Individual differences, dual processes, and induction

Aidan Feeney

University of Durham

20 June 2006

To appear in Inductive Reasoning (A. Feeney & E. Heit, Eds.), Cambridge University Press

Please address correspondence to
Aidan Feeney
Applied Psychology
University of Durham Queen's Campus
Thornaby
Stockton-on-Tees TS17 6BH
United Kingdom

aidan.feeney@durham.ac.uk

In this chapter I hope to demonstrate that answering the question 'who does what in reasoning experiments?' can also help us to answer fundamental questions about the nature of thought. Because it is nearly always the case that some experimental participants display phenomena of interest and others don't, almost every experiment run by cognitive psychologists produces data on individual differences. But, as most psychologists do not wish to take differences between their participants as the starting point of their investigations, individual differences tend to be ignored. Thus, the means that experimental psychologists report when they describe their data abstract across differences between individuals, with only the standard deviation or standard error indicating the extent to which participants varied in their responses. In this chapter I will address what the study of differences between individuals might tell us about a range of issues in inductive reasoning including what constitutes a good inference, what processes underlie induction, and how induction differs from deduction.

My primary focus in this chapter will be on the general question of how individual differences data supports dual process theories of thinking (Evans & Over, 1996; Stanovich, 1999; Sloman, 1996). Dual process theories have been applied to deduction, decision-making and induction. I will outline the dual process approach and describe evidence in its favour, paying particular attention to individual differences. I will also describe how individual differences methodology has been used to arbitrate between different normative accounts of specific thinking tasks (see Stanovich & West, 1998b&c). With this background covered I will consider what individual differences might tell us about category-based induction. In particular, what can the method tell us above and beyond more standard experimental methods about people's inductive reasoning abilities? Does the method have anything to say about how a dual process approach might be applied to induction across categories? And can it inform the debate about the normative status of inductive phenomena? I will describe the results of two studies designed with these questions in mind.

I will conclude with a consideration of other questions about induction to which the individual differences methodology might help provide some answers. I will briefly describe a recent study designed to investigate who is susceptible to inductive reasoning fallacies, that is judgements about the strength of inductive arguments that depart from the prescriptions of some normative theory of induction. In addition, I will consider recent claims about dissociations between induction and deduction (Rips, 2001, this volume; Oaksford, this volume; Heit & Rotello, 2005). I will outline a dual-process perspective on this dissociation and argue that an individual differences approach allows one to test the dual-process account. First however, as dual-process theories of thinking have motivated much of the individual differences work I will describe, it is important to review them in some detail here.

Dual-process theories of thinking

It is now a relatively common claim that there are two types of thinking. Roughly speaking, these types correspond to an intuitive mode of thought and a more analytic way of thinking. Recently, there have been at least three attempts to put this claim on a sound theoretical footing (Evans & Over, 1996; Sloman, 1996; Stanovich, 1999; for recent reviews see Evans, 2003, and Osman, 2005). All versions of the theory agree that one type of thinking is fast, associative, and automatic while the other is slow, rule-based and, to some degree at least, controlled. I will refer to the first of these types of thinking as

Type 1 and to the second as Type 2. All three theories claim that the two types of thinking overlap and that they are not confined to any particular thinking task. For this reason the distinction between types of reasoning is orthogonal to the philosopher's distinction between induction and deduction (Sloman, 1996; but see Oaksford, this volume; and Rips, this volume). They also agree that when engaged in the same thinking task the systems can produce conflicting responses. Sloman enumerates a range of thinking tasks where he claims that conflicting responses are available. These range from Tversky & Kahneman's (1983) conjunction fallacy to the case of belief bias in syllogistic reasoning (see Evans, Barston & Pollard, 1983). Evans & Over and Stanovich explicitly claim that the second type of thinking calls on working memory resources and hence, is likely to be predicted by measures of cognitive ability. As Sloman equates Type 2 thinking with symbol manipulation, this claim is implicit in his account.

Although there are marked similarities between what these three theories have to say about different types of reasoning, there are also differences. Sloman is at pains to place his framework on a sound representational footing. Thus, he supports the distinction he draws between types of thinking with arguments taking from the literature on connectionist representation (e.g. Hinton, 1990). In general, his version of the dichotomy is the most clearly specified from a computational point of view. Evans & Over and Stanovich focus on the implications of the dichotomy for conceptions of human rationality. They are also more concerned with the evolutionary basis for their proposals than is Sloman. The claims made by Evans and Over & Stanovich are greater than that made by Sloman because the former authors claim that there are two separate systems for reasoning which developed at different times in human evolutionary history.

Given that claims about the existence of dual processes are almost always contentious, I will briefly review the evidence most often cited to support dual process claims about thinking. Conveniently for the purposes of this overview, the task used to measure belief bias in syllogistic reasoning (Evans, Barston & Pollard, 1983) has been explored using a variety of methodologies and participants. In the syllogisms used in belief bias tasks, participants are presented with two quantified premises and are asked to verify a conclusion provided by the experimenter. For example:

All beekeepers are artists — Argument 1

All artists are archers

Therefore all beekeepers are archers

Participants are presented with syllogisms whose conclusions either necessarily follow or do not follow from the premises and which are believable or unbelievable given background knowledge. A believable and logically valid problem taken from Evans et al. (1983) is as follows:

No police dogs are vicious — Argument 2

Some highly trained dogs are vicious

Therefore, some highly trained dogs are not police dogs

Problem 3 is an unbelievable invalid problem taken from the same experiment.

No millionaires are hard workers — Argument 3

Some rich people are hard workers

Therefore, some millionaires are not rich people

Evans and colleagues found main effects of logic and of conclusion believability in their study. That is, participants were more likely to accept conclusions that were either believable or logically necessary given the premises. Furthermore, they showed that the results were not due to the existence of separate groups of logical and belief-based responders in their sample. Sometimes individual participants responded on the basis of belief and sometimes on the basis of what was necessarily true. This is one of the best examples of thinking tasks that give rise to conflicting responses (see Sloman, 1996). That is Type 1 and Type 2 processes produce different answers in competition with one another. Sometimes the Type 1 answer is accepted whereas on other occasions the Type 2 response overcomes the Type 1 response.

There are a number of lines of investigation that support the dual-process interpretation of Evans et al's results. For example, Gilinsky & Judd (1994) showed that susceptibility to effects of belief on conflict problems increases with age. This makes sense if logical or Type 2 thinking is associated with working memory and I.Q., both of which decline as we get older (e.g. Hasher & Zacks, 1988; Horn & Cattell, 1967). In another line of investigation it has been shown that susceptibility to belief bias on conflict problems may be reduced by stressing the concept of logical necessity in the instructions (Evans, Allen, Newstead & Pollard, 1997). This result is predicted because of the controlled nature of Type 2 thinking processes. Further evidence for the dual-process interpretation comes from a recent study by Evans & Curtis-Holmes (2005) using a speeded response paradigm (for a related use of this paradigm see Shafto & Coley, this volume). When participants were required to make a speeded response they were more susceptible to the effects of belief in conflict problems. To summarise then, responses based on Type 1 processes are primary but responses based on Type 2 processing can dominate if participants are instructed to suppress Type 1 outputs, if they have sufficient time for the slower Type 2 processes to operate, or if they are young.

Before proceeding it is worthwhile to consider two other recent sources of particularly compelling evidence. Goel and Dolan (2003) have reported the results of an MRI study of belief bias in which they found very strong evidence for separate types of thinking. When participants gave belief-based responses on conflict problems they observed increased activation in ventral medial prefrontal cortex (VMPFC). When these same participants gave logical responses on conflict problems they observed increased activation in right inferior prefrontal cortex. Goel and Dolan argue that VMPFC is associated with the effects of beliefs in reasoning (see also Adolphs, Tranel, Bechara, Damasio & Damasio, 1996) whereas activation in inferior prefrontal cortex is due to the inhibition of a response based on Type 1 thinking processes in favour of a Type 2 or logical response. The notion that inhibition is important to logical reasoning is further supported by some data on children's reasoning reported by Handley, Capon, Beveridge, Dennis & Evans (2004). These authors found that scores on the conflict index in children aged ten years were correlated with performance on a task designed to measure ability to inhibit a prepotent response. Children who were more likely to respond logically had higher inhibitory control. Both of these studies suggest the existence of two systems for reasoning, one of which is dominant and whose effects must be suppressed by the other.

Handley et al's study is interesting in the current context because it employed an individual differences methodology. However, the individual differences variable measured was response inhibition and, although there is currently considerable interest in the relationship between measures of inhibition and working memory (see Kane & Engle, 2003), there has been little other work on inhibition and thinking. By far the most popular individual differences variable to have been employed by researchers interested in dual process theories of thinking is general cognitive ability. It is to a review of the literature on cognitive ability and thinking that I will now turn

Individual differences in cognitive ability and thinking

In a series of papers Keith Stanovich and his colleagues have described the results of a large-scale research program designed to investigate the relationship between cognitive ability and thinking. For the most part using SAT scores as an indicator of ability, these researchers have investigated which tasks are predicted by ability and which are not. This program of research is notable because of the range of deductive, decision-making, and (to some extent) inductive tasks it has investigated. It is motivated by two arguments, the first of which is that individual differences studies allow one to test dual process theories of thinking. Explicit in Evans and Over’s and Stanovich’s formulation of the theory (and implicit in Sloman’s) is the claim that Type 2 processes are subject to capacity limitations whereas Type 1 processes are not. Thus, on tasks where Type 2 processes are required in order for the normatively correct answer to be produced, participants who are lower in cognitive resources will tend to perform more poorly than those who are high in cognitive resources.

There are a number of lines of work in the literature linking IQ measures to general cognitive constraints. First, performance on intelligence tests is held to be constrained by basic cognitive abilities. For example, Carpenter, Just & Shell (1990) gave an account of individual differences in the series completion problems used in Raven’s Progressive Matrices test. Their account implicated variations in the ability to induce abstract relations and the ability to coordinate goals in working memory as the basis for individual differences. In support of this account, Unsworth & Engle (2005) have recently shown that individual differences in working memory predict performance on Raven. Although other evidence (e.g. Fry & Hale, 1996) suggests that processing speed mediates the relationship between memory and IQ, nonetheless there is very good evidence for relationships between constraints on basic cognitive abilities and performance on tests of ability. As SAT scores are known to be very closely related to IQ (Frey & Detterman, 2004) it is reasonable to conclude that the same constraints limit SAT performance.

According to Stanovich (1999) the finding that correct performance on certain tasks, but not others, is associated with SAT scores lends support to the suggestion that there are at least two types of thinking and that each is involved to different extents in the accomplishment of a variety of thinking tasks. Stanovich (1999, pg. 143) claims that large differences in performance due to cognitive ability will only be found for thinking tasks that engage both systems. When a thinking task engages the associative system only then no difference will be observed. In addition, individual differences are more likely to be found when the systems produce conflicting responses.

Stanovich illustrates this argument with reference to performance on deontic and indicative versions of Wason's selection task (see Stanovich & West, 1998b). In the selection task participants are presented with a conditional rule that is said to govern a set of instances. In standard indicative versions of the task the rule governs what is printed on each side of sets of cards. For example "If there is an A on one side of the card (p), then there is a 3 on the other side (q)". Participants are shown four cards. On the visible sides are printed an A (p), an L (not-p), a 3 (q) and a 7 (not-q). Participants are asked which card or cards need to be turned over to test the rule. Deontic versions of the selection task concern what should or ought to be done. Most often, they concern social laws, obligations or permissions. For example, “If someone is drinking alcohol in a bar (p), then they must be over 18 (q)”. Printed on the visible side of the cards in this example might by “Beer” (p), “Coke” (not-p), “21” (q) and “16” (not-q).

One of the earliest findings in the literature on the indicative selection task (see Wason, 1966) was that people tended not to select the cards that could falsify the rule (the p and the not-q cards in the example above). However, when deontic content is included (see Johnson-Laird, Legrenzi & Legrenzi, 1972) people regularly select the falsifying cases. There are many explanations of why people behave differently on the indicative and deontic tasks (see Cheng & Holyoak, 1985; Cosmides, 1986; Manktelow & Over, 1991; Oaksford and Chater, 1994). Stanovich argues that most explanations suggest that people’s selections on the deontic task are not determined by conscious, analytic reasoning but by largely unconscious, heuristic processes. Correct responding on the indicative task, on the other hand, requires that the analytic system overcome the unconscious associative system which, in most participants, is responsible for the selection of those cards that match the items in the rule.

If Stanovich is right, then correct performance on the indicative task should be associated with cognitive ability whereas correct performance on the deontic task should not. Stanovich & West (1998b) confirmed this prediction. They found clear evidence that people who reported higher SAT results were more likely to select the potentially falsifying cards on the indicative selection task. There was much less evidence for such an association on the deontic task. Broadly similar results have been reported by Newstead, Handley, Harley, Wright & Farrelly (2004) who also found that whether associations between ability and performance on deontic and indicative tasks are found depend on the ability range of the participants sampled.

The second argument that Stanovich uses to motivate the individual differences approach relates to what counts as the correct normative response on any particular task. There has been substantial debate about which is the correct normative account of tasks as varied as Kahneman & Tversky's (1983) Linda problem and Wason's selection task (see Cohen, 1981; Oaksford and Chater, 1994). Stanovich & West (1998b&c) argued that an association between cognitive ability and certain patterns of performance on these tasks might favour one normative account over another (see also Cohen, 1981). For example, Oaksford and Chater (1994) have recast Wason's selection task as one of deciding which card or cards is likely to be most informative. From a Bayesian perspective, given certain assumptions about the probability of items mentioned in the conditional rule, it turns out that the most informative cards on the indicative task are the p and q cards. Oaksford and Chater argue that their Bayesian reanalysis of the task is closer to a correct normative analysis than is the Popperian notion of falsification which claims that the 'correct' solution is to select the p and not-q cards.

To arbitrate between these competing normative accounts Stanovich and West looked for associations between cognitive ability and the tendency to select cards in accordance with each of these distinct normative approaches. They found that more able participants were more likely to select the falsifying cards than were less able participants. Less able participants, on the other hand, were more likely to select the p and q cards. Stanovich concludes that the individual differences data suggest that from a normative point of view, it is more appropriate to think about the indicative selection task as one requiring falsification rather than decisions about expected information gain. Of course this argument may not always hold. For example, if one happens to sample from the bottom of the range of ability scores then the most able participants in a particular sample may very well perform in non-normative ways on a particular thinking task (see Newstead et al., 2004). However, if one consistently finds, across a range of samples and abilities, that the most able participants consistently favour one response over another, then this lends support to the claim that the response favoured by these participants is the one that is normatively correct.

Unfortunately, although Stanovich and colleagues have investigated a number of phenomena (including belief bias) very little data has been published on the relationship between inductive reasoning and general cognitive ability. One exception is work by Stanovich & West (1998a) on a task where participants make a decision in the face of conflicting evidence comprised of a vivid testimonial from a friend and a statistical summary of the experiences of a number of people (see also Jepson, Krantz & Nisbett, 1983; Fong, Krantz & Nisbett, 1986). This task satisfies Stanovich’s (1999) conditions for the observation of an association between intelligence and normatively correct responses. This is because in order to resist a response based on the vivid single case, Type 2 processes are required to decontextualise the problem. A response based on the vivid non-statistical information is likely to be delivered by context-sensitive Type 1 processes. Stanovich & West (1998a) report correlations of .33 and .38 between the tendency to resist the effects of the salient single case in favour of the statistical evidence and a composite measure of thinking ability. This association suggests that sensitivity to evidence quality in inductive reasoning is related to cognitive ability and hence, reflects the operation of Type 2 thinking processes.

Individual differences and Induction: The Case of Diversity

What might an individual differences approach tell us about category-based induction where people are asked about their willingness to project a (usually) blank property from members of one or more premise categories to members of another conclusion category? One potential use of the methodology is to explore the diversity effect (see Heit, this volume; Shafto, Coley & Vitkin, this volume). Consider Arguments 4 and 5 below.

Tigers have property X — Argument 4

Cows have property X
____

All mammals have property X

Tigers have property X — Argument 5

Lions have property X
____

All mammals have property X

Participants are said to be sensitive to evidential diversity if they prefer arguments based on dissimilar rather than similar premise categories. Because the premises in Argument 4 concern more diverse categories than do the premises in Argument 5, Argument 4 is said to be stronger than Argument 5. As we will see, an individual differences approach may be useful in approaching normative and psychological questions about diversity.

The normative argument for diversity

Although there is general agreement that, all else being equal, diverse evidence is strong evidence, it has been difficult to formally prove this intuition to be true. As a consequence there has been disagreement in the literature about whether diversity is a sound normative principle for inductive reasoning. Sensitivity to diversity is advocated by a range of philosophers of science (Nagel, 1939; Carnap, 1950; Hempel, 1966), and in the psychological literature the soundness of diversity is assumed in a variety of models (see Osherson et al., 1990; Sloman, 1993; Heit, 1998). In addition, a variety of Bayesian justifications for its soundness have been offered (see Howson & Urbach, 1989).

However, Bayesian accounts of diversity have been called into question (Wayne, 1995; for a reply see Myrvold, 1995), and arguments against the generality of a diversity principle have been made by Lo, Sides, Rozelle & Osherson (2002) and by Medin, Coley, Storms & Hayes (2003). Medin et al have demonstrated a number of exceptions to diversity (for an experimental reappraisal of one of these exceptions see Heit & Feeney, 2005). Other exceptions to diversity include those found in cross-cultural work comparing Itzaj-Mayan Indians to North American students. Although the students displayed sensitivity to diversity (see Lopez, Atran, Coley, Medin & Smith, 1997), members of the Itzaj tribe did not. Proffitt, Coley & Medin (2000) showed that some tree experts do not show strong diversity effects when reasoning about trees. Reviewing all of this work, Heit, Hahn & Feeney (2005) have argued that these exceptions to diversity do not invalidate the diversity principle. Instead, they may involve the sensible use of background knowledge beyond simple similarity information when evaluating arguments. Knowledge about diversity is but one of many sources of knowledge that might be used to evaluate inductive inferences.

Heit et al's attempt at reconciliation notwithstanding, there is clearly some disagreement about the normative status of the diversity principle. In some ways (but not in others), this disagreement parallels the normative debates we encountered when considering Stanovich's individual differences approach to thinking. Below I will describe two individual difference studies that apply Stanovich's logic to the diversity effect. These studies examined whether there is a positive correlation between sensitivity to evidential diversity and cognitive ability. If such a correlation were to be observed, then, by Stanovich's argument, we would have additional reason to believe that, all things being equal, sensitivity to diversity is normatively sensible strategy.

Dual-process predictions

Using an individual differences approach to study diversity will also enable us to give preliminary consideration to how a dual process view might be applied to category-based induction. Certainly, at least one theory of induction makes clear dual-process predictions. Sloman - a dual process theorist - has also put forward an important account of category-based induction called the feature-based model (Sloman, 1993). According to this model, category-based induction is explicable without the need to assume a stable category structure. In the FBM the premises of an argument are encoded by training a connectionist network to associate the property to be projected with the features of the categories in the premises. Argument strength is determined by the amount of activation of the target property in the presence of the conclusion category. The greater the featural overlap between premise and conclusion categories, the stronger the argument. In this approach, diverse premises lead to associations with a wider variety of features and hence are more likely to lead to strong activation of the target property in the presence of the conclusion category. This similarity-based model is associative in nature. It claims that category-based induction is achieved by Type 1 processes and, therefore, that the tendency to be sensitive to evidential diversity should not be associated with a measure of cognitive ability. Our individual differences studies will allow us to test this prediction of the FBM.

It is somewhat more difficult to see how dual process theory might map onto other theories of category-based induction. For example, according to Osherson and colleagues' Similarity-Coverage Model (SCM), the strength of a category-based inductive argument is determined by the similarity of the categories in the premises to the category in the conclusion and the degree to which the categories in the premises cover the superordinate category that includes all categories mentioned in the argument (Osherson, Smith, Wilkie, Lopez & Shafir, 1990). It is the coverage component of the model that accounts for diversity effects. The coverage measure is the average maximum similarity between the categories in the premises and instances of the superordinate category that come to mind when assessing the argument. As diverse categories in the premises are likely to be similar to a wider variety of instances from the superordinate, the coverage measure in the model tends to be elevated by the presence of diverse premise categories. This model is different from the FBM because it assumes the existence of a stable category structure. It also predicts that for arguments like 6 and 7, where the conclusion category is specific rather than general, in order to display sensitivity to diversity, people must generate the lowest level superordinate category that contains the categories in the premises and the conclusion.

Bears have property X Argument 6

Gorillas have property X

Tigers have property X

When the conclusion in the argument is general, however, as it is in Argument 4, then people do not have to generate a covering category.

Bears have property X Argument 7

Mice have property X

Tigers have property X

The need to generate a covering category has been implicated by Lopez, Gelman, Gutheil & Smith (1992) as an important factor in the development of category based induction. Lopez et al. found developmental trends where sensitivity to diversity emerged for general arguments before specific arguments. An individual differences study will allow for an alternative test of Lopez's argument. If generating a covering category is an additional source of difficulty then sensitivity to diversity for specific arguments should be related to cognitive ability. In addition, we should observe less sensitivity to diversity when the conclusion category is specific rather than general. Perhaps we might also expect to observe a weaker relationship between sensitivity to diversity and ability with specific conclusions, as only the most able people in any sample may display such sensitivity at above chance levels.

Medin and colleagues' relevance account (Medin, Coley, Storms & Hayes, 2003) attributes diversity effects to participants' hypotheses about the nature of the blank feature. Although they focus on the roles played by causal and environmental knowledge in producing exceptions to diversity, when no such knowledge is available then participants will use the degree of similarity between the categories in the premises to guide their hypothesis formation. Dissimilar categories lead to hypotheses that the feature is general and hence highly projectible, whereas similar categories lead to hypotheses about specific and hence non-projectible features (see also Heit, 1998).

The relevance account gives no grounds for dual process predictions in its current form. However, the Relevance theory of linguistic pragmatics (Sperber & Wilson, 1986/1995) upon which Medin et al's account is based, uses logical rules for non-demonstrative inference (see Sperber & Wilson, 1995, chapter 2). Thus, people work out the implications of an utterance by deriving deductive conclusions from assumptions made available by the context and what is explicated by the utterance. For such an approach to be applied to category-based induction, it is likely that people would have to explicitly represent principles of inductive inference and draw inferences about what follows from those rules and information in the utterance. For example, given dissimilar premise categories, people might infer that the blank property is general and, based on this interim conclusion and the further premise that general properties are highly projectible, might conclude that members of the category in the conclusion are also likely to possess the property. Certainly Medin et al. do not commit themselves to such a thorough application of relevance theory. Nonetheless, under such an application category-based induction might come to have a rather explicit or Type 2 character and one might predict that sensitivity to an inductive principle such as diversity would be associated with cognitive ability.

Finally, Heit (1998) has offered a Bayesian account of induction pitched at a computational rather than an algorithmic level. As a computational-level account specifies what the goals of people's cognition should be (see Marr, 1982), The Bayesian account might make the simple prediction that if there are differences in people's sensitivity to diversity, then they should be predicted by ability. If there is variability in the extent to which people attain their cognitive goals then people higher in cognitive capacity should be more likely to attain them.

Some Recent Evidence

In both of the studies to be described here we presented participants with a series of category-based inductive arguments and the AH4 (Heims, 1968), a test of cognitive ability previously used to study individual differences in induction (see Newstead et al., 2004). The test has verbal/numerical and spatial components, each of which has 65 items and is administered separately under timed conditions.

In the first study 100 students at the University of Durham were presented with 48 inductive arguments. Half of these had general conclusions and half had specific conclusions. Half of the items were designed to measure sensitivity to diversity. The other items were control items designed to examine monotonicity effects and will not be described here. There were twelve pairs of diversity items. One item in each pair concerned similar premise categories while the other concerned dissimilar premise categories. One of the categories in the pair always remained the same. For example, in one pair the diverse argument concerned hippos and hamsters while the non-diverse argument concerned hippos and rhinos. All of the arguments concerned blank premises such as Property X or Property B. Participants' task was to indicate on a percentage scale the extent to which they thought the conclusion was supported by the premises. For the purposes of statistical analyses we counted the number of pairs where the strength rating for the diverse argument was stronger than for the non-diverse argument.

There are two analyses of interest here. In the first we examined whether overall, participants were sensitive to diversity at a rate greater than chance. We found significant diversity effects for arguments with general conclusions (sensitivity in 58% of trials) but not for arguments with specific conclusions (sensitivity in 46% of trials). In general, people do not appear to be highly sensitive to premise diversity when evaluating arguments. The fact that they were significantly less sensitive for arguments with specific conclusions, $t(99) = 4.05$, $p < .001$, lends some support to the claim made by the SCM that the generation of a superordinate category for arguments with a specific conclusion is an extra source of difficulty.

The mean AH4 score for this study was 95.6 (S.D. = 14.4). This is close to the norm for this kind of sample (mean 96.4, S.D. = 15.01). When collapsed across conclusion type there was a significant association between AH4 score and sensitivity to diversity ($r = .27$, $p < .01$). This association was significant when items with general conclusions were considered on their own ($r = .27$, $p < .01$). However, when items with specific conclusions only were analysed the association between AH4 scores and sensitivity to diversity was no longer significant ($r = .16$, $p > .1$). Closer analysis of response patterns revealed that not one participant displayed diversity for all six pairs of arguments with specific conclusions whereas 10 participants displayed maximum sensitivity in the general condition. The requirement to generate a superordinate covering category seems to have rendered sensitivity to diversity very difficult for the vast majority of participants in this sample

In a second study we achieved greater control over our experimental materials than we did in the first. In the first study we used different materials for the diversity-specific and diversity-general conditions. In the second study we used the same materials for each condition. One hundred and fifteen participants at the University if Durham completed the AH4 and 36 reasoning problems. Eighteen of these problems concerned monotonicity and will not be discussed here. The set of 36 problems was composed of six sets of six. Each set of six employed the same five mammal categories three of which were premise categories (mice, dolphins, squirrels), one of which was a specific conclusion category (dogs), and one a general conclusion category (we used mammal for all six problem sets). For the specific and general diverse items, two taxonomically dissimilar mammal categories were presented in the premises (mice, dolphins) whereas for the non-diverse items two taxonomically similar items were presented (mice, squirrels). Problems were presented in one of four different random orders.

Once again we found that sensitivity to diversity was significantly greater than chance for arguments with general conclusions (59%) but not for arguments with specific conclusions (52%). The significant difference in sensitivity to diversity due to conclusion type, $t(114) = 2.96$, $p < .005$, is further support for the claim that the requirement to generate a covering category interferes with diversity-based reasoning.

Although the raw diversity findings from this study are similar to those obtained in Study 1, there are differences in the individual differences data. Mean performance on the AH4 was 106.3 (S.D. = 10.65). There are two things to note about this result. First, it is significantly greater than the norm for a sample of university students (mean = 96.4; S.D. = 15.01). As different degree courses have different academic entry requirements, the difference between the mean AH4 scores observed in these studies is most likely due to the degree course from which the sample for each study was drawn. Second, the standard deviation of the mean for this sample is considerably smaller than the normal standard deviation. Accordingly, all of the correlation coefficients reported below will be adjusted for truncated variance.

When collapsed across conclusion type we observed a significant association between sensitivity to diversity and AH4 scores ($r(adj) = .28$, $p < .005$). The association between reasoning performance and ability is significant when we consider arguments with general conclusions on their own ($r(adj) = .23$, $p < .02$), and when we consider arguments with specific conclusions on their own ($r(adj) = .25$, $p < .01$).

Interpreting the data

Across these two studies we have obtained evidence that bears on theories of inductive reasoning, specific claims about the normative basis for the diversity effect, and on the general claim that there are two types of thinking. First, the finding that people tend not to be sensitive to premise diversity when the conclusion is specific suggests that the SCM may be correct in its prediction that having to generate a covering category is a source of difficulty in induction. An examination of overall sensitivity to diversity across both studies might tempt one to argue that people never reliably generate the covering category and therefore, to conclude that evidence for the diversity effect with specific conclusions has been overstated. However, the individual differences analysis of Study 2 allowed us to observe a significant association between cognitive ability and sensitivity to diversity in specific arguments. Although we did not find evidence for such an association in Study 1, we appear to have sampled a higher part of the distribution of ability scores in Study 2. As Newstead et al. (2004) have demonstrated in the case of the selection task, whether an association is observed between ability and performance on difficult thinking tasks often depends on the nature of the sample. In a high ability sample, those who are highest in ability tend to be sensitive to premise diversity with specific conclusions. This appears to support the prediction from the SCM that the requirement to generate a covering category is an additional source of difficulty in induction. Sensitivity to diversity under these circumstances appears to be relatively rare.

The association that we observed between sensitivity to diversity and intelligence, by Stanovich's normative argument, provides some evidence that, all things being equal, sensitivity to diversity is normatively sound. One concern about this argument relates to the point in the previous paragraph: in any particular study whether certain effects are found to be associated with cognitive ability will depend on the make up of the sample. In Study 1 we found an association between sensitivity to diversity and ability. However, the sample in Study 1 was of average ability. It was possible that, for the higher ability sample in Study 2, no association would be found. Although we did find a relationship in Study 2, a failure to do so would have been very problematic for the normative argument. In principle, it is always possible that an association will not be found with a sample that is of higher average ability than the current sample. For this reason, the normative argument is strongest for those phenomena that have been tested by a number of studies across a range of abilities.

Our data also have implications for dual-process theories. In particular, they have implications for the prediction made by the FBM that category-based induction is achieved as a result of Type 1 processes and may be modelled wholly in associationist terms (Sloman, 1993). We have seen here that individual differences in cognitive ability predict sensitivity to diversity. Associations such as this one are taken by dual-process theorists to indicate that Type 2 processes are at play. In this case, such a conclusion is problematic for the FBM. It is also possible that people high in cognitive ability may have richer representations of categories than people lower in ability. This would mean that their associative systems had more effective knowledge bases to make inferences over. It should also be pointed out that Sloman (1996; 1998) in discussions of dual processes for reasoning…..Type 2 processes might be involved in the diversity effect because people may explicitly represent principles of inductive reasoning. Upon noticing that the categories in the premises are dissimilar, people may bring to mind the principle that diverse evidence leads to strong inductive inferences. This possibility has also been suggested by McDonald, Samuels & Rispoli (1996), and, as we noted above, is consistent with certain applications of relevance theory to category-based induction.

Other Applications of the Approach

As we have seen, the individual differences paradigm may be very useful in helping us to answer questions about diversity. There are a range of other questions raised in other chapters of this book to which the approach might usefully be applied. Some of these questions concern normative issues whilst others are more psychological.

Normative issues

There exist several cases in the literature on induction where the findings predicted by a particular normative account are not observed or, alternatively, phenomena are observed that are not predicted by the normative account. Consider Murphy and Ross's (this volume) description of data from inductive tasks involving uncertain categorisation that appear to be inconsistent with a rational analysis. For example, Murphy, Malt & Ross (1995) asked participants to say how likely an individual described as walking towards a house would be to check the sturdiness of the doors of the house. In one condition the individual was either an estate agent or a cable TV repairman. In another he was either an estate agent or a burglar. In both conditions the individual was most likely to be an estate agent. Because pretests had shown that people believe that burglars would be more likely to check the doors than a TV repairman, participants should have rated more highly the likelihood that the uncategorized individual will check the windows when there is a possibility that he is a burglar. However, in this experiment and many others involving abstract as well as concrete materials, the alternative category made no difference to people's likelihood judgements.

A related demonstration concerns people's inductive inferences based on cross-classifiable categories such as 'bagel' which may be classified as a bread or a breakfast food. Here, Ross & Murphy (1999) find that although a majority of participants rely on just one of the categories when evaluating the likelihood that a target will possess some property, a minority take both possible classifications into account.

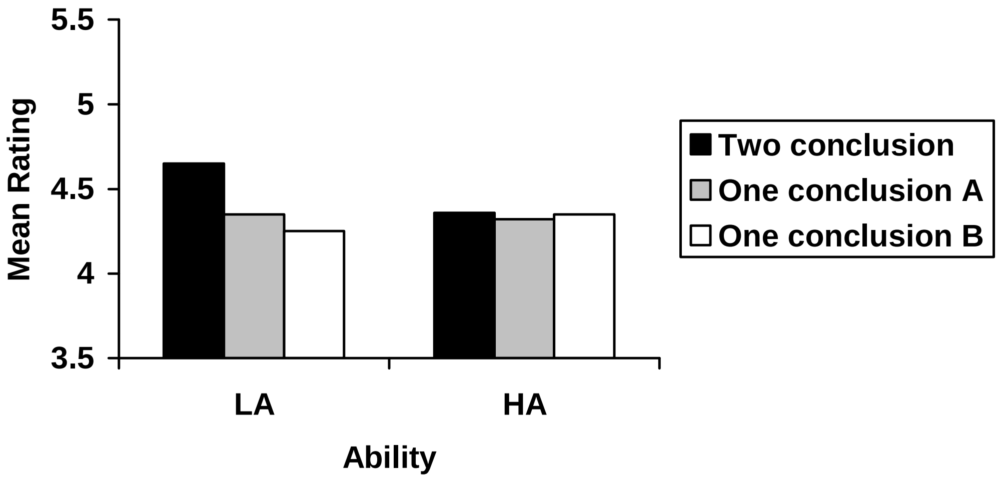
These findings pose a problem for Bayesian accounts of induction because they suggest that people are, in certain respects, poor Bayesians. However, as Murphy and Ross remark in their chapter, their problems resemble some of the problems, such as the taxi-cab problem, used by Kahneman & Tversky to study decision making. Just as Ross and Murphy find that increasing the relevance of the alternative category increases people's tendency to take it into consideration, so Kahneman & Tversky observed that making the base rate causally relevant increased the rate at which it played a role in people's judgements. Stanovich & West (1998a) have demonstrated an association between cognitive ability and base rate neglect so that the most able participants in a sample are less likely to ignore the base rate. A similar question may be asked about induction under uncertainty. Will the tendency to ignore the alternative category be predicted by ability? Perhaps Murphy and Ross's means hide predominantly high ability participants who take the alternative into account. This may be especially true of the cross-classification studies where a minority take both classifications into account. Perhaps this minority tend to be highest in ability. If so, then we would have a situation, similar to that which pertains in several other areas of high-level cognition (for a review see Stanovich, 1999) where only a minority of participants behave in accordance with the norm for the task. Because those participants tend to be the most able, the observation of non-normative behaviour overall is less problematic.

Another example of an experimental effect that is not predicted by Bayesian accounts of induction is the conjunction fallacy (Medin et al., 2003), where adding categories to the conclusion causes the argument to become stronger. Medin et al. observed two types of conjunction fallacy. The conjunction fallacy by property reinforcement is observed when the addition of a second category to the conclusion of an argument causes the feature shared by the categories in the argument to become more relevant. For example Medin et al. observed that the argument from chickens to cows and pigs was stronger than that from chickens to cows or from chickens to pigs. Medin et al. account for this finding by suggesting that the environment shared by members of all three categories becomes more available when all three are present in the argument. The causal conjunction fallacy occurs when the addition of an extra category to the conclusion causes the causal relationship between the categories in the argument to become relevant. Thus, the argument from grain to mice and owls is stronger than that from grain to mice or from grain to owls. Of course, in probabilistic terms this is non-normative because the probability of a conjunction of events can never be higher than the probability of either one of the conjuncts.

Again, it may be the case that these fallacies are resisted by participants who are highest in ability. Indeed, Stanovich & West found an association between ability and resistance to the conjunction fallacy in Kahneman & Tversky's famous Linda problem where many participants estimate it to be more likely that Linda is a feminist bankteller than that she is a feminist or a bankteller. As has been the case with other types of thinking, an individual differences approach may very well resolve apparent discrepancies between average behaviour and the predictions of normative or computational-level analyses.

A recent study of category-based induction by Feeney, Shafto & Dunning (under review) applied the individual differences approach to both types of conjunction fallacy. One hundred and thirty participants completed the AH4 and 36 category-based arguments. Half of these arguments were designed to test susceptibility to the conjunction fallacy by property reinforcement and the other half to test susceptibility to the causal conjunction fallacy. The eighteen property reinforcement arguments consisted of six items, each comprised of three arguments. One of the arguments in each item concerned all three categories with one category in the premise and the other two in the conclusion. The conclusion of the other two arguments concerned one or other of the conclusion categories from the first argument. These single conclusion category arguments were randomly labelled A or B. Each of the corresponding six causal items also contained a three-category argument and a pair of two-category arguments. Following Medin et al.

(a)



(b)

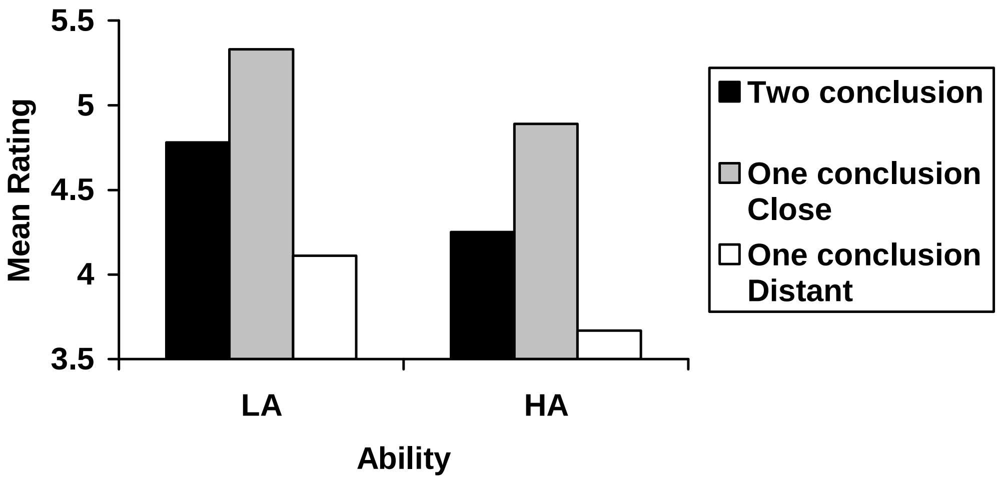


Figure 1: Interaction between Ability and Conclusion Type from Feeney, Shafto & Dunning. (a) refers to property reinforcement materials and (b) to causal materials.

the categories in the three premise arguments always made up a causal chain. For example, grain, mice and owls may share a property because mice eat grain and owls eat mice. Accordingly, because grain is closer to mice in this chain than it is to owls, we labelled one type of single conclusion argument Close (in this example, grain to mice), and the other (grain to owls) as Distant.

Mean score on the AH4 was 94.57 (S.D. = 14.32). As we saw earlier, this is close to the norm for a sample of university students. Our data contained clear evidence for the existence of a conjunction fallacy in category-based induction. Preliminary analysis revealed that participants were susceptible to the fallacy by property reinforcement on 47% of trials and to the causal version of the fallacy on 53% of trials. However, there was a striking difference between the pattern of results for the different versions of the fallacy. For the property reinforcement materials, an ANOVA on ratings of inductive strength revealed a significant interaction between ability and number of conclusions in the argument. However, the equivalent interaction from the analysis of strength ratings for the causal materials did not approach significance. The means in these interactions are to be seen in Figure 1 where it may be seen that high ability participants did not tend to distinguish between two-conclusion category and one-conclusion category property reinforcement arguments whereas participants lower in ability did. Figure 1 also shows that all participants, regardless of ability, tended to respect causal distance in their ratings of argument strength. So, for example, the argument from grain to mice was rated stronger than that from grain to owls. In fact, participants were more likely with these materials to respect causal distance in this way than they were to commit the causal conjunction fallacy.

The results are intriguing. First of all, because the conjunction fallacy by property reinforcement is not observed in more able participants, our results suggest that it may be relatively unproblematic for Bayesian accounts (see Heit, 1998) of category-based induction. However, both high and low ability participants were equally likely to commit the causal conjunction fallacy and to respect causal distance in their ratings of argument strength. The causal conjunction fallacy remains, therefore, highly problematic for Bayesian accounts of induction. Causal knowledge appears to be very compelling when people are evaluating inductive arguments. In this respect our results are similar to those reported by other researchers who have shown the pre-eminence of causal information in induction (see Rehder & Hastie, 2001; Rehder, this volume).

The Relationship Between Induction and Deduction

The final issue to which we will briefly apply the individual differences method is the distinction made by philosophers between deductive or necessary inferences and inductive or plausible inferences. One important question about this distinction is whether it corresponds to a psychological dissociation between the two types of reasoning. There are a variety of views on this issue. One view is that induction and deduction are achieved by the same processes (see Johnson-Laird, Legrenzi, Girotto, Legrenzi & Caverni, 1999). Another view is that there is a clear dissociation between induction and deduction (see Rips, 2001). Support for this view comes from an experiment by Rips (2001; see also Heit & Rotello, 2005) where participants were asked to evaluate the same set of inferences under inductive or deductive instructions. There were four types of inference: valid and causally strong; valid but causally weak; invalid but causally strong;

invalid and causally strong. An example of each of these inferences is to be seen in Table 1. For valid but causally weak and invalid but causally strong problems, Rips observed an

Table 1: Examples of the materials used by Rips (2001) to demonstrate a dissociation between induction and deduction

Valid Strong	*Valid Weak*	*Invalid Strong*	*Invalid Weak*
If Car #10 runs into a brick wall it will stop.	If car #10 runs into a brick wall it will speed up.		
Car #10 runs into a brick wall.	Car #10 runs into a brick wall.	Car #10 runs into a brick wall.	Car #10 runs into a brick wall.
-------------------------	-------------------------	-------------------------	-------------------------
Car #10 will stop.	Car #10 will speed up.	Car #10 will stop.	Car #10 will speed up.

interaction such that the former were rated stronger than the latter under deductive instructions, whereas under deductive instructions this pattern was reversed. Rips argued that this finding constitutes a psychological dissociation between induction and deduction.

A dual process view, on the other hand, is that although there are distinct psychological processes for thinking, these processes do not map onto induction and deduction. According to a dual process view, both Type 1 and Type 2 processes may be involved to a greater or lesser extent in any thinking task, inductive or deductive. When participants distinguish between deductive and inductive reasoning instructions, they do so because they have explicit knowledge about the difference between concepts such as necessity and plausibility and, given different instructions, Type 2 processes allow then to apply different standards for evaluating arguments.

Support for the dual process view comes from an experiment by Feeney, Dunning & Over (in preparation) where 81 students from the University of Durham completed a subset of Rips' problems under induction conditions. Their task was to decide whether the conclusion of each argument was plausible given the information in the premises. Causal strength and validity were independently manipulated and participants saw four arguments of each type. They circled 'strong' if they thought that the argument was plausible and 'not strong' if they thought that it was implausible. As with other studies described in this chapter, participants in this study completed the AH4. Mean score was 94.40 (S.D. = 12.94). We carried out a median split on our data by I.Q. score. The mean score of the 41 participants in the low ability group was 84.24 (S.D. = 8.20) whilst the mean score of the 40 participants in the high ability group was 104.8 (S.D. = 7.42).

A 2(Ability) x 2 (Validity) x 2(Strength) mixed design ANOVA on the proportion of plausible judgements in each condition revealed significant main effects of Validity, $F(1, 79) = 114.59$, $MSe = .07$, $p < .001$, and of Strength, $F(1, 79) = 172.45$, $MSE = .06$, $p < .001$. Whilst the interaction between Ability and Strength was non-significant, $F(1, 79) = 1.05$, $MSE = .06$, $p > .3$, Ability did interact significantly with Validity, $F(1, 79) = 5.14$, $MSE = .07$, $p < .03$. The means involved in this interaction are presented in Figure 2 where it may be seen that High Ability participants judge a greater proportion of valid arguments to be strong than do Low Ability participants, and they judge a grater proportion of invalid arguments to be 'Not Strong' than do Low Ability participants.

It is clear from Figure 2 that overall, logical validity plays an important role in people's judgements of inductive strength. However, what is most striking is that validity plays a greater role in the judgements of participants high in cognitive ability than in the decisions made by participants lower in ability. These results are consistent with the claim that inductive judgements are mediated by two types of process. One type of process is sensitive to content and context whereas the other is sensitive to structure. The first type of process underlies effects of belief in reasoning but because it does not

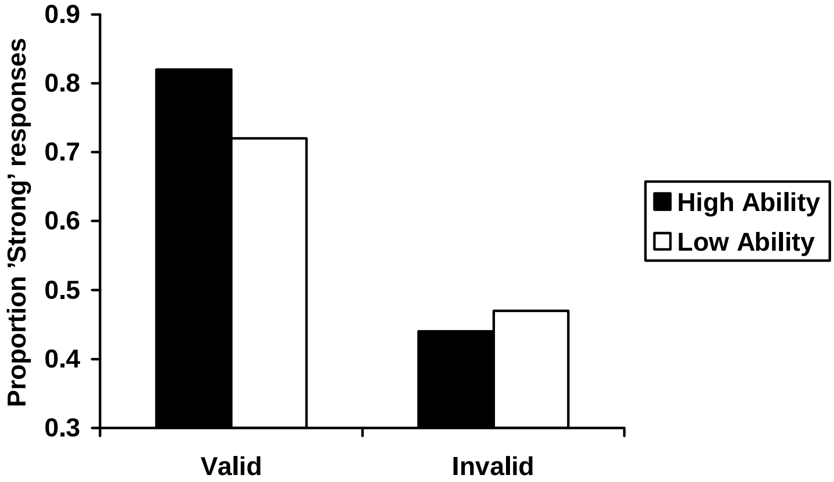


Figure 2: Means involved in the interaction between Ability and Validity from Feeney & Dunning.

depend on cognitive capacity, effects of belief do not interact with cognitive ability. The second type of process underlies our sensitivity to logical validity. Because it is dependent on cognitive capacity, sensitivity to logical validity in this experiment interacts with cognitive ability.

Conclusions

In this chapter I have described how results from studies using the individual difference method have informed recent dual-process theorising about thinking. I have shown how the method may provide valuable insights into the processes involved in induction. For example, sensitivity to evidential diversity is associated with cognitive ability and, in the presence of a specific conclusion, is almost impossible for all but the most able participants. These results suggest that the diversity effect is not explicable in wholly associationist terms and that the generation of a covering category is, as it is predicted to be by SCM, an extra source of difficulty in induction. The method may also have uses in testing Bayesian accounts of inductive reasoning and dual process accounts of the relationship between induction and deduction.

Individual differences are but one of several methods for testing models of induction, and as a consequence, individual differences studies are unlikely to lead to the formulation of an entirely novel theory of induction. However, as the range of approaches described in this volume attests, there is currently a broad range of theories of induction formulated at a number of levels of explanation. One might argue that currently the field does not need new theories so much as it does evidence with which to discriminate between existing theories and approaches. Individual differences studies are likely to be a rich source of discriminating evidence.

References

Adolphs, R., Tranel, D., Bechara, A., Damasio, H., & Damasio, A. R. (1996). Neuropsychological approaches to reasoning and decision-making. In A. R. Damasio, Neurobiology of decision-making. Berlin: Springer-Verlag.

Carnap, R. (1950). Logical foundations of probability. University of Chicago Press.

Carpenter, P. A., Just, M. A. & Shell, P. (1990) What one intelligence test measures: A theoretical account of processing in the Raven Progressive Matrices test. Psychological Review, 97, 404-431.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Cohen, L.J. (1981). Can human irrationality be experimentally demonstrated? Behavioral & Brain Sciences, 4, 317-370.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition, 31, 187-276.

Evans, J.St.B.T. (2003). In two minds: dual process accounts of reasoning. Trends in Cognitive Sciences, 7, 454-459.

Evans, J.St.B.T., Barston, J.L. and Pollard,P. (1983) On the conflict between logic and belief in syllogistic reasoning. Memory & Cognition, 11, 295-306.

Evans, J.St.B.T. & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual process theory of reasoning. Thinking & Reasoning, 11, 382-389.

Evans, J.St.B.T., Newstead, S.E., Allen,J.L. & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. European Journal of Cognitive Psychology, 6, 263-285.

Evans, J.St.B.T. & Over, D.E. (1996). Rationality and reasoning. Hove, UK: Psychology Press.

Feeney, A., Shafto, P. & Dunning, D. (in submission). Who is susceptible to conjunction fallacies in category-based induction?

Fong, G.T., Krantz, D.H. & Nisbett, R.E. (1986). The effects of statistical training on thinking about everyday problems. Cognitive Psychology, 18, 253-292.

Frey, M.C. & Detterman, D.K. (2004) Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. Psychological Science,15, 373—378

Fry, A. F. & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: evidence for a developmental cascade. Psychological Science, *7*, 237-241.

Gilinsky, A.S. and Judd, B.B. (1994) Working memory and bias in reasoning across the life-span. Psychology and Aging, 9, 356–371

Goel, V. and Dolan, R.J. (2003) Explaining modulation of reasoning by belief. Cognition, 87, B11–B22

Handley, S.J., Capon, A., Beveridge, M., Dennis, I., & Evans, J.St.B.T. (2004). Working memory and inhibitory control in the development of children's reasoning. Thinking and Reasoning, 10, 175-195.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), The Psychology of Learning and Motivation, 22, 193-225

Heim, A.W. (1967). AH4 group test of intelligence [Manual]. London: National Foundation for Educational Research.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), Rational models of cognition (pp. 248-274). Oxford: Oxford University Press.

Heit, E. & Feeney, A. (2005). Relations between premise similarity and inductive strength. Psychonomic Bulletin and Review, 12, 340-344.

Heit, E., Hahn, U., & Feeney, A. (2005). Defending diversity. In W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), Categorization inside and outside of the laboratory: Essays in honor of Douglas L. Medin, 87-99. APA.

Heit, E. & Rotello, C. (2005). Are there two kinds of reasoning? Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.

Hempel, C. G. (1966). Philosophy of natural science. Englewood Cliffs, NJ: Prentice Hall.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. Acta Psychologica, 26, 107-129.

Howson, C., & Urbach, P. (1993). Scientific reasoning: The Bayesian approach. Chicago: Open Court.

Jepson, C., Krantz, D. H. & Nisbett, R. E. (1983). Inductive reasoning: Competence or skill? Behavioral and Brain Sciences, 6, 494-501.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M., and Caverni, J-P. (1999). Naive probability: a mental model theory of extensional reasoning. Psychological Review, 106, 62-88.

Johnson-Laird, P. Legrenzi, P. & Legrenzi, S. (1972) Reasoning and a sense of reality. British Journal of Psychology, *63*, 395–400.

Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General, *132,* 47-70.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. Cognitive Science, 26, 181-206.

Lopez, A., Atran, S., Coley, J.D., Medin, D.L. & Smith, E.E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. Cognitive Psychology, 32, 251-295.

López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. Child Development, 63, 1070-1090.

McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis assessment model of categorical argument strength. Cognition, 59, 199-217.

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 646-661.

Manktelow, K. I. & Over ,D. E. (1991). Social roles and utilities in reasoning with deontic conditions. Cognition, 39, 85-105.

Marr, D. (1982). Vision. San Francisco, CA: H. Freeman & Co.

Medin, D. L., Coley, J., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. Psychonomic Bulletin & Review, *10*, 517-532

Murphy, G.L. & Ross, B.H. (1999). Induction with cross-classified categories. Memory & Cognition, 27, 1024-1041.

Myrvold, W.C. (1995). Bayesianism and diverse evidence: A reply to Andrew Wayne. Philosophy of Science, 63, 661-665

Nagel, E. (1939). Principles of the theory of probability. Chicago: University of Chicago Press.

Newstead, S.E., Handley, S. J., Harley, C., Wright, H. & Farrelly, D. (2004). Individual differences in deductive reasoning. Quarterly Journal of Experimental Psychology, 57, 33-60.

Oaksford, M & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, 101, 608-631.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. Psychological Review, *97*, 185- 200.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. Psychonomic Bulletin & Review, 11,988-1010.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 811-828.

Rehder, B. & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. Journal of Experimental Psychology: General, 130, 323-360.

Rips, L. J. (2001). Two kinds of reasoning. Psychological Science, 12, 129-134.

Ross, B.H. & Murphy, G.L. (1999). Food for thought: Cross-classification and category organization on a complex real-world domain. Cognitive Psychology, 38, 495-553.

Sloman, S. A. (1993). Feature-based induction. Cognitive Psychology, 25, 231-280.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22.

Sperber, D. & Wilson, D. (1986/1995). Relevance: Communication and cognition. Oxford: Basil Blackwell.

Stanovich, K. E. (1999). Who is rational? Studies of individual differences in reasoning. Mahwah, NJ: Erlbaum .

Stanovich, K.E. & West, R.F. (1998a). Individual differences in rational thought. Journal of Experimental Psychology: General, 127, 161-188.

Stanovich, K. E., & West, R. F. (1998b). Cognitive ability and variation in selection task performance. Thinking and Reasoning, 4, 193-230.

Stanovich, K.E. & West, R.F. (1998c). Individual differences in framing and conjunction effects. Thinking and Reasoning, 4, 289-317.

Tversky, A., & Kahneman, D. (1983) Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-3l5.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. Intelligence, 33, 67-81.

Wason, P. (1966) Reasoning. In Foss, B. (ed.) New horizons in Psychology. Harmondsworth: Penguin

Wayne, P. (1995). Bayesianism and diverse evidence. Philosophy of Science, 62, 111-121.