd say don't focus on placement rates, or the number of students who have jobs lined up by graduation. You can never be sure that two universities are calculating that number in the same way. Plus they know that people look at those numbers in choosing where to go, so they have an incentive to inflate them. I just chose the university I liked more, and I had a job locked up six months before graduating."
>
> **Person 4:** "Yeah, I'd think hard about that one. I've been out of university for over a year now, and I still haven't found a job."
>
> **Person 5:** "That's a sore subject. I didn't have a job when I graduated, and it took me a while to find one. Now, I'm really struggling to repay my student loans."

Please answer the following questions:

Based on the above responses, in this situation does university prestige have any relationship to placement in a job by graduation? Please select one of the following responses: yes, no or can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if university prestige has any relationship to placement in a job by graduation in this situation?

PROBLEM REPRESENTATION

INFORMATION SHEET 1

(Text presented to alternate participants)

Life is full of problems, and part of being successful is being good at solving those problems. Of course, there are many reasons why people fail to solve problems. Sometimes, for instance, they ignore one or more important aspects of a problem. Let's say that I'm overweight, and since I want to lose weight, I exercise more. But what if my weight problem has more to do with my diet than how much I exercise? If that's the case, I will most likely fail to lose weight, for I will have failed to take the role of my diet in weight loss into account.

In addition to failing to emphasize things that do matter, sometimes we mistakenly emphasize factors that don't, at least not at first. For example, let's say I'm considering the purchase of a rust prevention treatment for my car. While climate, let's say, can impact the occurrence of rust, I'd be wrong to consider it *before* looking at whether cars that have been treated rusted any less than cars that haven't. It's only by looking at the occurrence of rust in cars with *and without* the treatment that we can tell whether the treatment is effective and worth purchasing. In short, in deciding whether to purchase the treatment, climate doesn't matter.

Please answer the following questions:

1) What factor or factors did I fail to consider in representing my weight problem described above?

E) Exercise
F) Diet
G) My being overweight
H) A goal

2) In the rust prevention treatment example above, which factor or factors below matter in the treatment purchase decision?

E) The occurrence of rust on cars that have had the treatment
F) The climate
G) The occurrence of rust on cars that have not had the treatment
H) The car dealer's reputation

PROBLEM REPRESENTATION

INFORMATION SHEET 2

(Text presented to participants that did not receive Information Sheet 1)

Life is full of problems, and part of being successful is being good at solving those problems. Of course, there are many reasons why people fail to solve problems. Sometimes, for instance, they ignore one or more important aspects of a problem. Let's say that I'm overweight, and since I want to lose weight, I exercise more. But what if my weight problem has more to do with my diet than how much I exercise? If that's the case, I will most likely fail to lose weight, for I will have failed to take the role of my diet in weight loss into account.

In general terms, problem solving begins with *problem representations*, and they are critical to the process. A representation is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we'd rather be; and the *obstacles* and *constraints* that lie between them. Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. And when we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues.

Please answer the following questions:

1) Which one of the following is not a component of a problem representation?

E) An obstacle
F) A goal
G) The problem
H) A friend's opinion

2) What factor did I fail to consider in representing my weight problem described above?

A) Exercise
B) Diet
C) My being overweight
D) A goal

SCENARIO RE-PRESENTED

Recently, the daughter of a friend of mine was accepted into two universities. She thinks she knows what she wants to study, and the programs for her subject seem equally good between the universities. Attendance at both universities will cost roughly the same. Yet one university is considered more prestigious than the other.

So, my friend's daughter isn't sure which university to attend. While she liked the people she met at the less prestigious university more, not to mention the milder winters of its climate, she can't help but think attending the more prestigious university will be more beneficial when it comes to finding a job. And having a job at graduation will be very important given her level of student loans.

Over the next few weeks, my friend's daughter asked some recent alumni of the less prestigious university what their job search experience had been. Here's what each of those people had to say:

> **Person 1:** "I had a job at graduation. I went straight from university to work, and it's a good job too."
>
> **Person 2:** "I didn't have a job by graduation. That being said, I had a lot of fun at university, so maybe that's the trade-off."
>
> **Person 3:** "Well, I'd say don't focus on placement rates, or the number of students who have jobs lined up by graduation. You can never be sure that two universities are calculating that number in the same way. Plus they know that people look at those numbers in choosing where to go, so they have an incentive to inflate them. I just chose the university I liked more, and I had a job locked up six months before graduating."
>
> **Person 4:** "Yeah, I'd think hard about that one. I've been out of university for over a year now, and I still haven't found a job."
>
> **Person 5:** "That's a sore subject. I didn't have a job when I graduated, and it took me a while to find one. Now, I'm really struggling to repay my student loans."

Please answer the following questions:

Do you wish to change your original response(s) to this scenario? If yes, please continue to the next question; if no, you will be taken to the final page of the study.

Based on the above responses, in this situation does university prestige have any relationship to placement in a job by graduation? Please select one of the following responses: yes, no or can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if university prestige has any relationship to placement in a job by graduation in this situation?

H. Investigation 7 Scenario

SCENARIO

A friend of mine collects art, and she prefers to buy from one dealer in particular. She likes this dealer for two reasons: she seems to have access to the "best" works by the artists my friend likes, and she seems to price the works she sells more reasonably than other dealers.

However, another dealer now has a painting that my friend covets. And while my friend loves art, she also primarily views art as an investment, so a work's "price-to-value" is of utmost importance to her.

As the art world is relatively small, my friend has decided to ask some people she knows who have dealt with this other dealer whether the dealer offers good value for money. Here's what each of those people had to say:

> **Person 1:** "She has made me rich! I deal with her almost exclusively, and my purchases from her have been the best investments I've made."
>
> **Person 2:** "I don't know. Maybe it's that the artists I like too few other people do, or maybe I've paid too much for them. They seemed like good values at the time I purchased them from her, but none of them have turned out to be."
>
> **Person 3:** "She's not stupid. It's as if she knows the exact value of what she sells. The thing is, though, that she has a good eye, so whatever I've purchased from her, regardless of how much I spent to get it, has turned out to be a good value over time."
>
> **Person 4:** "Oh, she definitely offers good value for money—to those longstanding clients she likes. But I'm not one of them."
>
> **Person 5:** "She's fair, and that reputation precedes her. Rest assured that if you see something you like at her gallery, it will be a good value at whatever price she asks for it."

Please answer the following questions:

Based on the above responses, is there a relationship between the other dealer and value for money? Please select one of the following responses: yes, no or can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if this other dealer has any relationship to value for money in her offerings in this situation?

PROBLEM REPRESENTATION

INFORMATION SHEET 1

(Text presented to alternate participants)

Life is full of problems, and part of being successful is being good at solving those problems. Of course, there are many reasons why people fail to solve problems. Sometimes, for instance, they ignore one or more important aspects of a problem. Let's say that I'm overweight, and since I want to lose weight, I exercise more. But what if my weight problem has more to do with my diet than how much I exercise? If that's the case, I will most likely fail to lose weight, for I will have failed to take the role of my diet in weight loss into account.

In addition to being a case of neglecting to emphasize something that matters, this example is also a case of overemphasis. Let's say, for instance, that I'm considering a specific diet I've seen advertised. So I ask people who have been on the diet what their experience has been. And let's say most of them lost weight. So given that majority, I conclude that the diet is worth trying. Conversely, given that the diet wasn't successful for *everyone*, I could also conclude it *doesn't* work. Yet I'd be incorrect in drawing either of those conclusions. In the end, it's only by looking at the incidence of weight loss with *and without* the diet that we can tell whether the diet is effective and worth purchasing.

Please answer the following questions:

1) What factor did I fail to consider in representing my weight problem described above?

E) Exercise
F) Diet
G) My being overweight
H) A goal

2) In the diet example above, which factor or factors below matter in the decision of whether or not to try the diet?

I) The occurrence of weight loss in people that have used the diet
J) The number of ads I've seen for the diet, each of which showcases a handful of "success stories"
K) The occurrence of weight loss in people that have not used the diet
L) The fact that my best friend used it and failed to lose weight

PROBLEM REPRESENTATION

INFORMATION SHEET 2

(Text presented to participants that did not receive Information Sheet 1)

Life is full of problems, and part of being successful is being good at solving those problems. Of course, there are many reasons why people fail to solve problems. Sometimes, for instance, they ignore one or more important aspects of a problem. Let's say that I'm overweight, and since I want to lose weight, I exercise more. But what if my weight problem has more to do with my diet than how much I exercise? If that's the case, I will most likely fail to lose weight, for I will have failed to take the role of my diet in weight loss into account.

In general terms, problem solving begins with *problem representations*, and they are critical to the process. A representation is a mental model consisting of four components: the *problem*, or what it is that we want to change; the *goal*, or where it is we'd rather be; and the *obstacles* and *constraints* that lie between them. Problems are *misrepresented* whenever we fail to accurately clarify one or more of those components. And when we misrepresent a problem, we end up trying to solve a problem different from the one intended, and the original problem continues.

Please answer the following questions:

1) Which one of the following is not a component of a problem representation?

A) An obstacle
B) A goal
C) The problem
D) A friend's opinion

2) What factor did I fail to consider in representing my weight problem described above?

A) Exercise
B) Diet
C) My being overweight
D) A goal

SCENARIO RE-PRESENTED

A friend of mine collects art, and she prefers to buy from one dealer in particular. She likes this dealer for two reasons: she seems to have access to the "best" works by the artists my friend likes, and she seems to price the works she sells more reasonably than other dealers.

However, another dealer now has a painting that my friend covets. And while my friend loves art, she also primarily views art as an investment, so a work's "price-to-value" is of utmost importance to her.

As the art world is relatively small, my friend has decided to ask some people she knows who have dealt with this other dealer whether the dealer offers good value for money. Here's what each of those people had to say:

> **Person 1:** "She has made me rich! I deal with her almost exclusively, and my purchases from her have been the best investments I've made."
>
> **Person 2:** "I don't know. Maybe it's that the artists I like too few other people do, or maybe I've paid too much for them. They seemed like good values at the time I purchased them from her, but none of them have turned out to be."
>
> **Person 3:** "She's not stupid. It's as if she knows the exact value of what she sells. The thing is, though, that she has a good eye, so whatever I've purchased from her, regardless of how much I spent to get it, has turned out to be a good value over time."
>
> **Person 4:** "Oh, she definitely offers good value for money—to those longstanding clients she likes. But I'm not one of them."
>
> **Person 5:** "She's fair, and that reputation precedes her. Rest assured that if you see something you like at her gallery, it will be a good value at whatever price she asks for it."

Please answer the following questions:

Do you wish to change your original response(s) to this scenario?

Based on the above responses, is there a relationship between the other dealer and value for money? Please select one of the following responses: yes, no or can't tell.

If yes, why?

If no, why?

If you can't tell, what else would you need to know before you'd be able to tell if this other dealer has any relationship to value for money in her offerings in this situation?

I. Investigation 8 Scenarios

SCENARIO 1

In some of the most developed countries, 15-year-olds perform no better than average on international standardized assessment tests despite higher-than-average annual education spending per student. A group of researchers claim that developed-country wealth causes more money to be invested less thoughtfully.

On a scale of 1 to 5, please rate each evidence item below with regard to its strength in support of the researchers' claim, 1 being strong and 5 being weak/not evidence:

Item	Strength				
	1 (Strong)	2	3	4	5 (Weak/Non)
The researchers are so far removed from the international standardized assessment test problem that they are not really in a position to understand its true cause or causes.					
In most developed countries, 15-year-olds' performance on international standardized assessment tests does not improve with annual education-spending-per-student increases.					
In one of the world's most developed countries, 15-year-olds perform better than average on a variety of international standardized assessment tests.					
15-year-olds in some of the most developed countries perform no better than average on international standardized assessment tests because the tests measure the wrong things.					
In one of the world's least developed countries, 15-year-olds perform no better than average on a variety of international standardized assessment tests.					

SCENARIO 2

A friend of mine is purchasing a new computer, and the salesperson has recommended that she buy an extended warranty that will cover repairs for the two-year period after the manufacturer's first-year warranty expires.

On a scale of 1 to 5, please rate each evidence item below with regard to its strength in support of the salesperson's claim, 1 being strong and 5 being weak/not evidence:

Item	Strength				
	1 (Strong)	2	3	4	5 (Weak/Non)
One of my friend's colleagues bought the same computer and the warranty, and he was glad he did. The extended warranty covered the problems he had two years after purchase.					
My friend has never had a problem with her computers before, so she's inclined to believe that nothing will go wrong with her new one and won't need the extended warranty.					
My friend's sister bought the same computer and the warranty, but she regretted buying the warranty. She's had the computer for 4 years, and she's never had a problem with it.					
A survey of consumers compared a number of computers, including my friend's. Of 100 people who purchased hers, 64 had problems in the second or third year of ownership.					
My friend said that the salesperson was "overly friendly." My friend also found out from a former store employee that the salesperson receives a commission on every extended warranty she sells.					

J. Investigation 9 Scenarios

SCENARIO 1

In 2014, 13% of adults worldwide were obese. In what is arguably one of the world's most developed countries, the obesity rate amongst adults is currently at 40%. A group of scientists in that country claim that the primary reason for this much-higher-than-average obesity rate is lack of exercise.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the scientists' claim:

1) The news tells us that scientific reports often aren't reliable, so while this group of scientists claim lack of exercise, others will likely claim something else.
2) These scientists work at some of the world's top universities, so their findings are more reliable than others are likely to be.
3) In this country, 60% of adults do not exercise. A global study shows that in countries with higher-than-average adult obesity rates, more than 50% of adults do not exercise.
4) In a different developed country, more than 70% of adults do not exercise and 40% of its adults are obese.
5) In yet another developed country, the adult obesity rate is 20%, and 70% of its adults do not exercise regularly.
6) None of the above

SCENARIO 2

Ellen, a friend of mine, just purchased a new home. Her estate agent has suggested that she have it treated to prevent termite infestation. My friend is on a budget, yet she can still afford the treatment, and she will buy it if it is necessary. She's just not convinced that the treatment is indeed necessary.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) that supports the estate agent's recommendation of treatment to prevent termite infestation:

1) Another one of Ellen's friends bought a house in the same area, and he has not had a termite problem.
2) Ellen's new home is on the edge of a wood, and given all of the trees, there are likely to be termites nearby.
3) A local surveyor told Ellen that half of the homes in her neighborhood received treatment, and 1% of those got termites, and 70% of those untreated got termites.
4) Ellen's new neighbor had her house treated right before she moved in, and she has not had a termite problem.
5) The estate agent that sold Ellen her new home is not affiliated with the termite prevention service she has recommended.
6) None of the above

SCENARIO 3

You are trying to lose weight. A new diet is extremely popular, and you are tempted to try it, but you've been on many diets, and for one reason or another, they never seem to work for you. Still, you personally know several people that have tried this new diet, and the results have been exceptional. Even the NHS recommends it as an effective way to lose weight.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the NHS's claim that the treatment will be effective in losing weight:

1) One of your friends has been on the diet for several months now. He has lost the weight he intended to lose, and he has kept it off.
2) Most diets are simply marketing gimmicks designed to sell books and do nothing but put money in the authors' and publishers' pockets.
3) Another one of your friends has been on the diet for several months, too, but she has gained weight despite her efforts.
4) A medical study found that of 500 people who went on the diet, 400 lost weight, and of 500 who did not go on the diet, 20 lost weight.
5) It's common knowledge that the best way to lose weight is to diet, and we're always learning, so newer diets are probably more effective than older ones.
6) None of the above

SCENARIO 4

You have been trying to quit smoking for years. You've tried everything but hypnosis, and several people you know have used it successfully, and they swear that it will work for you too. While you're open to trying it, it's expensive, and you don't want to try it until you're certain it will be effective. A group of psychologists claim that it is likely to work.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the psychologists' claim that hypnosis will help you stop smoking:

1) A study shows that when 100 patients received hypnosis to stop smoking, 70 of them indeed quit; of 100 who chose not to undergo hypnosis, only 10 successfully quit.
2) Your friend Max smoked for 20 years, and he also struggled to quit. He decided to give hypnosis a try, and after a handful of sessions, he finally stopped—for good.
3) Everyone knows that hypnosis can be used to effectively help people stop all kinds of undesirable habits.
4) Your partner is a smoker, so the odds of you quitting are against you, no matter what you try to do to stop.
5) Nobody is forcing you to smoke. You can quit whenever you want, if you want, and hypnosis will help you do that.
6) None of the above

SCENARIO 5

You have been losing hair rapidly, and you're upset. You say that until you started losing it, you never realized how much your hair mattered to you. A new hair regrowth treatment came on the market last year, however. You'd like to try it, but it's expensive. You'll have to give up other things in order to afford it. So, you'd like some assurance that it will work. A group of independent scientists claim that the treatment works.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the user scientists' claim that the treatment will successfully regrow hair:

1) One of your friends has used it since the treatment came out, and his hair has regrown.
2) Of 200 users surveyed, 150 of them claimed hair regrowth after six months.
3) Of 200 users surveyed, 50 of them claimed they had experienced no hair regrowth.
4) Another one of your friends has also used the treatment, and his hair has regrown.
5) In a survey of 200 non-users, 75 experienced hair regrowth over the course of a year.
6) None of the above

SCENARIO 6

The average global murder rate is between 6 and 8 per 100,000 people. One developed country, a country not a war, however, has a rate of 24 per 100,000 people. A group of senior police officers claim that the reason for this high murder rate is a shortage of police officers.

From the list below, which item or items (you may select more than one) serve(s) as standalone evidence (meaning in the absence of any other information) in support of the senior police officers' claim:

1) For police officers to argue anything other that a shortage of police officers suggests they're not doing their job well.
2) Police officer jobs attract few applicants in the country. A global study shows that countries with many applicants per police officer job have lower murder rates.
3) Another country has the same number of police officers per person, and its murder rate is between 6 and 8 per 100,000 people per year.
4) The murder rate is 24 per 100,000 people in the developed country in question because there is a shortage of police officers in the country.
5) Yet another developed country has the same number of police officers per person, and its murder rate is 30 per 100,000 people.
6) None of the above

K. Trait Emotional Intelligence Questionnaire Version 1.50 (Petrides, 2009)

Reprinted with written permission from K. V. Petrides, the copyright owner.

Instructions

- Please complete this questionnaire on your own and in quiet conditions.
- Please answer each statement below by putting a circle around the number that best reflects your degree of agreement or disagreement with that statement. *There are no right or wrong answers*.
- Work quickly, and don't think too long about the exact meaning of the statements.

		DISAGREE COMPLETELY						AGREE COMPLETELY
1.	I'm usually able to control other people	1	2	3	4	5	6	7
2.	Generally, I don't take notice of other people's emotions	1	2	3	4	5	6	7
3.	When I receive wonderful news, I find it difficult to calm down quickly	1	2	3	4	5	6	7
4.	I tend to see difficulties in every opportunity rather than opportunities in every difficulty	1	2	3	4	5	6	7
5.	On the whole, I have a gloomy perspective on most things	1	2	3	4	5	6	7
6.	I don't have a lot of happy memories	1	2	3	4	5	6	7
7.	Understanding the needs and desires of others is not a problem for me	1	2	3	4	5	6	7
8.	I generally believe that things will work out fine in my life	1	2	3	4	5	6	7
9.	I often find it difficult to recognise what emotion I'm feeling	1	2	3	4	5	6	7
10.	I'm not socially skilled	1	2	3	4	5	6	7
11.	I find it difficult to tell others that I love them even when I want to	1	2	3	4	5	6	7
12.	Others admire me for being relaxed	1	2	3	4	5	6	7
13.	I rarely think about old friends from the past	1	2	3	4	5	6	7
14.	Generally, I find it easy to tell others how much they really mean to me	1	2	3	4	5	6	7
15.	Generally, I must be under pressure to really work hard	1	2	3	4	5	6	7
16.	I tend to get involved in things I later wish I could get out of	1	2	3	4	5	6	7
17.	I'm able to "read" most people's feelings like an open book	1	2	3	4	5	6	7
18.	I'm usually able to influence the way other people feel	1	2	3	4	5	6	7
19.	I normally find it difficult to calm angry people down	1	2	3	4	5	6	7
20.	I find it difficult to take control of situations at home	1	2	3	4	5	6	7
21.	I generally hope for the best	1	2	3	4	5	6	7
22.	Others tell me that they admire me for my integrity	1	2	3	4	5	6	7
23.	I really don't like listening to my friends' problems	1	2	3	4	5	6	7
24.	I'm normally able to "get into someone's shoes" and experience their emotions	1	2	3	4	5	6	7
25.	I believe I'm full of personal weaknesses	1	2	3	4	5	6	7
26.	I find it difficult to give up things I know and like	1	2	3	4	5	6	7
27.	I always find ways to express my affection to others when I want to	1	2	3	4	5	6	7
28.	I feel that I have a number of good qualities	1	2	3	4	5	6	7
29.	I tend to rush into things without much planning	1	2	3	4	5	6	7
30.	I find it difficult to speak about my intimate feelings even to my closest friends	1	2	3	4	5	6	7
31.	I'm not able to do things as well as most people	1	2	3	4	5	6	7
32.	I'm never really sure what I'm feeling	1	2	3	4	5	6	7
33.	I'm usually able to express my emotions when I want to	1	2	3	4	5	6	7
34.	When I disagree with someone, I usually find it easy to say so	1	2	3	4	5	6	7
35.	I normally find it difficult to keep myself motivated	1	2	3	4	5	6	7
36.	I know how to snap out of my negative moods	1	2	3	4	5	6	7

37.	On the whole, I find it difficult to describe my feelings	1	2	3	4	5	6	7
38.	I find it difficult not to feel sad when someone tells me about something bad that happened to them	1	2	3	4	5	6	7
39.	When something surprises me, I find it difficult to get it out of my mind	1	2	3	4	5	6	7
40.	I often pause and think about my feelings	1	2	3	4	5	6	7
41.	I tend to see the glass as half-empty rather than as half-full	1	2	3	4	5	6	7
42.	I often find it difficult to see things from another person's viewpoint	1	2	3	4	5	6	7
43.	I'm a follower, not a leader	1	2	3	4	5	6	7
44.	Those close to me often complain that I don't treat them right	1	2	3	4	5	6	7
45.	Many times, I can't figure out what emotion I'm feeling	1	2	3	4	5	6	7
46.	I couldn't affect other people's feelings even if I wanted to	1	2	3	4	5	6	7
47.	If I'm jealous of someone, I find it difficult not to behave badly towards them	1	2	3	4	5	6	7
48.	I get stressed by situations that others find comfortable	1	2	3	4	5	6	7
49.	I find it difficult to sympathize with other people's plights	1	2	3	4	5	6	7
50.	In the past, I have taken credit for someone else's input	1	2	3	4	5	6	7
51.	On the whole, I can cope with change effectively	1	2	3	4	5	6	7
52.	I don't seem to have any power at all over other people's feelings	1	2	3	4	5	6	7
53.	I have many reasons for not giving up easily	1	2	3	4	5	6	7
54.	I like putting effort even into things that are not really important	1	2	3	4	5	6	7
55.	I always take responsibility when I do something wrong	1	2	3	4	5	6	7
56.	I tend to change my mind frequently	1	2	3	4	5	6	7
57.	When I argue with someone, I can only see my point of view	1	2	3	4	5	6	7
58.	Things tend to turn out right in the end	1	2	3	4	5	6	7
59.	When I disagree with someone, I generally prefer to remain silent rather than make a scene	1	2	3	4	5	6	7
60.	If I wanted to, it would be easy for me to make someone feel bad	1	2	3	4	5	6	7
61.	I would describe myself as a calm person	1	2	3	4	5	6	7
62.	I often find it difficult to show my affection to those close to me	1	2	3	4	5	6	7
63.	There are many reasons to expect the worst in life	1	2	3	4	5	6	7
64.	I usually find it difficult to express myself clearly	1	2	3	4	5	6	7
65.	I don't mind frequently changing my daily routine	1	2	3	4	5	6	7
66.	Most people are better liked than I am	1	2	3	4	5	6	7
67.	Those close to me rarely complain about how I behave toward them	1	2	3	4	5	6	7
68.	I usually find it difficult to express my emotions the way would like to	1	2	3	4	5	6	7
69.	Generally, I'm able to adapt to new environments	1	2	3	4	5	6	7
70.	I often find it difficult to adjust my life according to the circumstances	1	2	3	4	5	6	7
71.	I would describe myself as a good negotiator	1	2	3	4	5	6	7
72.	I can deal effectively with people	1	2	3	4	5	6	7
73.	On the whole, I'm a highly motivated person	1	2	3	4	5	6	7
74.	I have stolen things as a child	1	2	3	4	5	6	7
75.	On the whole, I'm pleased with my life	1	2	3	4	5	6	7
76.	I find it difficult to control myself when I'm extremely happy	1	2	3	4	5	6	7
77.	Sometimes, it feels like I'm producing a lot of good work effortlessly	1	2	3	4	5	6	7
78.	When I take a decision, I'm always sure it is the right one	1	2	3	4	5	6	7
79.	If I went on a blind date, the other person would be disappointed with my looks	1	2	3	4	5	6	7
80.	I normally find it difficult to adjust my behaviour according to the people I'm with	1	2	3	4	5	6	7
81.	On the whole, I'm able to identify myself with others	1	2	3	4	5	6	7
82.	I try to regulate pressures in order to control my stress levels	1	2	3	4	5	6	7
83.	I don't think I'm a useless person	1	2	3	4	5	6	7
84.	I usually find it difficult to regulate my emotions	1	2	3	4	5	6	7
85.	I can handle most difficulties in my life in a cool and composed manner	1	2	3	4	5	6	7
86.	If I wanted to, it would be easy for me to make someone angry	1	2	3	4	5	6	7

87.	On the whole, I like myself	1	2	3	4	5	6	7
88.	I believe I'm full of personal strengths	1	2	3	4	5	6	7
89.	I generally don't find life enjoyable	1	2	3	4	5	6	7
90.	I'm usually able to calm down quickly after I've got mad at someone	1	2	3	4	5	6	7
91.	I can remain calm even when I'm extremely happy	1	2	3	4	5	6	7
92.	Generally, I'm not good at consoling others when they feel bad	1	2	3	4	5	6	7
93.	I'm usually able to settle disputes	1	2	3	4	5	6	7
94.	I never put pleasure before business	1	2	3	4	5	6	7
95.	Imagining myself in someone else's position is not a problem for me	1	2	3	4	5	6	7
96.	I need a lot of self-control to keep myself out of trouble	1	2	3	4	5	6	7
97.	It is easy for me to find the right words to describe my feelings	1	2	3	4	5	6	7
98.	I expect that most of my life will be enjoyable	1	2	3	4	5	6	7
99.	I am an ordinary person	1	2	3	4	5	6	7
100.	I tend to get "carried away" easily	1	2	3	4	5	6	7
101.	I usually try to resist negative thoughts and think of positive alternatives	1	2	3	4	5	6	7
102.	I don't like planning ahead	1	2	3	4	5	6	7
103.	Just by looking at somebody, I can understand what he or she feels	1	2	3	4	5	6	7
104.	Life is beautiful	1	2	3	4	5	6	7
105.	I normally find it easy to calm down after I have been scared	1	2	3	4	5	6	7
106.	I want to be in command of things	1	2	3	4	5	6	7
107.	I usually find it difficult to change other people's opinions	1	2	3	4	5	6	7
108.	I'm generally good at social chit-chat	1	2	3	4	5	6	7
109.	Controlling my urges is not a big problem for me	1	2	3	4	5	6	7
110.	I really don't like my physical appearance	1	2	3	4	5	6	7
111.	I tend to speak well and clearly	1	2	3	4	5	6	7
112.	On the whole, I'm not satisfied with how I tackle stress	1	2	3	4	5	6	7
113.	Most of the time, I know exactly why I feel the way I do	1	2	3	4	5	6	7
114.	I find it difficult to calm down after I have been strongly surprised	1	2	3	4	5	6	7
115.	On the whole, I would describe myself as assertive	1	2	3	4	5	6	7
116.	On the whole, I'm not a happy person	1	2	3	4	5	6	7
117.	When someone offends me, I'm usually able to remain calm	1	2	3	4	5	6	7
118.	Most of the things I manage to do well seem to require a lot of effort	1	2	3	4	5	6	7
119.	I have never lied to spare someone else's feelings	1	2	3	4	5	6	7
120.	I find it difficult to bond well even with those close to me	1	2	3	4	5	6	7
121.	I consider all the advantages and disadvantages before making up my mind	1	2	3	4	5	6	7
122.	I don't know how to make others feel better when they need it	1	2	3	4	5	6	7
123.	I usually find it difficult to change my attitudes and views	1	2	3	4	5	6	7
124.	Others tell me that I rarely speak about how I feel	1	2	3	4	5	6	7
125.	On the whole, I'm satisfied with my close relationships	1	2	3	4	5	6	7
126.	I can identify an emotion from the moment it starts to develop in me	1	2	3	4	5	6	7
127.	On the whole, I like to put other people's interests above mine	1	2	3	4	5	6	7
128.	Most days, I feel great to be alive	1	2	3	4	5	6	7
129.	I tend to get a lot of pleasure just from doing something well	1	2	3	4	5	6	7
130.	It is very important to me to get along with all my close friends and family	1	2	3	4	5	6	7
131.	I frequently have happy thoughts	1	2	3	4	5	6	7
132.	I have many fierce arguments with those close to me	1	2	3	4	5	6	7
133.	Expressing my emotions with words is not a problem for me	1	2	3	4	5	6	7
134.	I find it difficult to take pleasure in life	1	2	3	4	5	6	7
135.	I'm usually able to influence other people	1	2	3	4	5	6	7
136.	When I'm under pressure, I tend to lose my cool	1	2	3	4	5	6	7
137.	I usually find it difficult to change my behaviour	1	2	3	4	5	6	7
138.	Others look up to me	1	2	3	4	5	6	7
139.	Others tell me that I get stressed very easily	1	2	3	4	5	6	7

140.	I'm usually able to find ways to control my emotions when I want to	1	2	3	4	5	6	7
141.	I believe that I would make a good salesperson	1	2	3	4	5	6	7
142.	I lose interest in what I do quite easily	1	2	3	4	5	6	7
143.	On the whole, I'm a creature of habit	1	2	3	4	5	6	7
144.	I would normally defend my opinions even if it meant arguing with important people	1	2	3	4	5	6	7
145.	I would describe myself as a flexible person	1	2	3	4	5	6	7
146.	Generally, I need a lot of incentives in order to do my best	1	2	3	4	5	6	7
147.	Even when I'm arguing with someone, I'm usually able to take their perspective	1	2	3	4	5	6	7
148.	On the whole, I'm able to deal with stress	1	2	3	4	5	6	7
149.	I try to avoid people who may stress me out	1	2	3	4	5	6	7
150.	I often indulge without considering all the consequences	1	2	3	4	5	6	7
151.	I tend to "back down" even if I know I'm right	1	2	3	4	5	6	7
152.	I find it difficult to take control of situations at work	1	2	3	4	5	6	7
153.	Some of my responses on this questionnaire are not 100% honest	1	2	3	4	5	6	7

L. The Empathy Quotient Questionnaire (Baron-Cohen & Wheelwright, 2004)

Below is a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree with it. There are no right or wrong answers, and there are no trick questions.

	Strongly Agree	*Slightly Agree*	*Slightly Disagree*	*Strongly Disagree*
I can easily tell if someone else wants to enter a conversation.				
I find it difficult to explain to others things that I understand easily when they don't understand it the first time.				
I really enjoy caring for other people.				
I find it hard to know what to do in a social situation.				
People often tell me that I went too far in driving my point home in a discussion.				
It doesn't bother me too much if I am late meeting a friend.				
Friendships and relationships are just too difficult, so I tend not to bother with them.				
I often find it difficult to judge if something is rude or polite.				
In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.				
When I was a child, I enjoyed cutting up worms to see what would happen.				
I can pick up quickly if someone says one thing but means another.				
It is hard for me to see why some things upset people so much.				
I find it easy to put myself in somebody else's shoes.				
I am good at predicting how someone will feel.				

	Strongly Agree	*Slightly Agree*	*Slightly Disagree*	*Strongly Disagree*
I am quick to spot when someone in a group is feeling awkward or uncomfortable.				
If I say something that someone else is offended by, I think that that's their problem, not mine.				
If anyone asked me if I liked their haircut, I would reply truthfully, even if I didn't like it.				
I can't always see why someone should have felt offended by a remark.				
Seeing people cry doesn't really upset me.				
I am very blunt, which some people take to be rudeness, even though this is unintentional.				
I don't tend to find social situations confusing.				
Other people tell me I am good at understanding how they are feeling and what they are thinking.				
When I talk to people, I tend to talk about their experiences rather than my own.				
It upsets me to see an animal in pain.				
I am able to make decisions without being influenced by people's feelings.				
I can easily tell if someone else is interested or bored with what I am saying.				
I get upset if I see people suffering on news programmes.				
Friends usually talk to me about their problems as they say that I am very understanding.				
I can sense if I am intruding, even if the other person doesn't tell me.				
People sometimes tell me that I have gone too far with teasing.				

	Strongly Agree	*Slightly Agree*	*Slightly Disagree*	*Strongly Disagree*
Other people often say that I am insensitive, though I don't always see why.				
If I see a stranger in a group, I think that it is up to them to make an effort to join in.				
I usually stay emotionally detached when watching a film.				
I can tune into how someone else feels rapidly and intuitively.				
I can easily work out what another person might want to talk about.				
I can tell if someone is masking their true emotion.				
I don't consciously work out the rules of social situations.				
I am good at predicting what someone will do.				
I tend to get emotionally involved with a friend's problems.				
I can usually appreciate the other person's viewpoint, even if I don't agree with it.				

Acknowledgements

A number of people have been instrumental in the writing of this dissertation, both directly and indirectly, and to them I am forever grateful. First, there are the people who supported me at Birkbeck: Professor Rick Cooper, my PhD Supervisor; Ruben Zamora, who provided outstanding technical support; and Ida Akhtar, who ensured all went smoothly from an administrative perspective. From my time at Cambridge, I am grateful to my MEd supervisors, Professor Christine Howe and Dr. Michelle Ellefson, who challenged and encouraged me; Dr. Sara Baker, our Route Co-ordinator, who provided me with guidance and support; and my thesis assessors, Dr. Ros McClellan and Dr. Julia Flutter, whose appreciation for my work spurred me on. Importantly, I am also grateful to my studies' participants, who, as promised, will remain nameless. And lastly, I am most grateful to my partner, Brian McCarter, without whom I could not have done this work.