How Scientists Reason:
The Use of Unexpected Findings

Lisa M. Baker
Department of Psychology
McGill University, Montreal
May, 1994

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for
the degree of Master of Arts

Abstract

While there is much data in the experimental cognitive psychology literature reporting that subjects working on science-like tasks ignore findings inconsistent with their hypotheses, much cognitive science research has found that reasoners focus on unexpected findings. To study how real-world scientists deal with unexpected findings, data was collected from a prominent immunology laboratory. Four lab meetings were analyzed using a standardized coding procedure. The amount of reasoning, interactions, and new hypotheses about unexpected versus expected findings was analyzed. Presenters at the meetings reasoned more about unexpected than expected findings, and group members reasoned and interacted extensively about unexpected findings. Both presenter and group members formed more new hypotheses about unexpected than about expected findings. These results are consistent with the finding in cognitive science research that reasoners focus on unexpected data. It is proposed that several heuristics influence which unexpected findings scientists pay attention to.

Résumé

Dans la littérature expérimentale de psychologie cognitive, plusieurs études rapportent que les sujets travaillant sur des tâches a caractère scientifique ignorent les données qui ne supportent pas leur hypothèses. Parcontre, une part importante de la recherche en science cognitive a démonstré que les gens pendant que pensant logiquement sont particulièrement attentifs aux résultats suprenants. Dans le but d'étudier comment les vrais scientifiques traitent leur résultats imprévus, des données provenant d'un laboratoire d'immunologie reconnu ont été receuillies. Quatre rencontres de laboratoire ont été analysées à partir d'une procédure de codage standardisée. La quantité de raisonement, interactions, et nouvelles hypothèses générées à partir de résultats imprévus versus prévisibles ont été analysées. Les chercheurs présentant leur résultats lors d'une rencontre de laboratoire raisonnent plus à propos des données imprevues, et les membres d'un groupe raisonnent et interagissent considérablement à propos des résultats imprévus. Les conférenciers ainsi que les membres d'un groupe génèrent plus d'hypothèses à propos de résultats suprenants qu'à propos de résultats prévisibles. Ces résultats sont compatibles avec les résultats de la recherche en sciences cognitives. Il semble que plusieurs stratégies influencent à quel résultat suprenant les scientifiques portent attention.

Acknowledgements

I wish to thank my adviser, Professor Kevin Dunbar, for giving me the opportunity to work on this project, for effective guidance on this thesis, and for teaching me, over the last year and a half, how to do psychology research. I would also like to thank Professors Tony Marley and Vimla Patel for helpful feedback on an earlier draft of this thesis, and Professor Paul Pietrowski for insightful comments on the originally submitted version.

All the people who have worked in Professor Dunbar's lab in the last two years have contributed to this ongoing research on real-world scientific reasoning, and I have valued their insights and support. In particular, I would like to thank Bernadette Han and Faraaz Siddiqi for assisting with the coding of some data analyzed in this thesis. In addition, I would like to thank Martine Turgeon for translating the abstract into French.

Finally, I would like to thank my husband, Jerry Quinn, who made very real sacrifices so I could come to Montreal and study at McGill, which was the best academic choice I have ever made.

I wish to acknowledge the David Stewart Memorial (Major) Fellowship of McGill University, which supported my research this year. In addition, I wish to acknowledge a grant from the Spencer Foundation to Kevin Dunbar and grant #OGP0037356 from the Natural Sciences and Engineering Research Council of Canada to Kevin Dunbar, which supported this research.

Table of Contents

1. Literature Review

Researchers in many disciplines have devoted considerable effort to investigating how scientists reason and make discoveries. For example, cognitive scientists and psychologists have developed theories of scientific reasoning (e g., Holland, Holyoak, Nisbett, & Thagard, 1986; Klahr & Dunbar, 1988; Qin & Simon, 1990) and have conducted laboratory experiments where subjects were asked to discover concepts, rules, or theories (e.g., Dunbar, 1993; Klahr, Fay, & Dunbar, 1993; Mynatt, Doherty, & Tweney, 1977; Wason, 1960). In addition, a number of computational models have been developed to mimic the heuristics scientists used in making discoveries (e.g., Kulkarni & Simon, 1988; Thagard, 1989). Research on scientific reasoning has also been conducted from a historical perspective; historians of science and psychologists have analyzed diary and notebook records of scientists' work (e.g., Tweney, 1989). All of these approaches have contributed greatly to our understanding of how science is conducted.

Cognitive theories of scientific reasoning and discovery have focused on many aspects of the scientific reasoning process; however, the bulk of the work on scientific reasoning has been concerned with induction (e.g., Bruner, Goodnow, & Austin, 1956, Holland et al , 1986, Langley, Simon, Bradshaw, & Zytkow, 1987). This focus came about because researchers were concerned with describing how people build theories from data Induction, broadly defined, is "all inferential processes that expand knowledge in the face of uncertainty" (Holland et al., 1986). Numerous cognitive theories of scientific reasoning focus on the induction of theory from patterns of data (e.g., Bruner et al., 1956; Langley et al., 1978). Other theories that encompass such aspects of scientific reasoning as designing experiments and searching for plausible hypotheses have similarly been motivated by historical evidence, laboratory studies of "science-like" tasks, or prior theory (e.g., Kulkarni & Simon, 1988, 1990; Thagard, 1989). While these approaches have resulted in novel and important theories of scientific reasoning, little is known about how contemporary scientists actually do reason when formulating hypotheses, designing experiments, interpreting findings, and carrying out other components of their work. A particular issue in science is the need to constrain induction, because the range of hypotheses that could be proposed and the types of data that could be investigated are virtually unbounded (cf. Goodman, 1972). Of particular interest, therefore, is empirical and theoretical work on how subjects constrain the inductive process (e.g., Gorman, 1986; Holland et al., 1986, Tweney et al., 1980).

In this chapter, a review of the major empirical and theoretical work concerning scientific reasoning and, in particular, constraints on induction in scientific settings will be given. This discussion of the literature will be followed a description of current research that investigated how real-world scientists reason. This research is the focus of this thesis.

1.1. Empirical Studies of Induction

Classic cognitive research has investigated how subjects induce theories from data A major line of research (e.g , Wason, 1960) has focused on subjects' strategies for testing hypotheses This work has led to the investigation of the strategies that subjects use when they are confronted with data inconsistent with a hypothesis (e.g., Mynatt et al., 1977). In addition, particular strategies that reasoners use have been described, strategies that might cause them not to make effective use of inconsistent data (e.g., Chinn & Brewer, 1992) A description of the hypothesis formation work will be followed by a discussion of various heuristics and constraints that have been identified in the experimental literature as affecting subjects' use of inconsistent data

1.1 1. Induction and Hypothesis Testing A long tradition of experimental cognitive psychology research has focused on how subjects choose experiments to test their hypotheses and how they react to evidence that does not support their hypotheses (e.g , Mynatt et al., 1977; Wason, 1960). The results of much of this research support the conclusion that reasoners exhibit a "confirmation bias " That is, subjects are said to seek out evidence to support their hypotheses and to ignore evidence that does not support their hypotheses Part of the motivation for research on confirmation bias can be found in Popper's (1963) argument that scientists should attempt to "falsify" their hypotheses—search for evidence that could disprove them—rather than trying to confirm their hypotheses. It was in comparison with Popper's ideal scientist that subjects were found wanting.

Confirmation bias research falls into two broad categories. One claim has been that people tend to *seek out* evidence that will confirm their hypotheses. Results from experiments using Wason's (1960) "2–4–6 task" supported this conclusion. In this task, subjects are told that the experimenter has in mind a rule concerning sequences of three numbers, that they are to guess the rule the experimenter has in mind, and that they may propose sequences of numbers and the experimenter will tell them whether the sequences fit the rule. Subjects are also told that the sequence "2–4–6" conforms to the rule. Typically, studies have found that subjects hypothesize a fairly specific rule, such as "all sequences of numbers ascending by twos" and then generate test sequences *that conform to their hypothesized rule* (e.g., Miller, 1967; Snyder & White, 1981; Wason, 1960, 1968). This tendency of subjects to try to "confirm" their hypotheses has been labeled "confirmation bias." Usually, the rule the experimenter actually has in mind is a broad one, "all ascending sequences," that is consistent with a wide variety of instances. Subjects assume that, once several sequences conform to their hypothesized rule, then the rule must be correct. They therefore fail to arrive at the correct rule because they receive no evidence to contradict their initial hypotheses. Subjects' failure to arrive at the correct rule is typically attributed to their strategy of testing sequences that they believe will "confirm" their hypotheses, rather than actively attempting to falsify their hypotheses, as Popper had advocated.

Recently, Klayman and Ha (1987) have suggested a reconceptualization of confirmation bias. They proposed that subjects' strategy of searching for instances that support their hypotheses is more appropriately called a "positive test" strategy. Klayman and Ha argued that in many real-world situations a positive test strategy is more efficient than a negative test strategy at arriving at the correct hypothesis. They pointed out that using a positive test strategy does not mean that only evidence consistent with the hypothesis will be obtained. In the case of typical proposed ("all ascending sequences of even numbers") and target ("all ascending sequences") rules for the 2–4–6 task, where the proposed rule is a subset of the target rule, the positive test strategy does not result in the discovery of evidence inconsistent with the proposed rule. However, in cases where the target rule is a subset of the proposed rule, or where the two rules overlap, use of the positive test strategy would often result in obtaining inconsistent evidence. Klayman and Ha argued that the latter two cases might occur more often in real-world reasoning, and therefore the positive test strategy would in general be an efficient reasoning heuristic. The confirmation bias research described above led to the conclusion that reasoners, and in particular scientists, are subject to a bias that is a major impediment to proposing new theories. However, Klayman and Ha argued that a positive test strategy is instead a useful and often successful tactic for guiding induction.

The second line of inquiry that is sometimes classified as "confirmation bias" work investigates whether people ignore or make use of evidence inconsistent[1] with their hypotheses, once they do obtain it. Mynatt et al. (1977) investigated use of inconsistent evidence by studying how subjects go about discovering rules that govern a computer-modeled "simulated universe." They found that most subjects who obtained unambiguous evidence inconsistent with their hypotheses changed to a new hypothesis. They concluded: "When confronted with unambiguous falsifying evidence, [our subjects] utilized it in precisely the correct way—by rejecting their incorrect hypotheses just as Popper said they should" (Mynatt et al., 1977, p. 94). However, in a follow-up study, Mynatt, Doherty, and Tweney (1978) claimed that subjects who worked with a more complex version of the same task "discounted or disregarded" inconsistent evidence. In most cases, subjects maintained their current hypotheses despite

---

[1]In the experimental psychology literature, the terms "disconfirmatory" and "falsifying" are often used to describe inconsistent data. They will be avoided here, because they carry the psychological connotation that the researcher will recognize the data as failing to confirm or as falsifying a hypothesis, which is not necessarily the case. Instead, the term "inconsistent" will be used.

inconsistent data. A closer examination of Mynatt et al.'s (1978) data carried out by Dunbar (1993) demonstrated that subjects actually did make some use of inconsistent evidence. Dunbar noted that the subjects in the Mynatt et al. studies used inconsistent evidence not only to abandon hypotheses permanently but in some cases to revise their hypotheses or to abandon them temporarily. However, the general conclusion that has been drawn from Mynatt et. al's (1978) work is that subjects ignore inconsistent evidence.

Subjects' treatment of evidence inconsistent with their hypotheses has been investigated in many empirical studies (e.g., Gorman, 1986; Tweney et al., 1980). One conclusion that has been drawn from this literature is that confronting inconsistent evidence forces subjects to make use of whatever inductive reasoning strategies they have at their disposal, because it is in this situation that old hypotheses are most likely to appear inadequate and new ones to be proposed. Chinn and Brewer (1992, 1993) have listed seven ways subjects may deal with inconsistent data. Based on a survey of psychological and historical literature, Chinn and Brewer concluded that subjects may (1) *ignore* inconsistent data, giving no indication of noticing it, (2) *reject* the data, giving an explanation for why they are doing so, (3) *exclude* the data, asserting that it is not relevant to the present theory, (4) hold an anomaly in *abeyance*, indicating that eventually their theory will account for it, (5) *reinterpret* the data to fit with their theory, (6) engage in *peripheral change* of their theory, modifying one part of their theory, or (7) engage in *theory change,* rejecting their current theory. This listing indicates the richness of this situation for observing subjects' inductive strategies. The general conclusion from the experimental literature reviewed in this section has been that subjects tend to ignore inconsistent data. However, not all subjects ignored inconsistent data and subjects' performance differed from task to task. The next section describes some strategies that influence use of inconsistent results.

1.1.2. Heuristics and Constraints. Much empirical work on scientific reasoning has investigated the strategies people use to induce theory in uncertain environments and the constraints that operate on their induction processes. The belief motivating such research is that particular features of a situation may affect whether subjects are likely to pay attention to inconsistent data. Some features that have been suggested are the presence of potential error in experimental feedback (e.g., Gorman, 1986) and whether the subject believes deeply in the current hypothesis (e.g., Chinn & Brewer, 1992).

Gorman (1986) investigated the effect of telling subjects there might be error in the feedback they received on a rule-discovery task. Although erroneous feedback was never in fact given, subjects instructed to use a "disconfirmation" strategy were significantly less likely to discover the correct rule when they believed there might be error in their data. This was true, Gorman reported, in part because subjects tended to attribute feedback inconsistent with their hypotheses to error and in part because they spent time double-checking potential errors. Gorman's finding demonstrates that the potential for error in data reporting, which is a feature of most real-world science, can have a significant impact on reasoning about data.

Chinn and Brewer (1992; Brewer & Chinn, 1991) have used another approach to investigating reasoners' strategies for dealing with inconsistent data. They performed several experiments in which subjects' knowledge about phenomena was manipulated, with the goal of determining whether prior knowledge or beliefs would have an effect on willingness to reject or modify theories in the face of inconsistent evidence. Brewer and Chinn (1991) presented undergraduates with texts that contradicted their previously held entrenched beliefs about special relativity and quantum mechanics. Brewer and Chinn reported that subjects understood the new theories but did not believe them, and refused to abandon their prior entrenched beliefs when confronted with inconsistent evidence. Chinn and Brewer (1992) directly manipulated entrenchment by giving some subjects large amounts of evidence supporting the meteor impact theory of dinosaur extinction Subjects in the entrenched condition were less likely to abandon the meteor impact theory than subjects in the non-entrenched condition, when both groups were presented with evidence that appeared to undermine the theory. As might be expected, strong belief in a theory affected whether a subject was willing to abandon it in a given situation.

1.1.3. Conclusion The empirical work discussed in this section leads to some general conclusions about inductive reasoning. Hypothesis testing research suggests that reasoners have definite strategies for examining data instances, and points to the issue of reasoners' reaction to inconsistent or surprising results as fundamental to illuminating these reasoning strategies. More fine-tuned empirical studies highlight specific strategies that may influence reasoners' use of such inconsistent data. In general, subjects were found to ignore inconsistent evidence, particularly when they held prior entrenched beliefs or when they thought their data might be in error. Empirical work has described general reasoning behaviors in inductive tasks. More theoretical work can give a framework for understanding these behaviors, and these theories will be explored in the next section.

1.2. Theories and Models of Induction

In addition to empirical work, a great amount of effort has been applied to developing full-scale cognitive theories of scientific reasoning, theories that to date share the approach of treating scientific discovery as an inductive task.[2] One of the two major theories

---

[2]Cognitive scientists from disciplines other than psychology, notably philosophy, have developed theories of scientific reasoning.

models scientific reasoning as a special case of problem-solving behavior (e.g., Qin & Simon, 1990) and the other describes heuristics and constraints that guide induction (Holland et al., 1986). In both theories, induction is viewed as complex behavior involving search for explanations of phenomena. The problem-solving model focuses on search through a space of solutions. In particular, it is argued that reasoners search through spaces of hypotheses and data (e.g., Klahr & Dunbar, 1988). In Holland et al.'s description of induction, reasoners search for successful theories, with search guided and constrained by various heuristics and limitations. Both types of models seek to describe strategies that reasoners use to overcome the potentially insurmountable difficulties involved in reasoning about large amounts of data.

1.2.1. Scientific Reasoning as Problem-Solving. Herbert Simon and his colleagues (Kulkarni & Simon, 1988, 1990; Langley et al., 1987, Newell & Simon, 1972; Qin & Simon, 1990; Simon, 1989; Simon & Lea, 1974) have described scientific reasoning as a form of problem-solving behavior. Problem solvers are viewed as beginning from an initial

---

However, in this thesis psychology-oriented cognitive science theories will be concentrated on. For a philosophy-oriented cognitive science approach to science, see Thagard (1988).

state, which includes a representation of the problem. The problem solver uses heuristics to search through a space of problem states, applying operators to transform the initial state into intermediate states and eventually, if successful, into the goal state. This search process can be used to model induction; reasoners use heuristics to guide the search for successful solutions to a problem given various inputs. These heuristics include, for example, means-end analysis, in which reasoners recursively set the goal of minimizing the distance between the current state and a goal state and apply operators to cover that distance (cf. Newell & Simon, 1972).

Simon and his colleagues have designed computational models to simulate problem-solving behavior in scientific environments. The first of these programs were BACON and other computational models derived from it (Langley et al., 1987). These models are data-driven: beginning with a set of observations, the programs attempt to formulate a theory to explain the data. Langley et al. argued that such data-driven discovery has played a major role in historic scientific discoveries. For example, Kepler derived his Third Law—that the cube of the distance of a planet from the sun is directly proportional to the square of the period of revolution—with no pre-existing theory about the relationship. Instead, he worked directly from data collected by Brahe. BACON was successful in rediscovering Kepler's Third Law from data similar to the data Kepler had available. BACON accomplished this by employing the heuristic of searching for a relation among variables that was invariant over the set of data. If two variables vary inversely (directly), BACON then tests their product (ratio, respectively) for invariance. In the case of Kepler's Third Law, BACON notes a series of inverse and direct variances among the distances (D) and periods of revolution (P) of various planets, and products and ratios of these variables. These observations eventually lead the program to create the new variable D^3/P^2, which is constant over the entire set of data.

More recent computational models have simulated aspects of scientific discovery other than data-driven induction of hypotheses (see the volume edited by Shrager & Langley, 1990, for several examples). Kulkarni and Simon's (1988, 1990) KEKADA program, like BACON and its successors, is based on Newell and Simon's theory of problem solving. However, KEKADA not only generalizes from data to theory but also chooses problems, creates hypotheses, designs and chooses experiments to run, and sets expectations for the experiments' outcomes. KEKADA was designed to simulate the process by which the scientist Hans Krebs discovered the urea (Krebs) cycle in 1932, as recorded in his laboratory notebooks and described by Holmes (1980). At the time Krebs made the discovery, he was attempting to understand the process by which urea is synthesized in the human body. He believed initially that ammonia was involved in the reaction, and also that it was likely that one or more amino acids was involved While running different series of experiments, Krebs and his associate Henseleit observed a rate of urea production using a combination of ammonium chloride and the amino acid ornithine that was higher than they had expected based on their experience with similar compounds. Krebs immediately turned his attention to this unexpected effect and eventually concluded that ornithine initiates the chemical reaction that produces urea According to Kulkarni and Simon, a crucial heuristic used by Krebs was to focus on surprising results. They built this heuristic into their model, KEKADA, which attends to and exploits surprising findings, that is, findings that violate the expectations of the program for each experiment. This heuristic is similar to the tactic of attending to inconsistent data, which was identified as a crucial issue in the experimental work described earlier in this chapter. In the KEKADA simulation of Krebs's discovery, the model must choose among different lines of inquiry, and it is the model's decision to focus on the surprisingly (relative to its expectations) high production of urea in the presence of ornithine that leads to the discovery of the urea cycle. The heuristic of focusing attention on surprises also figured in KEKADA's later success in replicating Krebs's discovery of glutamine synthesis (Kulkarni & Simon, 1990), thus demonstrating that Krebs may have used this heuristic in a variety of situations.

Qin and Simon (1990) conducted empirical work to investigate whether subjects solve scientific problems using the same approaches subjects use to solve other kinds of problems. Qin and Simon presented subjects with the same kind of data Kepler used to derive his Third Law, and that BACON used in simulating that discovery. This data, along with a calculator and knowledge about basic mathematical functions, constituted the initial state of the problem The goal as presented to subjects was to discover a formula relating the two sets of data. Qin and Simon found that subjects used problem-solving heuristics that were essentially identical to those described in Newell and Simon's (1972) account of problem solving, which supports the argument that scientific reasoning is similar to everyday problem solving. As in ordinary problem solving, subjects' behavior was characterized by a search process. Qin and Simon described subjects' behavior as a search through the space of possible rules (solution formulas) and also the space of data instances. Among the four subjects (of fourteen) who succeeded in rediscovering Kepler's Third Law, this search was conducted in a systematic way. Hypothesized rules were tested on instances, and information about how rules failed to predict data was used in forming new rules. This description of search in spaces of rules and instances echoed that originally proposed by Simon and Lea (1974).

Klahr and Dunbar (1988; see also Dunbar & Klahr, 1989; Fay, Klahr, & Dunbar, 1990; Klahr, Dunbar, & Fay, 1990; Klahr et al , 1993) further articulated the concept of scientific discovery as search in two spaces, which they referred to as spaces of hypotheses and experiments. In the Klahr and Dunbar studies, subjects worked with a robot named "Big Trak," which subjects were able to control by writing short programs directing the robot's movement and "firing" actions.

Subjects were asked to discover the function of a button on the robot labeled "RPT." Klahr and Dunbar described subjects' performance as a search in both a space of experiments (performed by writing and executing programs) and a space of hypotheses about what the key might do. Some subjects (whom Klahr and Dunbar labeled "experimenters") searched mainly in the experiment space, often trying sample programs without any hypothesis in mind and using the data obtained to derive hypotheses; others (the "theorists") used prior knowledge to propose hypotheses and used data only to constrain the hypotheses they could propose. The "experiment space" in the Klahr and Dunbar model is analogous to the "instance space" in Qin and Simon's (1990) work, and the "hypothesis space" is analogous to the "rule space." Subjects in Klahr and Dunbar's studies guided their search for a solution (a successful hypothesis) by using information from one space to guide search in the other. For instance, a successful search of the experiment space involves choosing experiments that are interpretable with respect to possible hypotheses.

Various manipulations of the Big Trak experiment highlighted different aspects of subjects' search strategies, some of which echo themes of the empirical work described in the previous section. Klahr et al. (1993) found that if subjects perceived a suggested hypothesis as plausible, they tended to use a positive test strategy to test that hypothesis. Thus, subjects used a positive test strategy with regard to hypotheses they believed were likely to be true, but not with hypotheses they doubted. Level of belief in a hypothesis also affected subjects' willingness to discard it when confronted with inconsistent evidence. Consistent with the findings of Chinn and Brewer (1992, Brewer & Chinn, 1991), subjects in both the Klahr et al. (1993) and Klahr et al. (1990) studies abandoned hypotheses given to them more readily than hypotheses they generated themselves, most likely because they had less investment in theories they did not derive themselves. When interpreting the results of experiments, subjects in both the Klahr et al. (1993) and Klahr et al (1990) studies displayed a tendency to focus on one dimension of a hypothesis at a time, with the result that they sometimes missed important features of an experimental result that related to other dimensions. In spite of this, many subjects in both the Klahr and Dunbar (1988) and Klahr et al. (1990) studies were able to make use of surprising findings. These subjects responded to surprising findings by shifting to a goal of explaining the surprising result, which was often followed by a shift to different types of experiments and generation of new hypotheses.

The description of scientific reasoning as problem solving provides a framework within which to investigate scientific discovery processes Work in this framework has shown that historical examples of scientific reasoning can be modeled quite successfully and has highlighted the importance of reasoners' having strategies to constrain their search for hypotheses and for data to induce over The theory described in the next section makes this issue of constraints more explicit.

1.2.2. Constraining the Inductive Process. Following the nineteenth-century philosopher Peirce, Holland et al. (1986; see also Holyoak & Nisbett, 1988) have argued that the "central puzzle of induction" is how induction is constrained. That is, given an infinite number of hypotheses that can be formed about the world, how does the reasoner decide which instances are worth attending to and which theories about them are most plausible?

Holland et al. (1986) identified a number of heuristics and constraints that reasoners use to maximize the efficiency of induction. The first set of aids to induction are heuristics: strategies that might be considered guides in the generation of theories about the world. People use, for example, various reasoning schemata or templates. A reasoning schema regarding causality might be that it is important to determine both necessary and sufficient conditions for an effect to occur. In addition to strategies, which guide search through instances, Holland et al. also identified cognitive limitations on the kinds of inferences people can draw. Among these are temporal and statistical impediments to drawing inferences. It is unlikely that reasoners will perceive events as related if they occur far apart in time, or if one occurs with the other only on some occasions and not others.

Some behaviors may function as both heuristics and limitations. Holland at al. (1986) identify a cognitive limitation of beginning the process of drawing inferences when there are certain "triggering conditions," such as events that are unexpected or problematic. This reflects the argument from the empirical work reviewed above that when reasoners encounter inconsistent results they are required to use their inductive reasoning skills to make sense of the findings. The "triggering condition" of an unexpected event is also the basis of a heuristic identified by Holland et al. as the "unusualness" heuristic. Holland et al. stated that when two unusual or unpredicted events occur in close proximity to each other, the reasoner is likely to conclude that they are related to each other. Not only is induction brought into play by unexpected findings, but also the findings themselves become the basis for induction. The behavior observed by Holland et al. of attending to unexpected events and then reasoning about them to derive new hypotheses is similar to the heuristic Kulkarni and Simon (1988, 1990) identified in Krebs's work of "focusing on surprising results." This behavior offers a rich example of how reasoners might constrain induction in complex environments.

Thagard (1989) offered other criteria for constraining induction. He argued that the goal in drawing inferences is to maximize the explanatory coherence of a theory. Thagard modeled a theory as a system of propositions, which cohere to the extent that they explain one another or form analog relations among themselves. Propositions that make up the theory gain independent acceptability if they describe the results of observation. When reasoners choose experiments and interpret findings in an attempt to maximize explanatory coherence, as Thagard has defined it, their inductive task is greatly constrained.

Thagard designed the program ECHO to model his theory of induction. One consequence of that theory, which is evident in ECHO, is that induction does not have to rely on the assumption that all data is valid. Hypotheses are judged on a relative basis, so a hypothesis is preferred if it explains more than its competitors, but it is possible that no one theory will explain all the available evidence. In such cases, some data may be discounted if it is only explained by theories that fail on other grounds and is not explained by theories that are successful on other grounds. This model provides a possible explanation of why reasoners might ignore inconsistent evidence in some situations. Another consequence of Thagard's theory is that hypotheses are judged relative to other hypotheses rather than being accepted or rejected independently. Judgment researchers Einhorn and Hogarth (1986) reached a similar conclusion while describing hypothesis evaluation as a "discounting" process where hypotheses are accepted or rejected in comparison with other possible causal explanations. Einhorn and Hogarth argued that causal judgments are updated in the face of alternative explanations, so evidence inconsistent with a hypothesis may not lead to its rejection if there are no more viable candidate hypotheses. In these accounts, induction is perceived as a search for the best obtainable theory, not necessarily an ideal theory.

Dunbar (1993) addressed another possible constraint on induction: the formulation of subgoals. Dunbar designed a computer-simulated molecular genetics laboratory in which subjects were given all the necessary tools to rediscover an important genetic control mechanism originally discovered by Jacob and Monod. In the first manipulation, subjects were shown a specific mechanism and were led to believe that it was important for genetic control. In the course of designing and conducting simulated genetic experiments in the computer, the subjects were confronted with evidence inconsistent with their originally hypothesized mechanism. All subjects noticed and used inconsistent evidence to search for instances that would support their initial hypothesis, but only a third were successful in using inconsistent evidence to hypothesize an alternate, correct mechanism of genetic regulation. All subjects who maintained their original hypothesis distorted inconsistent evidence to fit that hypothesis. In a second manipulation, Dunbar altered the molecular environment so that subjects could obtain evidence that confirmed their initial hypotheses before potentially going on to account for continuing discrepancies in the data. In this experiment, three-quarters of the subjects hypothesized the correct alternate mechanism, significantly more than the one-third who reached this hypothesis in the first study.

Dunbar (1993) argued that subjects in the first manipulation were blocked from using inconsistent evidence to generate new hypotheses because of the goal they had set of finding evidence for their initial hypotheses. This is similar to the finding in Klahr et al. (1990) and Klahr et al. (1993) that subjects focused on one dimension of data at a time, and sometimes failed to notice other aspects of a result that might contradict their hypotheses. Subjects in the second Dunbar (1993) manipulation were given the opportunity to reach their initial goal and were then observed to go on to take inconsistent evidence into account. Thus Dunbar's work suggests that setting subgoals is one heuristic reasoners may use to constrain the inductive search. This study also provides another example where an induction heuristic was identified by observing reasoners' responses to obtaining inconsistent evidence. Dunbar's research helps explain other researchers' findings in which people ignored inconsistent evidence. While these researchers in some cases did not know why subjects ignored inconsistent data, according to Dunbar it is because of the goals subjects set. These goals determine how reasoners interpret evidence: not always in relationship to their overall hypotheses but rather in relationship to goals they have set in the course of investigating those hypotheses.

1 2.3. Conclusion. The theoretical work discussed in this section has provided frameworks for investigating scientific reasoning and has highlighted some general points about subjects' approaches to the inductive problem of formulating viable theories with only experimental data and some prior knowledge to draw on. While some discoveries have been made, and no doubt continue to be made, using theory-free, data-driven induction, many other insights can only be derived by using hypotheses and experiments to guide search through large spaces of potential theories and data. Scientists may use techniques such as focusing on surprising findings to guide induction, and strategies like this one may conflict with other heuristics, such as being unwilling to relinquish deeply held prior beliefs or setting subgoals while carrying out an inductive search. It is clear that in dealing with problems of any degree of complexity, reasoners use a variety of strategies to guide inductive search.

In comparing the conclusions of the experimental literature (discussed in the first part of this chapter) with those of the cognitive science literature, it is important to realize that there are conceptual differences in the two bodies of work concerning what kind of evidence is considered to be *inconsistent* or *unexpected.* This distinction will be a fundamental one throughout this thesis. Unexpected findings are not necessarily inconsistent findings. For example, a scientist might hypothesize that a molecule will cause cell growth, and based on past experience might expect growth on the order of 20 to 30%. If the finding is that there was 150% growth, the result would be unexpected and yet still consistent with current hypotheses (the moelcule did cause cell growth). The two main bodies of literature reviewed in this chapter place different emphases on the inconsistency and unexpectedness of experimental findings. The experimental psychology literature focuses on how subjects respond to data that is inconsistent with their current hypotheses. Researchers are not interested in how the subjects respond whenever they are surprised; they are interested only in how the subjects respond whenever they confront data that is not consistent with their hypotheses. On the other hand, cognitive science researchers discussed in this chapter are interested in how reasoners respond to the "unexpected" (Holland et al., 1986) or "surprising" (Kulkarni and Simon, 1990). No specific reference is made to whether surprising or unexpected occurrences violate the reasoner's or scientist's current beliefs. In this thesis, analyses about scientists' response to data will categorize data as unexpected or inconsistent where appropriate.

1.3. The Paradox Predictions from the Two Literatures Conflict

Induction in reasoning has been studied from multiple perspectives, including hypothesis testing and problem solving. From both of these perspectives, researchers have identified heuristics and constraints on reasoning as critical to explaining people's performance in reasoning in information-rich environments. Despite agreement on the importance of constraints, however, there has been some disagreement about the form these constraints take. For example, an intriguing paradox has emerged concerning how people respond to unexpected or anomalous instances, or data. While the experimental literature led to the conclusion that subjects ignore inconsistent findings, more theoretical cognitive science work argues that reasoners focus on the unexpected This disagreement is important because responding to unexpected and inconsistent data poses a major challenge for induction. This section will recapitulate the evidence on both sides of the controversy and will discuss the kind of evidence that might be required to resolve the paradox.

Experimental psychologists have generally concluded that subjects ignore data inconsistent with their hypotheses. An initial impetus for this conclusion came from the work of Mynatt et al. (1978). In that study, subjects working with a complex version of the "simulated universe" task discounted or disregarded inconsistent evidence, instead maintaining their current hypotheses. Later experiments by Gorman (1986) and Brewer and Chinn (1991) introduced parameters that seemed to mimic aspects of real-world science, namely the presence of potential error in the data and entrenched belief in a current hypothesis, respectively. In both cases, subjects tended to disregard evidence inconsistent with their hypotheses. Therefore, the experimental literature seems to make a compelling case that human reasoners ignore inconsistent evidence, especially in circumstances most like those in real-world science.

In contrast, theoretical work by cognitive scientists makes a strong case for the importance in inductive reasoning of focusing on unexpected findings. Holland et al. (1986) argued that paying attention to unusual or unexpected events is a common reasoning heuristic. Kulkarni and Simon (1990) argued that it was crucial for the scientist

Krebs to focus on unexpected results in preference to all other findings, and they built this heuristic into their model of scientific reasoning.

The paradox is this: if, according to cognitive science research, reasoners focus on unexpected findings, what can be made of the fact that experimental subjects seem to do just the opposite when confronted with inconsistent evidence? One might be led to the conclusion that, at least on science-like problems, most humans are poor reasoners. If actual scientists' performance were similar to that of experimental subjects, it would be hard to explain how scientific advances occur at all.

However, the issue may be more complicated than this. The differering conclusions may stem from factors related to the types of experiments that have led to the conclusion that subjects ignore unexpected data. One indication that seemingly minor variations in the experimental situation can have profound effects on reasoning strategies is given by Dunbar's (1993) experiments using a simulated molecular biology laboratory. In an initial experiment, most subjects ignored inconsistent data as they struggled to achieve their initial experimental goal. However, in a subsequent manipulation, where subjects were allowed to achieve their initial goal before confronting inconsistent data, most subjects paid attention to and made use of inconsistent findings. Conflicting findings like these highlight the difficulty of designing experimental situations that accurately reflect what goes on in real-world science.

Dunbar (in press) presents several arguments regarding the danger of generalizing from experimental studies of scientific reasoning to the way real-world scientists reason. First, contemporary science takes place in a social setting, whereas most cognitive work has focused on individuals. Second, psychologists have almost exclusively used tasks that are not "real" scientific problems. Third, the subjects studied are generally non-scientists. Fourth, subjects in experimental studies are asked to work on problems that require a short amount of time to solve and require no extensive knowledge of the scientific topic.

While experimental psychology has focused attention on important issues about reasoning in science-like settings, it is not clear that it is a good model for what goes on in real-world science. Because the ultimate goal of experiments is to shed light on what happens in the real world, one way of providing a balance for experimental work is to conduct research on scientific reasoning in real-world laboratories. Following the terminology of Dunbar (in press), this type of real-world research may be termed *in vivo*,[3] which contrasts it against psychologists' experimental studies of subjects' reasoning strategies, such as those described earlier in this paper, which are referred to as *in vitro*.

---

[3]In biology, *in vitro* research is done outside of the living organism, for example in test tubes or with tissue slices. *In vivo* research is done with the living organism, for instance by injecting something.

The following chapters will describe *in vitro* research in which scientific reasoning is being investigated in real-world laboratories. Specific predictions regarding use of unexpected findings will be made, based on the literature reviewed in this chapter. The methodology used to analyze the real-world data will be described, and findings will then be presented and discussed.

2. Predictions

The data and the methodology used in the current research will be described in detail in the next chapter. Briefly, four laboratory meetings of a prominent immunology lab were tape recorded, transcribed, and coded along such dimensions as when reasoning occurred, when group members interacted, and when new hypotheses were proposed. Using this coded data, it has been possible to investigate questions about how scientists respond to unexpected findings.

The data analyzed in most of this thesis was categorized by whether findings were unexpected, as opposed to inconsistent. As was argued in the last chapter, these two types of findings are not necessarily the same. It may be that some unexpected findings are inconsistent and some unexpected findings are consistent with the researcher's current hypotheses. This discrepancy occurs because predictions for an experiment's outcome are based not always on what a hypothesis would predict but in some cases on other empirical or theoretical knowledge the scientist brings to bear. For clarity in the reporting of results, only one of these categories could be used to analyze data. Unexpected findings rather than inconsistent findings were analyzed for the following reasons: There is a direct link between expectations scientists set for an experiment and determining whether a finding meets those expectations. The predictions and results are reported at the same level and with the same language (e.g., "This antibody will prevent binding of cells to tissue," "This antibody prevented only some binding"), so it is straightforward to determine whether a finding matched a prediction; that is, whether the finding was expected or unexpected. In contrast, the determination of whether a finding was inconsistent with a hypothesis requires multiple steps. First, it must be determined what prediction a hypothesis (e.g., "This molecule mediates homing of these cells to this organ") would support for a particular finding, and only then can the actual finding (e.g., "This antibody prevented some binding") be compared with this prediction. Because the determination of whether a finding was unexpected required one straightforward step, and the determination of whether the finding was inconsistent required two steps, it was decided to analyze most data in terms of response to unexpected, rather than response to inconsistent, findings.

Literature reviewed in the previous chapter can be used to make predictions about how real-world scientists will respond to unexpected findings; that is, findings that are different from the predicted finding The questions and predictions to be addressed in this thesis are as follows:

2.1 Do Scientists Obtain Unexpected Findings?

Klayman and Ha (1987) have argued convincingly that the base rates of expected and unexpected findings obtained will affect experimentation strategies. Therefore, it is of theoretical as well as practical interest to determine how many unexpected findings researchers received. The literature reviewed in this paper does not provide much guidance as to what frequencies of unexpected findings would be expected in the current study.

In the experiments described in Chapter 1 that employed science-like tasks, whether subjects obtained inconsistent findings was to some extent a function of how the experimenter designed the task. For instance, Dunbar (1993) and Brewer and Chinn (1991) deliberately designed their tasks so subjects would be forced to confront inconsistent data. Therefore, it is difficult to predict from experimental data whether real-world scientists will obtain inconsistent findings frequently, occasionally, or almost never.

The theoretical work by cognitive scientists also does not make definite predictions about this question. However, this work does seem to include underlying assumptions about the relative frequencies of obtaining expected and unexpected data. Holland et al. (1986) refer to unexpected events as "triggering conditions" for induction to take place. This kind of reference seems to indicate that triggering conditions occur against a background of normal, expected occurrences and thus occur with regularity but are relatively infrequent. Kulkarni and Simon's (1988, 1991) KEKADA model had heuristics built in requiring the model to prioritize the following up of surprising findings over all other goals. This method of prioritizing seems to assume that surprising findings would be fairly uncommon but important when they did occur.

Because the literature does not allow for firm predictions about how frequently scientists will encounter unexpected findings, it will be important to determine how often this does occur in real-world data.

2.2 Do Scientists Focus on Unexpected Findings?

After determining whether scientists in the sample obtain unexpected findings, the next step is to ask whether they pay attention to these kinds of findings. Response to expected findings is used as a benchmark against which to judge level of response to unexpected findings. It is possible that scientists could pay less attention, similar amounts of attention, or more attention to unexpected findings than to expected findings. The specific variables used to measure whether scientists "pay attention to" unexpected findings will be described in Chapter 4.

From the literature reviewed in Chapter 1, it is possible to make definite, if conflicting, predictions about what scientists will do when they confront unexpected data. As was discussed in Section 1.4, experimental psychology literature (e.g., Brewer and Chinn, 1991; Gorman, 1986; Mynatt et al., 1978) has led to the general conclusion that subjects ignore inconsistent findings, preferring instead to maintain their current hypotheses. By contrast, theoretical cognitive science work (e.g., Holland et al., 1986; Kulkarni and Simon, 1988,

1990) has emphasized the importance of focusing on unexpected findings. Even without disregarding the distinction between inconsistent and unexpected findings, the predictions the two bodies of research make are quite different, so a major goal of the current research will be to answer the question of whether real-world scientists do in fact pay attention to unexpected findings or whether they ignore them.

The amount of reasoning and interacting about unexpected findings will be compared to the amount for expected findings. This analysis will be done separately for the presenter and for other group members. If scientists reason and interact little about unexpected findings, it will be concluded that the prediction from the experimental psychology literature is correct and that scientists ignore unexpected findings. If scientists reason and interact equally about unexpected and expected findings, it will be concluded that scientists do not ignore unexpected findings, but they also do not focus preferentially on them (that is, not more than they do on expected findings). If the scientists reason and interact significantly more about unexpected findings than expected findings, it will be conditionally concluded that the cognitive science prediction is correct and that scientists focus on unexpected more than expected findings.

The "conditional" conclusion in the last sentence reflects the fact that even large amounts of reasoning and interacting about an unexpected finding may not indicate that the scientists are truly interested in the implications of that finding. As Chinn and Brewer (1992, 1993) argue, scientists may reason about unexpected data but only to reject it, exclude it, or reinterpret it as consistent with their theories. These responses may be grouped together as attempts to "explain away" the finding rather than trying to make sense of it theoretically. If the scientists' only purpose in reasoning about unexpected findings was to find grounds for explaining them away, it would not be reasonable to conclude that they had truly focused on these findings. Therefore the goal of reasoning will be evaluated and reported before any conclusion is drawn about whether scientists focus on unexpected findings.

2.3 Do Scientists Make Use of Unexpected Findings?

In addition to asking whether scientists pay attention to unexpected findings, it will also be possible to ask whether they actually use these findings to advance their work. It is possible, for instance, that scientists might spend a lot of time talking about unexpected findings but only to dismiss their importance. It would be more interesting to know whether the scientists do in fact make use of unexpected findings to change their representations of biological phenomena.

For this question, predictions about how scientists will make use of expected and unexpected findings are difficult to make. The experimental literature is often not explicit in distinguishing between paying attention to and making use of inconsistent findings. In part this is because hypotheses and instances are often identical in this work (cf. Klahr & Dunbar, 1988), so noticing an unexpected finding should be tantamount to rejecting the current hypothesis. In experiments where the relationship between data and hypothesis is not so clear, subjects are often reported to make use of inconsistent findings to change their hypotheses if they did in fact notice the inconsistent data (e.g., Dunbar, 1993; Mynatt et al., 1977). The cognitive science work would lead to the prediction that reasoners will not only pay attention to unexpected events, but will also use them to change their beliefs about how the world works (Holland et al., 1986). The modeling of scientific research embodied in KEKADA would particularly support the prediction that successful scientists will reorient their entire research programs when they confront unexpected findings (Kulkarni and Simon, 1988, 1990). Therefore, a review of the literature would lead to the prediction that if scientists pay attention to unexpected findings, they will then make use of them.

The primary measure of "making use of" an unexpected finding that will be used in this thesis is whether scientists propose new hypotheses in response to the finding. If scientists generate new hypotheses in response to unexpected findings, it will be concluded that they made use of unexpected findings. The patterns of new hypotheses generated in response to expected and unexpected findings will also be compared, again separately for presenter and other group members. However, the meaning of a comparison between expected and unexpected findings is not immediately evident. It is not clear whether scientists should generate new hypotheses in response to expected findings. If the findings obtained are expected, it may indicate that the scientists' current hypotheses are correct, so new hypotheses are not required. The overall number of new hypotheses generated in response to unexpected findings is important because it indicates to what extent the scientists made use of these findings. The comparison of number of new hypotheses about unexpected compared to expected findings is less important because the interpretation of such a result is unclear. However, statistics about the comparison between expected and unexpected findings will be reported for completeness.

3. Method

In order to test the preceding ideas, it was necessary first of all to instantiate terms like "make use of" in particular events that could be identified when they occur during a meeting. For instance, the term "make use of" was instantiated as new hypotheses relating to a finding. Then these events (e.g., new hypotheses) had to be recorded each time they occurred; that is, the file had to be coded along these dimensions. Finally, it was necessary to count the number of times certain events occurred in conjunction with other events (for instance, the number of times a new hypothesis occurred in response to an unexpected finding). This chapter describes the data analyzed for this thesis and the method by which it was coded and analyzed.

3.1. Data

The current research makes use of data collected in world-class biology laboratories in 1991–92 (see Dunbar, in press, for a more detailed description). For this thesis, data from laboratory meetings of an immunology laboratory were analyzed. This laboratory, located at a major U.S. university, is run by a senior researcher who is a leader in his field. The laboratory included 22 postdoctoral fellows, 5 graduate students, and 4 technicians. For the purpose of maintaining confidentiality, the names of researchers and certain identifying information about their work were disguised in this report. However, despite the requirement that researchers on this project disguise in publications all identifying information about the labs, in gathering data Dunbar was given free access to all proceedings in the laboratory.

The following types of data were collected in this laboratory: tape recordings of laboratory meetings, tape recordings of interviews with individual researchers, grant proposals, and drafts of papers. Four laboratory meetings have been analyzed for this thesis. Two different postdoctoral fellows presented their work at these meetings, with each presenting at two of the meetings. One of the presenters was chosen for this thesis because an important discovery occurred during the second of the two meetings at which he presented. The other presenter was selected to provide a balance; no major theoretical advances occurred at either of the meetings analyzed where he was the presenter. While the tape-recorded lab meetings form the data set for this thesis, information from the other sources listed was also used to clarify the content of statements made during the lab meetings.

Laboratory meetings lasted from one to two hours and consisted of the presenter informing the lab about his current work, including describing the methodology of recent experiments and reporting experimental findings. In each of the meetings analyzed, the principal investigator of the lab was present, along with other lab members, who asked questions of the presenter and discussed his findings with him. The general role of lab meetings was to keep lab members informed about what each was doing and to provide feedback to the lab member who was presenting. Because of this "feedback" function of lab meetings, it is not uncommon for lab members to suggest new interpretations of data during them. In other words, some scientific theorizing that occurs in this laboratory occurred during lab meetings. Other theorizing probably occurred during one-on-one interactions with other lab members or when the researcher was working alone.

3.2. Overview of Procedure

In this section, a general overview of the procedure used to code the lab meetings is provided. In the next section, the coding scheme utilized is described in more detail.

Each of the laboratory meetings was transcribed from audiotape to typewritten form by a transcriber familiar with the names and corresponding voices of members of the lab and also familiar with terminology used in this lab. After being transcribed, each transcript was subdivided into verb phrases. Figure 1 depicts a portion of a transcribed file that has been separated by verb phrase.

Insert Figure 1 about here

After each transcribed file was separated by verb phrase, it was imported into a database program called MacSHAPA (Sanderson, 1993), as depicted in Figure 2.

Insert Figure 2 about here

Within the Basic coding file, verb phrases were coded along a variety of dimensions, which will be described in detail in the next chapter. The dimensions pictured in Figure 2 are Reasoning and Experimental Classification.[4] Each group of Text cells that comprises one coding on a particular variable is coded just once for that variable. For example, in Figure 2 the Text cells 478 and 479 were coded as Similarity by one Reasoning cell (cell 30). One Reasoning cell represents a Reasoning "block" that occurred in the Text cells; that is, one Reasoning coding typically encompassed more than one Text cell.

The final stage in the coding process was the creation of a Macro file, which stores information about multiple coded meetings. In the Macro file, new hypotheses proposed at each meeting were coded (Figure 3), and total counts of various events that were coded in the Basic databases are recorded (Figure 4).

---

[4] Names of coding variables will be capitalized in the Method chapter (Chapter 3) and in the parts of Chapter 4 where results are reported. Elsewhere, for ease of exposition, they will not be capitalized.

Insert Figures 3 and 4 about here

3.3. Coding Scheme

In addition to the variables in each file for recording Text of the meeting and Who was speaking, each meeting was coded over many other dimensions. The general coding scheme used for analyzing this data is very detailed and makes it possible to chart the step-by-step changes in reasoning and representations that occur. The codings are currently being used to build models of the changes in representation that occur in a meeting. This section describes the particular variables that are utilized in this thesis, which fall into five broad areas.

3.3.1. Interactions. The first major coding variable was Interactions, which was coded in the Basic coding file. This variable was used to code interpersonal exchanges during the meetings. Whenever someone other than the presenter said something, it was recorded as an Interaction. The Interaction block extended from the beginning of the group member's statement to the end of the presenter's response, if any. Portions of the meeting when the presenter was giving his pre-planned talk were not coded as being part of any Interaction.

3.3.2. Reasoning Types: Induction, Deduction, Similarity. The second major class of codings was Reasoning variables, of which three were analyzed for this thesis: Induction, Deduction, and Similarity. These were coded in the Basic coding file and were defined as follows:

(A) Induction. The speaker is trying to *abstract some common element* out of several examples of something.

(B) Deduction. The conclusion follows *logically* from the premise. There is no piece of evidence you can give that would invalidate the conclusion, as long as the premise is true. Often characterized by *if, then* statements (e.g., "If it's true that any antibody can block this binding, then this one should do it.").

(C) Similarity. Any instance where the speaker refers to this phenomenon as being "similar to" or "just like" something else, or "not like" or dissimilar to something else. Any comparison. There must be some explicit comparison/contrast drawn: X is "bigger than," "smaller than," "more important than," "clearer than" Y, or "both X and Y have property Z."

Induction and Deduction were treated as mutually exclusive codings, but either could co-occur with Similarity.

3.3.3. The Reasoning Goals. Each meeting was coded for the Goal of Interactions and Reasoning. The three Reasoning Goal categories were the following:

(A) Theory Build. This included statements that had the general goal of making theoretical sense out of findings; that is, using the information provided by experimental findings to propose biological Mechanisms that might be present.

(B) Explain Away. This included statements in which the speaker attempted to discount findings, for instance by arguing that the assumptions underlying an experimental manipulation were incorrect or that the manipulation was carried out incorrectly.

(C) Neither. These statements did not fit into either of the other two categories.

3.3.4. Experiments and Outcomes. The coding variable Classify (also referred to as Experimental Classification) was used to trace all references to various lines of experimentation. Use of this variable allowed all references to a specific manipulation to be followed from the initial description of a manipulation through reporting of the findings from the experiment and finally through initial reactions to and interpretation of the findings and group discussion of the meaning of the findings. This variable was coded in the Basic coding file and employed an alphanumeric code to classify all experiments referred to by the presenter in some level of detail. The classification was hierarchical, with the general line of experimentation being the highest level, followed by the particular experiment, the specific condition within the experiment, and so on. At the most basic level, a code describes an experimental condition. In turn, groups of conditions together make up an experiment. That is, within any given experiment there may be one or more control conditions and one or more experimental conditions.

All verb phrases were coded as referring to one or more Experimental Classifications, whenever an experiment was being referred to.

The Outcome of each experimental condition was coded in one of the following three categories:

(A) Expected. The results pertaining to this experimental classification were as the presenter predicted.

(B) Unexpected. The results pertaining to this experimental classification were different than the presenter predicted.

(C) Indeterminate. The condition was exploratory; that is, the presenter made no prediction about the outcome of this experimental condition, or the presenter's expectation could not be determined.

The presenter's prediction for an experimental finding was determined from pre-talk interviews, from statements during the presentation, and in a few cases from retrospective accounts of the experiments performed.

3.3.5. Mechanism and New Hypotheses. The coding variable Mechanism recorded the scientists' representation of the *in vitro* research. Mechanism was coded in the Basic coding file. The following categories were coded:

(A) Actor. The specific object that the scientists posit as being involved in an event.

(B) Action. The particular action that an actor performs.

(C) Recipient. The object that the actor performs the action upon.

For example, if the proposed Mechanism was "X cells bind to brain tissue," Actor would be coded as "X cells," Action would be coded "bind to," and Recipient would be coded "brain tissue."

Any or all of these slots were coded whenever a speaker was referring to an underlying biological mechanism. The codings utilized similar language to that used by the researcher.

The Mechanisms postulated by researchers constituted their current representation of how a biological mechanism functions, and changes in how classification slots were coded could indicate changes in the speakers' representation of the biological mechanism. When such a change in representation occurred, it was coded in the Macro file as a New Hypothesis. A change in Mechanism was coded as a New Hypothesis if it met all of the following criteria:

(A) Two of the slots remain substantively constant while another one changes substantively. (The use of the word "substantive" in the preceding sentence reflects the fact that minor changes in vocabulary do not always reflect a change in meaning.)

(B) The hypothesis has not been advanced earlier in the meeting or in an interview prior to the meeting.

(C) The hypothesis is not, to the best of the author's knowledge, the commonly accepted belief about this mechanism in the field.

New Hypotheses were advanced by both the presenter and by others present at the meeting.

3.4. Method of Obtaining Statistical Information from Coding of Data

After the meeting transcripts were coded, data was compiled for statistical analysis. The number of Interaction blocks, Reasoning blocks, and New Hypotheses referring to each Experimental Classification was counted. These counts were performed separately for Reasoning blocks and New Hypotheses offered by the presenter and by other group members. Therefore, each Experimental Classification had a total of five numbers associated with it (presenter New Hypotheses, group New Hypotheses, presenter Reasoning blocks, group Reasoning blocks, and Interactions). These five numbers for each Classification were stored in the Macro file (see Figure 4). The Experimental Classifications were categorized as to whether the Outcome of the Corresponding condition was Expected, Unexpected, or Indeterminate. Statistical tests (described in the next section) were performed using the Expected and Unexpected Classifications as categorizing variables and the five numbers associated with them as dependent measures.

For later analyses, the total number of Reasoning blocks about all Expected and all Unexpected findings was categorized by Reasoning Goal (Theory Build, Explain Away, Neutral).

3.5. Statistical Methods Utilized

The numerical data obtained using the procedure described in the last section made it possible to address the questions outlined in the previous chapter. The data were analyzed using standard nonparametric statistical techniques. Most of the analyses were performed using the Mann-Whitney rank test (with the U statistic), which is a test for nonparametric data that is analogous to a t-test for parametric data. The non-parametric test was used because the data obtained did not follow a normal distribution. Data in Tables 1 through 4 were analyzed using a contingency table test (with the Chi Square statistic).

In the Mann-Whitney tests, the condition in an experiment was used as the unit of analysis. The dependent measure was the number of times particular events (e.g., Interactions) occurred in relation to that condition. Because some Interactions (or Reasoning blocks, etc.) referred to more than one condition, the total number of Interactions entered into the analysis was greater than the actual total number of interactions in the four meetings (that is, some Interactions were double- or even triple-counted over more than one experimental condition). In addition, it is likely that the number of items such as Interactions for one experimental condition might be correlated (positively or negatively) with the number for another condition. For example, Interacting about one finding might be positively correlated with Interacting about a related finding and negatively correlated with Interacting about some unrelated finding. For these reasons, the data analyzed for this thesis do not fulfill the assumption of independence of observations underlying all standard statistical tests. However, the results of the analyses are still interpretable as long as this reservation is kept in mind, and the rank test is the best technique available for analyzing this type of data.

The significance level for all experimental analyses was set at $p < .05$. The p value used for Mann-Whitney analyses was adjusted for ties.

4. Results and Discussion

Using the methodology described in the previous chapter, it was possible to answer the questions posed in Chapter 2 using standard statistical techniques. Results are reported as to how frequently scientists obtained unexpected findings and how they responded when they obtained them. The first three sections of this chapter mirror those of the Predictions chapter; later sections address issues that arose out of the analysis.

4.1. Do Scientists Obtain Unexpected Findings?

The major question addressed in this thesis is how scientists respond to unexpected findings. Before addressing the scientists' responses to such findings it is important to determine whether and to what extent scientists obtain unexpected findings.

As described in the previous chapter, all experimental conditions reported on at the four meetings were classified by means of an alphanumeric code, and groups of conditions together made up experiments. In these four meetings, a total of 16 experiments were reported on, consisting of 70 experimental conditions (for a mean of 4.4 conditions in each experiment).

The result for each of the experimental conditions that the scientists used was classified as having either Expected, Unexpected, or Indeterminate Outcome. There were 22 Expected Outcomes, 27 Unexpected Outcomes, and 21 Indeterminate Outcomes. Only data on

Expected and Unexpected findings will be reported in this thesis, because the focus of the experimental questions is on response to Unexpected findings and the contrast with response to Expected findings.

These findings clearly demonstrate that obtaining unexpected outcomes is a normal part of scientific life in this laboratory. Twenty-seven conditions led to in unexpected findings, resulting in two-thirds of the experiments containing at least one condition with an unexpected result. Scientists are constantly faced with the issue of how to respond to unexpected findings.

4.2. Do Scientists Focus on Unexpected Findings?

In this section, the question of whether scientists focus on or ignore unexpected findings will be addressed. Two measures were used to determine how much attention scientists pay to expected and unexpected findings. The measures used were number of reasoning blocks and number of interactions about a finding. A reasoning block is a set of contiguous sentences in which a person reasons about a finding. The more reasoning blocks that refer to a finding, the more attention is paid to the finding. The number of times group members addressed a comment to the speaker about a finding was also used as a measure of attention paid to that finding. As argued in the Predictions chapter, the amounts of reasoning blocks and interactions can be compared for expected and unexpected findings to determine whether scientists pay more attention to one type of finding than the other. Results are first given for the presenter, with the measure being how much reasoning the presenter did about different types of findings. Next, the findings for group members are reported, including both reasoning by group members and number of interactions that group members initiate about a finding. Finally, data about reasoning goals is reported.

4.2.1. Presenter reasoning. The first set of findings concerns reasoning by the presenter. Experimental conditions to which the presenter referred in one or more Reasoning blocks were counted. The experimental conditions were categorized by Outcome. There were a total of 43 presenter Reasoning blocks about Expected findings and 164 about Unexpected findings. There were significantly more Reasoning blocks about Unexpected than about Expected findings, $U(1, N = 49) = 175$, $p < .05$. The mean rank for Expected findings was 19.46, while the mean rank for Unexpected findings was 29.52. Thus, according to this measure presenters not only pay attention to Unexpected findings, they pay more attention to Unexpected than Expected findings

4.2.2. Spontaneous and prompted presenter reasoning. The previous result shows that presenters Reason more about Unexpected than about Expected findings. However, it does not prove that presenters would, *of their own accord*, reason more about Unexpected findings. Another possibility is that group members direct the attention of the presenter toward Unexpected findings that the presenter might otherwise ignore. If that were the case, the preferential attention paid to Unexpected findings should be attributed to the group, not to the presenter.

To assess whether presenter Reasoning about Unexpected findings was spontaneously offered or was only given in response to prompting from the group, the following analysis was performed. Those Reasoning blocks in each macro file that occurred within an Interaction block were counted separately from Reasoning blocks that did not occur within an Interaction. That is, those Reasoning blocks in which the presenter was responding to a question or statement from somebody in the group by Reasoning were coded as Prompted Reasoning. Presenter Reasoning blocks that were not part of an Interaction were coded as Spontaneous Reasoning. In general, Spontaneous presenter Reasoning occurred during the pre-planned portion of the presenter's talk.

The numbers of Reasoning blocks offered Spontaneously by the presenter were 34 for Expected findings and 87 for Unexpected findings. There were significantly more Spontaneous presenter Reasoning blocks about Unexpected findings (mean rank 30.00) than Expected (mean rank 18.86) findings, $U(1, N = 49) = 162$, $p < .05$.

Turning to the data on Prompted Reasoning by the presenter, the presenter was prompted by group members to offer 9 Reasoning blocks about Expected findings and 77 about Unexpected findings. Although there is a tendency for the presenter to reason more about Unexpected findings in his Prompted statements, a Mann-Whitney test on this data failed to reach significance.[5]

Based on this analysis of spontaneous and prompted presenter reasoning, it can be concluded that presenters do not need to be prompted to reason about unexpected findings; they voluntarily reason about these findings. Furthermore, not only do presenters not ignore unexpected findings, they spontaneously reason more about them than about expected findings.

4.2.3. Group interactions and reasoning. In this section, data on whether group members ignored or focused on Unexpected findings is reported. Group response to a finding can be measured in two ways: by

[5]This negative result may at first seem surprising, because there were many more total Prompted Reasoning blocks about Unexpected than about Expected findings. However, the distribution of Reasoning blocks over findings was such that many findings in both categories had zero or very few Prompted Reasoning blocks associated with them. A few Unexpected findings received a great deal of Prompted Reasoning; however, a rank analysis does not take into account absolute size of the dependent measure on different findings, only relative rank ordering. Therefore, using the Mann-Whitney rank analysis, no significant difference in mean rank between Unexpected and Expected findings was found.

the number of Interactions that pertain to the finding and by the number of Reasoning blocks offered by a member of the group (not the presenter) that pertain to a finding.

There were 23 Interactions about Expected findings and 176 about Unexpected findings. There were significantly more Interactions about Unexpected (mean rank 28.33) than about Expected (mean rank 20.91) findings, $U(1, N = 49) = 207$, $p < .05$.

There were 4 group Reasoning blocks about Expected findings and 91 about Unexpected findings. Although the trend is toward more group Reasoning about Unexpected findings, the Mann-Whitney test failed to reach significance.[6]

Group members clearly did not ignore unexpected findings, but only on the measure of number of interactions did they focus on unexpected findings.

4.2.4. Goal of Reasoning. Before the findings presented thus far are discussed, the results of one final analysis will be reported. As described in the previous chapter, Reasoning blocks were coded as to the Goal of the Reasoning in one of the three following categories:

---

[6]As with the Prompted Reasoning measure, the distribution of Group Reasoning (with many findings in both categories receiving little or no Reasoning) was such that the Mann-Whitney test found no significant difference in Group Reasoning between the two categories.

Theory Build, Explain Away, and Neither. Table 1 shows the number of total (group and presenter) Reasoning blocks devoted to each Goal.

Insert Table 1 about here

The data in Table 1 indicate that the scientists at these meetings spent very little time trying to Explain Away experimental findings, be they Expected or Unexpected findings. Instead, the majority of their Reasoning used findings to Build Theory. To test whether the relative proportion of Explain Away Reasoning was the same for Expected and Unexpected findings, a 2 x 2 contingency table analysis was performed. The number of Reasoning blocks with an Explain Away Goal was compared to the number of Reasoning blocks with a Theory Build Goal.[7] There was a significant result, $c^2(1, N = 221) = 4.10$, $p < .05$, indicating that the scientists used proportionately fewer Reasoning blocks Explaining Away Unexpected as opposed to Expected findings.

4.2.5. Discussion The results reported in this section strongly support the conclusion that scientists *do not* ignore unexpected findings. In no case were unexpected findings reasoned about or interacted about less frequently than expected findings, by the

---

[7]That is, the contingency table had 33 and 161 in the columns of the first row and 9 and 18 in the second row.

presenter or by group members. Furthermore, most reasoning about unexpected findings was directed toward building theory with reference to the finding, not trying to explain away the finding. Therefore, it cannot be concluded that scientists in this lab ignored unexpected findings.

The next question is whether scientists actually focus on unexpected findings more than on expected findings. In order to conclude that scientists focus preferentially on unexpected findings the results would have to indicate that scientists reason more about unexpected than about expected findings and that the goal of this reasoning was not to explain away unexpected findings. The latter criterion was met; the goal of the vast majority of reasoning was to build theory, not to explain away unexpected findings. However, it is not as clear whether scientists reason more about unexpected than expected findings.

The presenter in these meetings did focus on unexpected findings more than expected findings. The presenter reasoned more about unexpected than about expected findings. With reference specifically to the results about spontaneous presenter reasoning, the clearest measure of the presenter's tendencies, the presenter again reasoned more about unexpected than expected findings.

The results about group members are not as clear. Although, by all measures (prompted presenter reasoning, group reasoning, and interactions), there was a trend toward more attention being paid to unexpected findings, in only one case did this trend reach significance: there were more interactions about unexpected than about expected findings. It can be concluded definitively that group members do not ignore unexpected findings, but there is only tentative support for the conclusion that group members focus preferentially on unexpected findings compared to expected findings.

Interpretation of this result is somewhat problematic because of the social nature of the laboratory meeting. Because presenters had already spontaneously reasoned extensively about unexpected findings (in their prepared talks), it is possible that it was not necessary for the group to focus attention on those findings. If the presenter had ignored unexpected findings, it is possible that the group would have reasoned about more of them.

4.3. Do Scientists Make Use Of Unexpected Findings?

In the previous section, it was shown that the majority of reasoning about unexpected findings in these lab meetings was directed toward theory building. The data presented in this section provide more direct measures of whether scientists use unexpected results to build theory: whether reasoning about unexpected findings was used to build theory and whether the findings result in new hypotheses being formed about how a biological mechanism functions. Again these analyses will be performed separately for presenter and group.

Presenters offered 31 Reasoning blocks about Expected findings that had a Goal of Theory Building, and they offered 114 Reasoning blocks about Unexpected findings with a Theory Build Goal. There were significantly more Theory Build Reasoning blocks about Unexpected (mean rank 28.72) than about Expected (mean rank 20.43) findings, $U(1, N = 49) = 197$, $p < .05$. Presenters formed a total of 0 New Hypotheses about Expected findings and 19 about Unexpected findings. There was a significant effect for Expected versus Unexpected findings, $U(1, N = 49) = 168$, $p < .05$, indicating more New Hypotheses were formed about Unexpected (mean rank 29.89) than about Expected (mean rank 19.00) findings.[8]

Group members offered 2 Reasoning blocks with a Theory Build Goal about Expected findings, and they offered 47 Reasoning blocks about Unexpected findings that had a Theory Build Goal. The Mann-Whitney analysis resulted found no significant difference. Turning to hypotheses offered by group members, there were a total of 0 New Hypotheses offered by group members about Expected findings and 22

---

[8]It may at first seem surprising that although each Expected finding had 0 New Hypotheses offered by the presenter, the mean rank for Expected findings was 19.00 rather than 0. However, this occurred because even findings with 0 New Hypotheses are ranked, so ranks (starting at 1 and increasing) were assigned to all Expected and Unexpected findings with 0 New Hypotheses.

about Unexpected findings. There were significantly more group New Hypotheses about Unexpected (mean rank 27.85) than about Expected (mean rank 21.50) findings, $U(1, N = 49) = 220$, $p < .05$.

A total of 14 out of 27 Unexpected findings resulted in New Hypotheses by presenter, group, or both.

The results reported in this section indicate that new hypotheses are offered about some but not all unexpected findings. Therefore, it may be conjectured that there are heuristics that the scientists use narrow down the range of unexpected findings that they focus on. Presenters offer significantly more theory build reasoning statements about unexpected than about expected findings, and significantly more new hypotheses are offered about unexpected than about expected findings by both presenter and group members. As was argued in the Predictions chapter, it is not clear what should be made of this finding To the extent that an expected finding indicates that a scientist's current hypothesis is correct, an expected finding would not necessarily lead a scientist to theorize or propose new hypotheses. Therefore, comparing the numbers of new hypotheses offered for expected and unexpected findings may not be meaningful.

4.4. Are Certain Unexpected Findings Focused on More?

As reported in the previous section, scientists do make use of a good number of the unexpected findings to propose new hypotheses. However, the scientists by no means propose new hypotheses about all unexpected findings. This section and the following section address the issues of how the pattern of total reasoning, interacting, and theorizing about unexpected findings is distributed over the various findings, and what criteria the scientists use to decide which findings will receive more attention.

There are a total of 27 unexpected findings that were reported at the four laboratory meetings analyzed. The number of Interactions. Reasoning blocks, and New Hypotheses related to each finding is illustrated in Figure 5. Findings were placed in the figure in order from greatest to least overall activity related to the finding.

Insert Figure 5 about here

A regression analysis of the data for unexpected findings yields regression coefficients of .897 between number of Interactions and number of New Hypotheses and .824 between number of Reasoning blocks and number of New Hypotheses.

From Figure 5, it can be seen that the numbers of interactions, reasoning blocks, and new hypotheses associated with a finding are converging measures that designate which unexpected findings were considered important. There was much more attention paid to some findings than others; in particular, there is a core group of four to ten findings about which there was much reasoning and interacting and at least a few new hypotheses offered. This result that some findings receive far more attention, by any measure, than other findings suggests that there must be heuristics the scientists use to decide which findings they will focus on.

4.5. What Distinguishes Findings that Receive More Attention?

The analysis in the previous section established that scientists focused on and made use of some unexpected findings far more than others. This immediately prompts the question of what heuristics the scientists used to decide which findings to focus on. In this section, unexpected findings are categorized in two ways that may have been important to the scientists in deciding which findings to focus on. The first distinction is between unexpected findings that are inconsistent or consistent with the researcher's current hypothesis. The second distinction is between unexpected findings obtained in control and non-control[9] conditions of an experiment.

4.5.1. Inconsistent and consistent unexpected findings It is possible that scientists treated unexpected findings differently depending on whether they were inconsistent with their current hypotheses The

---

[9]Non-control conditions are sometimes referred to as "experimental" conditions in the scientific literature. The term "experimental" will be not be used in this context to avoid confusion with other terminology used in this thesis.

distinction between unexpected findings and inconsistent findings was first addressed in Chapter 1. For purposes of data analysis in this paper, unexpected findings have been the primary focus. In this section, the distinction between inconsistent and consistent will be used. Unexpected findings were coded as to whether they were consistent or inconsistent with the researcher's current hypothesis.

There were 27 findings coded as Unexpected. In order to categorize these as Consistent or Inconsistent, it was necessary that the experimenter stated a hypothesis pertaining to that condition before the experiment was run. These hypotheses were obtained from interviews and meeting transcripts. Of the 27 findings, there were 18 for which the experimenter had a hypothesis, 7 for which there was no hypothesis, and 2 that could not be coded.[10] The 18 Unexpected findings for which there was a hypothesis were categorized as to whether they were Consistent or Inconsistent with the researcher's current hypothesis. Of the 18 Unexpected findings, 8 were Consistent and 10 were Inconsistent with the current hypothesis.

---

[10]Of the 7 Unexpected findings for which there was no hypotheses, 1 was a Non-control and 6 were Control conditions. Of the 18 findings for which the presenter had a hypothesis, 13 were Non-control and 5 were Control conditions. Presenters are more likely to have a hypothesis about Non-control than about Control conditions.

In order to make clear how a finding could be both Unexpected and Consistent, it is useful to distinguish between findings that are Quantitatively different from the predicted finding and those that are Qualitatively different. A Quantitatively different finding has a result that is in the direction expected but is of a different magnitude than expected; that is, the difference is one of degree. A Qualitatively different finding is an entirely different type of finding from that expected; for example, something was present that was not expected to be present. All of the 8 Consistent Unexpected findings were Quantitatively different from the predicted finding. Of the 10 Inconsistent Unexpected findings, 5 were Qualitatively and 5 were Quantitatively different from the predicted finding.

Several unexpected findings were consistent with the presenter's current hypothesis. This result is initially counterintuitive, but it becomes understandable when it is realized that all such results were quantitatively, not qualitatively, different from that predicted. An example of this type of finding is one where the scientist hypothesized that a certain molecule would protect an organism from the effects of radiation. The finding was that the molecule did confer radioprotection, but to a much greater extent than expected. The finding was consistent with the scientist's hypothesis, but unexpected in its magnitude.

In order to determine whether the scientists focused more on Consistent or Inconsistent findings, the Consistent and Inconsistent

Unexpected Findings were categorized as to whether the scientists made use of them (that is, whether there were New Hypotheses generated, based on that finding). The results are shown in Table 2.

Insert Table 2 about here

There were New Hypotheses for 8 of the 10 Inconsistent Unexpected findings and for 4 of 8 Consistent Unexpected findings. Although there is a trend in the direction of researchers' generating more New Hypotheses on the basis of Inconsistent Unexpected findings than Consistent Unexpected findings, a contingency table analysis of Table 2 did not reach significance.

4.5.2. Unexpected findings in control and non-control conditions
The second distinction among Unexpected findings that will be addressed is between those obtained on Control and on Non-control conditions. A Non-control condition is designed to look directly at the phenomenon of interest, whereas a Control condition is one against which the result of the Non-control condition is compared. Of the 27 Unexpected findings, 11 occurred on Control conditions and 16 on Non-control conditions. Findings in Control and Non-control conditions were categorized as to whether they were made use of, and the results are shown in Table 3.

Insert Table 3 about here

There is a trend toward Control conditions being hypothesized about more often, but this trend does not reach significance in a contingency table analysis.

Finally, because scientists might be attending to both the Inconsistent/Consistent distinction and the Control/Non-control condition simultaneously, data on the Control/Non-control and Inconsistent/ Consistent categorizations were combined in Table 4.

Insert Table 4 about here

Unexpected findings were coded under these classifications and also categorized as to whether they were the basis of New Hypotheses. Of the 6 findings that were coded both Inconsistent and Non-control, all 6 were made use of. For findings that were coded either Inconsistent or Non-control, but not both, a total of 6 out of 14 were made use of (middle two columns of Table 4). Of those findings coded neither Inconsistent nor Non-control, 2 out of 7 were made use of. A 2 x 2 contingency table analysis was performed to compare Unexpected findings coded Inconsistent and Non-control against all other

Unexpected findings.[11] There was a significant result, $c^2(1, N = 27) = 7.16$, $p < .05$, indicating that Unexpected findings that occur on a Non-control condition *and* are Inconsistent with the presenter's current hypotheses are more likely to be the source of New Hypotheses than other Unexpected findings.

The analyses presented in this section suggest two possible heuristics scientists may use to decide which findings to focus on. Namely, scientists may focus on unexpected findings that occur on non-control conditions and findings that are inconsistent with their current hypotheses. There was a combined effect that suggests scientists used both heuristics simultaneously. All unexpected findings that met both criteria (inconsistent and occurring in a non-control condition) were theorized about by the scientists, as well as about half of the findings that met one criterion or the other. Few findings that met neither criterion were sources of new hypotheses.

4.6. What is the Role of Exploratory Research?

An issue raised by the current research that has not received attention in the psychological literature is the role of exploratory research in science. Exploratory conditions were those run without any prediction, either theoretical or experience-based, about what the

---

[11]That is, the contingency table had 6 and 8 in the columns of the first row and 0 and 13 in the second row.

outcome would be. This type of condition made up about a third of the experimental conditions reported on in these four meetings. There has been little theoretical attention paid to the role of exploratory research in cognitive research, so virtually nothing is known about what purpose exploratory experimentation might serve. In many cases in these meetings, it appeared that the researcher had identified a class of items that might be involved in some phenomenon and was performing relevant tests to see what that role might be, if any. For instance, one scientist believed that one of a class of molecules must be involved in a binding phenomenon, and ran experiments on several molecules to see if any would bind. Another scientist was at the beginning stages of trying to figure out what effect a molecule had on a population of cells, and he ran several conditions where different properties of the cell population were measured before and after treatment. Exploratory conditions often seemed to be run in the early stages of an investigation, when the researcher was trying to discover what might be the important components of a phenomenon. After the exploratory conditions were run, those with interesting results were followed up on with theorizing and more directed experimentation

In terms of attention paid to the results of exploratory experiments, an initial analysis of the data from these meetings suggests that exploratory findings are reasoned about and made use of more than expected findings but less than unexpected findings. This impression has not been tested with statistical analyses, however. In principle, it might be expected that if a researcher used exploratory conditions, he would be interested in the results and would try to determine what they meant. Since the researchers did not reason about or form theory about all findings from exploratory conditions, it is likely that scientists use heuristics to decide which findings to focus on.

5. General Discussion

The main question addressed in this thesis is whether scientists ignore or focus on unexpected findings. Different literatures led to very different predictions of the answer to this question, and the first section summarizes the results that were obtained in this study. Further questions stem from this original question, involving how scientists decide which unexpected findings to focus on and how they respond to unexpected findings. These questions are also addressed in this general discussion, and the answers highlight the richness of heuristics and strategies real-world scientists bring to bear when dealing with unexpected data.

5.1. Do Scientists Focus on Unexpected Results?

This thesis grew out of a paradox. On the one hand, work by experimental psychologists (e.g., Brewer and Chinn, 1991; Gorman, 1986; Mynatt et al., 1978) had led researchers to the conclusion that subjects ignore unexpected data when working on science-like tasks. On the other hand, cognitive science researchers have argued that reasoners in general (e.g., Holland et al., 1986) and scientists in particular (Kulkarni and Simon, 1990) pay attention to and even focus on unexpected events. These two bodies of work led to contradictory conclusions about how people respond to unexpected data. Furthermore, not only were the conclusions contradictory, they claimed to pertain to a population that no one had yet systematically studied:

real-world reasoners. In this thesis, the responses of a group of real-world reasoners, contemporary immunologists, to unexpected and inconsistent data were systematically recorded and analyzed.

In this study, it was found that scientists making presentations paid more attention to unexpected than to expected findings. They reasoned more about the unexpected findings, especially in reasoning statements that they issued spontaneously (without group prompting). With respect to the group rather than the individual scientist, the analysis revealed that group members clearly did not ignore unexpected findings; there was no case in which group members paid less attention to unexpected than to expected findings. However, there was only one measure (number of interactions) on which group members paid significantly more attention to unexpected than expected findings, so there was only tentative support for the conclusion that group members paid more attention to unexpected than expected findings. The meaning of this finding is somewhat unclear, however: because presenters had already reasoned extensively about unexpected findings it is possible that group members did not feel obliged to focus attention on the unexpected findings.

With respect to the measure of making use of unexpected findings, presenters and group members between them formed new hypotheses about 14 out of 27 unexpected findings, indicating that they made use of half the unexpected findings they obtained. Presenters offered more theory-building reasoning about unexpected findings than about expected findings, and presenters and group members formed significantly more new hypotheses about unexpected than about expected findings. This result may merely reflect the fact that theorizing and forming new hypotheses are not necessarily useful responses to expected findings.

Amounts of reasoning, interactions, and new hypotheses about findings were converging measures that designated which findings the scientists considered most important or interesting. The pattern of attention to unexpected findings was found to be extremely variable, with some findings receiving a great deal of attention and theorizing and others very little. Scientists' heuristics regarding which findings to pay attention to may reflect such issues as whether a finding was on a non-control or control condition and whether the finding was consistent or inconsistent with the current hypothesis.

Experimental psychologists have believed, following Popper (1963), that subjects should attend to data inconsistent with their current hypotheses. From this perspective, the finding in the experimental psychology literature that subjects ignored inconsistent findings during science-like tasks raised the very real question of how scientists would ever make new discoveries if real-world scientists performed as poorly in this area as subjects did. The data reported in this study of real-world scientists demonstrates that scientists did not ignore inconsistent findings. Scientists spent a considerable amount of time reasoning about unexpected findings, especially unexpected inconsistent findings, and they made use of many unexpected findings to form new hypotheses.

Cognitive science researchers have argued that reasoners in general and scientists in particular focus on unexpected events. Scientists in this study did not apply this strategy to all unexpected findings. They reasoned and theorized about some unexpected findings much more than others. This pattern of response raises the question of whether the scientists' reactions were suboptimal or were appropriate given the context in which they were working. In one sense, an ignored unexpected finding may represent a discovery that was not made. However, it may simply not be possible for scientists to respond to every unexpected finding they obtain. Klayman and Ha (1987) argued that whether a positive or negative test strategy is most successful depends on the likelihood that a reasoner's current hypothesis is a subset or superset of the correct hypothesis. Similarly, it is likely that the most successful strategy for dealing with unexpected findings may depend on the frequency with which such findings are obtained. In this sample of four laboratory meetings, scientists obtained more unexpected than expected findings. Twenty-seven unexpected findings over four meetings is a lot of unexpected data to deal with, and it is possible that the optimal strategy for scientists is not to focus on all unexpected findings but rather to focus on some that are most important. This raises the question of what heuristics scientists use to decide which findings to focus on. Several heuristics that scientists use are discussed in the next section.

5.2. Under What Conditions Do Scientists Make Use of an Unexpected Finding?

Kulkarni and Simon (1988, 1990) have argued that the scientist Krebs focused his attention on surprising findings. In this thesis, the question of whether scientists use "focusing on unexpected findings" as a general-purpose heuristic was investigated. It was found that despite the large quantity of unexpected findings obtained, it was very rare that an unexpected finding was ignored completely, and it was also rare that scientists tried to explain away unexpected findings (cf. Chinn and Brewer, 1992, 1993). However, some findings were clearly reasoned about and theorized about more than others. Rather than using "focus on unexpected findings" as a general heuristic, the scientists appeared to use a variety of heuristics to decide *which* unexpected findings to focus on. Several of these heuristics are discussed in this section.

5.2.1. Focus on unexpected findings that are inconsistent As was discussed at the end of the previous chapter, not all unexpected findings are inconsistent with the scientist's current hypotheses. Some unexpected findings may be consistent with the current hypothesis and still be unexpected, for instance in the degree of the effect obtained. Scientists are most likely to pursue the unexpected inconsistent findings, because these conflict directly with their current theory. Attending to and making use of an unexpected *consistent* finding is in some sense "optional" for the scientists. This is because without focusing on the unexpected finding, they could continue with their current research plans; nothing has happened to challenge their current hypotheses. In each such instance of obtaining unexpected consistent findings, the scientists have to decide whether the benefits of following up the unexpected finding are more compelling than the benefits to be gained by staying with the current research plan. Therefore, scientists may make use of some but not all unexpected consistent findings.

5 2 2 Focus on unexpected findings from non-control conditions. The role of control conditions has received very little attention in the psychological literature. Control conditions are used for at least two purposes: (1) Control conditions are used to verify that the experimental methodology is sound. That is, if all the machinery and the growing conditions for cell cultures and so on are set up properly, the scientist expects the experimental condition to give a different result than the control condition. (2) Control conditions are used to indicate the range of validity of the hypothesis. For example, if the hypothesis is that some molecule will cause an effect, then a control condition might replace that molecule with another molecule of the same class If both cause the effect, it may be that all molecules of that class cause the effect

In this study, scientists did not theorize about many unexpected results pertaining to control conditions. Because there is often no hypothesis that pertains directly to a control condition, an unexpected finding on a control condition may not directly refute the validity of the scientist's current theory. In contrast, there are almost always hypotheses that pertain directly to non-control conditions, and unexpected results on these conditions often have immediate implications for current theory. Scientists therefore are likely to pay more attention to unexpected results on non-control conditions. In this study, the combination of this heuristic and the one described in the previous section was particularly powerful: scientists theorized about all unexpected inconsistent findings that occurred in non-control conditions.

It should be noted that not focusing on unexpected findings on control conditions is a strategy that can easily backfire on the scientists, particularly if the control condition was of the first type described above. If the unexpected result on the control condition indicates that the manipulation did not work properly, ignoring that result could lead the scientist to make erroneous conclusions about the non-control condition.

5.2.3. Pay attention to whether unexpected findings are qualitatively or quantitatively different from those predicted Qualitatively different findings are those that differ from the predicted finding in an absolute sense. Quantitatively different findings are those that differ from the predicted finding only in degree. The two presenters analyzed for this study differed in the kinds of unexpected findings they tended to obtain when categorized by this measure. One presenter was investigating the relative changes in immune protection when a certain molecule was introduced. Therefore the kinds of unexpected findings he tended to obtain were quantitatively different from those predicted (e.g., 70% of mice were protected as opposed to the 90% predicted). The other presenter was investigating which molecule mediated binding of immune cells to certain body tissues. The kinds of unexpected results he tended to obtain were qualitatively different from those predicted (e.g., a molecule that was not expected to cause binding did). Both researchers appeared to use a version of Holland et al.'s (1986) "unusualnesss" heuristic: in the few cases when they obtained findings that were different in kind than those they usually obtained (for instance, obtaining a qualitatively different finding when quantitatively different findings were usually obtained), they tended to focus on those findings. Either kind of finding can be important, depending on the researcher's goals. However, findings that are different in kind from those usually obtained appear to act as triggers for theorizing.

5.2.4 Focus on unexpected findings that available theoretical knowledge makes interpretable. As this thesis has established, scientists obtain unexpected findings constantly. In many cases, there is no hypothesis pertaining directly to the finding, or the finding is not inconsistent with current hypotheses. In those cases, the scientist must decide whether to pursue the theoretical implications of the finding. The scientist can only do this in a meaningful way if he has a knowledge base within which the finding makes sense. In this study, there were examples of findings that no one at the meeting seemed to know what to say about—they were unexpected, but no one could explain why they might have occurred. In those cases, the scientists were unable to form hypotheses to explain the finding. In other cases the scientists immediately recognized the theoretical implications of an unexpected finding; for example, the finding disproved a long-standing assumption in the field. In those cases, there was abundant theorizing about what the finding might mean.

5.2.5. Focus on unexpected findings that fit overall research goals
Dunbar (1993) argued that experimental subjects were inhibited from making use of unexpected findings when they had not yet fulfilled their experimental goals. In this study, whether a scientist followed up on an unexpected finding depended in part on what his overall goals were. For example, an unexpected finding that had broad theoretical implications was not aggressively pursued by one researcher because his ultimate goal was more practical, to find a cure for a disease. As discussed below, one way scientists can respond to an unexpected finding is to change their short-term or long-term goals, but if they are reluctant to give up their current goals they may not be motivated to

make theoretical sense of an unexpected finding that has no relevance to their current goals.

5.2.6. Focus on unexpected findings with a potential payoff. There are often practical considerations that make it more or less attractive for a particular scientist at a particular time to follow up on an unexpected finding. For example, if there is a large payoff (in terms of reputation, monetary reward, etc.) for making a discovery in a certain area of research, the scientists may be more likely to pursue an unexpected finding that has implications for that area. Alternatively, the scientist may be less inclined to follow up on a finding if the only methodologies available for doing so are very expensive. The scientist's assessment of whether pursuing a certain finding is likely to result in a discovery will also affect whether that finding is made use of: if the chances of success are high, the scientist will be more likely to pursue it.

5.2 7 Conclusions Cognitive science researchers have argued that some discoveries have been made by focusing on unexpected findings. However, given the rate at which unexpected findings are obtained in real-world laboratories, it is impossible for scientists to focus on all unexpected findings. When more than a third of findings obtained are unexpected, and focusing on each of those would inevitably lead to obtaining more unexpected findings, it is literally impossible for scientists to focus on all of them. The scientists must develop heuristics for deciding which unexpected findings to make use of, even though any of the heuristics they use may backfire by causing the scientist not to focus on some finding that might have turned out to be important. Scientists have no alternative but to focus their energies on the unexpected findings that their heuristics lead them to believe are most important.

5.3. Why Do the Current Results Differ from Those in the Experimental Literature?

The findings reported in this thesis differ in many respects from findings obtained in cognitive psychology experiments testing subjects' use of inconsistent evidence. Unlike experimental subjects, who ignored inconsistent evidence, scientists in this study focused on and formed new hypotheses about inconsistent findings. Furthermore, some factors that were predicted in the experimental literature to inhibit the use of inconsistent evidence did not appear to have that effect in the real-world data. For example, Gorman (1986) found that when potential error was present in data, subjects were likely to attribute inconsistent findings to error and therefore fail to find the correct rule. However, despite the fact that the potential for error is constantly present in real-world laboratories, scientists showed little tendency to explain away inconsistent findings as being due to error

Another factor that was found by Chinn and Brewer (1992; Brewer and Chinn, 1991) to prevent subjects from attending to inconsistent data was the degree of "entrenchment" of the subjects' belief in their current theory. Subjects who had stronger beliefs in their current theories were less likely to question the theories when presented with inconsistent data. It might be expected that real-world scientists would have much more entrenched belief in their current theories than subjects in a two-hour lab experiment. However, scientists in this study seemed quite willing to question their current hypotheses. In part this may be due to the fact that many of their theories were not fully formed and could be easily modified. It was in the case of a fully formed theory that one scientist in this study was most reluctant to abandon the theory in the face of inconsistent evidence. However, scientists were also seen to question fundamental assumptions of their field when data appeared inconsistent with those assumptions. On the whole, the issue of "entrenchment" appeared to be moot in this study; scientists rarely appeared to be strongly committed to any one theory.

In summary, some issues raised in the experimental literature appear to have little applicability to the real-world situation investigated in this thesis. The question arises as to why the experimental literature comes to such different conclusions from those reported here. Dunbar (in press) put forth several arguments about why experimental findings might not be consistent with what goes on in the real world. First, contemporary science takes place in a social setting, whereas most cognitive work has focused on individuals. Second, psychologists have almost exclusively used tasks that are not "real" scientific problems. Third, the subjects studied are generally non-scientists. Fourth, subjects in experimental studies are asked to work on problems that require a short amount of time to solve and require no extensive knowledge of the scientific topic. All of these are extremely important differences between the experimental setting and real-world science. For example, with respect to the use of non-scientists as subjects, Arocha and Patel (in press) have found that more senior medical students were less likely to ignore inconsistent data than less experienced students, suggesting that expertise plays a role in this area. Aside from the differences differences between the real-world and experimental settings noted by Dunbar, additional differences can be pointed out based on issues discussed in this thesis. Two of these are discussed in this section: the greater availability of strategies for dealing with unexpected findings and the discrepancy between what was measured in this study and in the experimental studies.

First, scientists have well-developed strategies for deciding what unexpected findings they will focus on (for example, those listed in the previous section) and how they will respond to unexpected findings Most of these involve actively making use of the unexpected finding, rather than ignoring or disregarding the finding. For example, three of the strategies for responding to unexpected findings listed by Chinn and Brewer (1992, 1993)—ignore, reject, and reinterpret—were rarely employed by scientists in this study Scientists can respond to unexpected findings in any of the following ways:

(1) Use a different methodology. One strategy used by scientists is to maintain the current goal or hypothesis and to conduct a new experiment using a different methodology to try to find some support for the current hypothesis. For example, one scientist in this study who obtained results that were not entirely consistent with his current theory responded by using a different experimental technique to try to obtain evidence consistent with that theory. Only after he again failed to obtain evidence consistent with his current hypothesis using the new methodology did he abandon the theory.

(2) Modify current theory. Scientists often used unexpected findings to limit the range of applicability of or otherwise modify their current hypotheses. Many of the theories being developed by the scientists were not fully articulated; that is, the current hypothesis was vague and open to further specification. For example, one presenter theorized that a certain molecule would protect mice from the harmful effects of radiation. When he later found that the radioprotective effects were greater in one organ of the body than they were in another organ, he modified his theory to account for this. The process of building and articulating a theory takes place over time and in response to many findings, both expected and unexpected. In the experimental literature, subjects typically have only fully articulated hypotheses. When faced with inconsistent evidence, they either have to abandon their theory or ignore the evidence. When the subjects have little knowledge that would allow them to develop alternate

hypotheses, they may feel their only choice is to maintain their current hypothesis.

The strategy of modifying the current theory is somewhat similar to the "peripheral theory change" strategy described by Chinn and Brewer (1992, 1993). However, the idea that scientists will change unimportant or non-central components of a hypothesis in response to unexpected findings presumes that those non-central parts of the theory are well articulated to begin with. In this study, it was much more common that the scientists' theories were not fully articulated, so modifying a theory was more likely to involve adding details than changing details.

(3) Abandon current theory. This strategy is most appropriate when a theory is fully articulated. Because fully articulated theories were uncommon in this data set, it was also uncommon to abandon a theory entirely. However, this did happen twice in the four meetings investigated.

(4) Switch to a new line of research. This strategy is also fairly rare, but when it happens it is of great importance. It is most likely to occur when an unexpected finding does not pertain directly to the current hypothesis but instead suggests that fundamental underlying assumptions have been in error. Initiating research to explain this finding might involve abandoning current research that the scientist has already invested time and energy in, so the switch would only take

place if the scientist felt the new line of research was more promising than the old.

Having a broad range of possible responses available for dealing with unexpected findings makes it possible for the scientists to choose the most promising and appropriate response in the current circumstances, and it also gives them alternate strategies to pursue if one does not seem to be working. Experimental subjects have no strategies for dealing with unexpected findings and no conceptual structure they can use to make sense of unexpected findings, so they ignore them.

A second possible cause of the differences in conclusions drawn from the current study and conclusions drawn from the experimental work is the difference in what was measured. Experimentalists concluded that subjects ignored inconsistent evidence because the subjects did not immediately abandon their current theories. In this study, more fine-grained measures were used to determine whether scientists attended to unexpected and inconsistent findings. Scientists reasoned about virtually all unexpected findings and formed new hypotheses about half of them, so it was concluded that scientists do not ignore unexpected findings. However, if the only response that had qualified as "not ignoring" an unexpected finding was to immediately abandon the current theory, it would have been concluded that scientists in this study "ignored" many unexpected findings, just as subjects were said to "ignore" inconsistent data.

This sort of all-or-none criterion for deciding whether subjects ignore or make use of inconsistent data is inadequate for capturing the rich and varied range of responses scientists can make to an unexpected finding. Scientists can reason about findings and then apply many heuristics to decide whether or not to focus on them. They can offer new hypotheses about experimental findings and use these to modify and build theory, without abandoning their current theory completely. They can use alternate methodologies or switch to entirely new lines of research. Unlike subjects in a lab, who have few heuristics and few strategic alternatives available to them, scientists have many strategies for dealing with unexpected and inconsistent data. Real-world scientists do not have to choose simply to ignore or not ignore unexpected findings, because they have developed a rich array of heuristics and strategies for dealing with data of this kind.

References

Arocha, J. F., & Patel, V. L. (in press). Coordinating theory and evidence in diagnostic reasoning by novices. *Journal of the Learning Sciences.*

Brewer, W. F., & Chinn, C A. (1991). Entrenched beliefs, inconsistent information, and knowledge change. *Proceedings of the 1991 international conference on the learning sciences*, 67–73. Charlottesville, VA: Association for the Advancement of Computing in Education.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: NY Science Editions.

Chinn, C. A., & Brewer, W. F. (1992). Psychological responses to anomalous data. *Proceedings of the fourteenth annual conference of the Cognitive Science Society*, 165–170. New Jersey: Lawrence Erlbaum Associates.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Education Research*, *63*(1), 1–49

Dunbar, K. (1993) Concept discovery in a scientific domain. *Cognitive Science, 17,* 397–434.

Dunbar, K. (in press). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *Insight*. Cambridge, MA: MIT Press.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery strategies. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing The impact of Herbert A Simon*. Hillsdale, NJ: Erlbaum.

Einhorn, H. J. & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 319.

Goodman, N. (1972). Seven strictures on similarity. In Goodman, N. (Ed.), *Problems and projects* New York: Bobbs-Merrill.

Gorman, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology*, *77*, 85–96.

Holland, J., Holyoak, K , Nisbett, R. E., & Thagard, P (1986) *Induction Processes of inference, learning, and discovery.* Cambridge, MA MIT Press.

Holmes, F. L. (1980). Hans Krebs and the discovery of the ornithine cycle. *Federation Proceedings*, *39*, 216–225.

Holyoak, K. J., & Nisbett, R. E. (1988) Induction. In R. J Sternberg & E. E. Smith (Eds), *The psychology of human thought* Cambridge Cambridge University Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–48.

Klahr, D., Dunbar, K., & Fay, A L. (1990). Designing good experiments to test bad hypotheses In J Shrager & P Langley (Eds), *Computational models of discovery and theory formation*. San Mateo, CA: Morgan-Kaufmann.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*, 111–146.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, *12*, 139–175.

Kulkarni, D., & Simon, H. A. (1990). Experimentation in machine discovery. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann Publishers.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: The MIT Press.

Miller, G. A. (1967). *The psychology of communication*. New York: Basic Books.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85–95.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology, 30,* 395–406.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ. Prentice-Hall, Inc.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper & Row.

Qin, Y., & Simon, H. A. (1990). Laboratory replication of scientific discovery processes. *Cognitive Science, 14,* 281–312.

Sanderson, P. M. (1993). Designing for simplicity of inference in observational studies of process control: ESDA and MacSHAPA. *Proceedings of the Fourth European Conference on Cognitive Science Approaches to Process Control (CSAPC '93): Designing for Simplicity.* August 25–27, Copenhagen, Denmark.

Shrager, J., & Langley, P. (Eds.) (1990). *Computational models of scientific discovery and theory formation.* San Mateo, CA: Morgan Kaufmann Publishers.

Simon, H. A. (1989). Science and thought. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A Simon*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction. In H. A. Simon (Ed.), *Models of thought*. New Haven, CT: Yale University Press.

Snyder, M., & White, P. (1981). Testing hypotheses about other people: Strategies of verification and falsification. *Personality and Social Psychology*, 7, 39–43.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.

Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Sciences*, *12*, 435 502.

Tweney, R. D. (1989). A framework for the cognitive psychology of science. In B. Gholson, A. Houts, R. A. Neimayer, & W. Shadish (Eds.), *Psychology of science and metascience*. Cambridge: Cambridge Univ. Press.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, *32*, 109–123.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Wason, P. C. (1968). "On the failure to eliminate hypotheses . . . "—A second look. In P. C. Wason & P. N. Johnson-Laird (eds.), *Thinking and reasoning*. Middlesex, England: Penguin Books.

Tables

Table 1

Categorization of Reasoning Blocks by Goal

	Expected	Unexpected
Theory Build	33	161
Explain Away	9	18
Neither	5	76

Table 2

Consistent and Inconsistent Unexpected Findings Categorized by Whether Finding was the Source of New Hypotheses

	Consistent	Inconsistent
New Hypotheses	4	8
No New Hypoths	4	2

Table 3

Unexpected Findings on Control and Non-control Conditions Categorized by Whether Finding was the Source of New Hypotheses

	Control	Non-control
New Hypotheses	4	10
No New Hypoths	7	6

Table 4

Unexpected Findings Categorized by both Inconsistent and Control Status and Whether Finding was the Source of New Hypotheses

	Inconsistent, Non-control	Inconsistent, Control	Not Incons., Non-control	Not Incons., Control
New Hypotheses	6	2	4	2
No New Hypoths	0	2	6	5

Figures

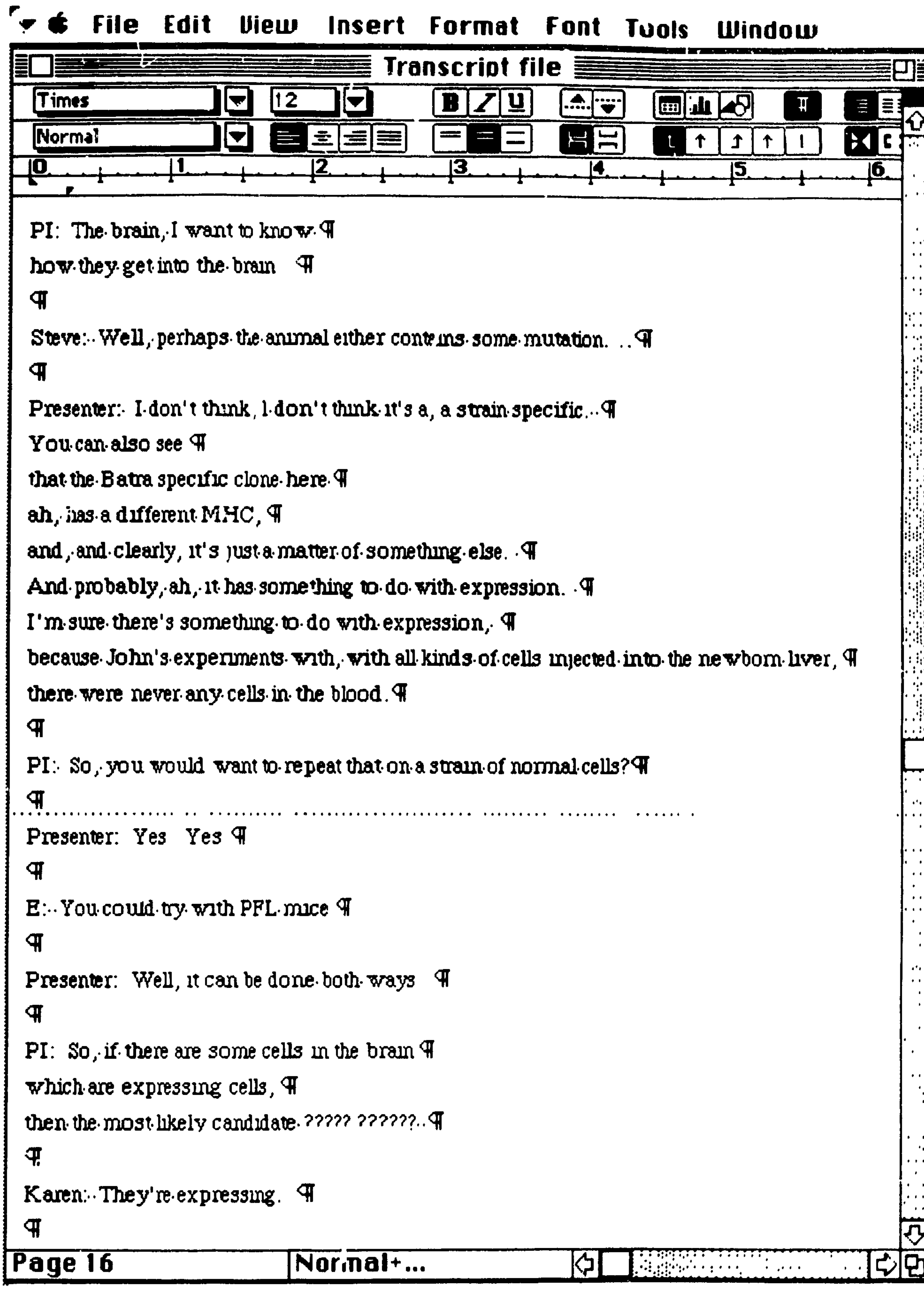
Figure 1. Example of meeting transcript, delineated by verb phrase.

Figure 2. Example of basic coding of meeting transcript depicted in the last figure. Transcript text has been placed in "TEXT" column, with one cell allocated for each verb phrase. Cells are coded for Reasoning types, and Classification and Speaker are coded for each Reasoning block.

Figure 3. Example of macro coding file, which contains information from several meetings. New Hypotheses suggested in each meeting are coded according to the experimental condition(s) they were offered in response to and by hypothesis type.

Figure 4. Another piece of macro file. Numbers of New Hypotheses, Reasoning blocks, and Interactions are counted from earlier codings (Figures 2 and 3) and recorded.

Figure 5. Number of Interactions, Reasoning blocks, and New Hypotheses related to each Unexpected finding. Findings are ordered by decreasing activity.


File Edit View Insert Format Font Tools Window
Transcript file
Times
12
Normal
PI: The brain, I want to know ¶
how they get into the brain ¶
¶
Steve: Well, perhaps the animal either contains some mutation. . ¶
¶
Presenter: I don't think, I don't think it's a, a strain specific. ¶
You can also see ¶
that the Batra specific clone here ¶
ah, has a different MHC, ¶
and, and clearly, it's just a matter of something else. ¶
And probably, ah, it has something to do with expression. ¶
I'm sure there's something to do with expression, ¶
because John's experiments with, with all kinds of cells injected into the newborn liver, ¶
there were never any cells in the blood. ¶
¶
PI: So, you would want to repeat that on a strain of normal cells? ¶
¶
Presenter: Yes Yes ¶
¶
E: You could try with PFL mice ¶
¶
Presenter: Well, it can be done both ways ¶
¶
PI: So, if there are some cells in the brain ¶
which are expressing cells, ¶
then the most likely candidate ????? ??????. ¶
¶
Karen: They're expressing. ¶
¶
Page 16
Normal+...

File Edit Spreadsheet Query Reports Import Windows Sys

Basic file

TEXT	REASONING	CLASSIFY	WHO
474 00.06.54.00 PI: The brain, I want to know			
475 00·06 55:00 how they get into the brain			
476 00 06 56 00 Steve: Well, perhaps the animal either contains some mutation....	28 00:06:56:00 INDUCTION	84 00:06:56.00 (B,2,b,ii,c)	50 00 Steve
477 00·06·57·00 Presenter: I don't think, I don't think it's a, a strain specific	29 00:06:57:00 DEDUCTION	85 00:06:57:00 (B,2,b,ii,c)	51 00 Presenter
478 00·06·59:00 The Batra specific clone here	30 00:06·59:00 SIMILARITY	86 00·06·59·00 (B,2,b,ii,c)	52 00 Presenter
479 00.07 00:00 ah, has a different MHC,			
480 00 07 01 00 and, and clearly, it's just a matter of something else.			
481 00 07 02.00 And probably, ah, of course it has something to do with expression	31 00:07·02:00 INDUCTION	87 00:07·02 00 (B,2,b,ii,c)	53 00 Presenter
482 00 07 03 00 I'm sure there's something to do with expression,			

File Edit Spreadsheet Query Reports Import Windows Sys

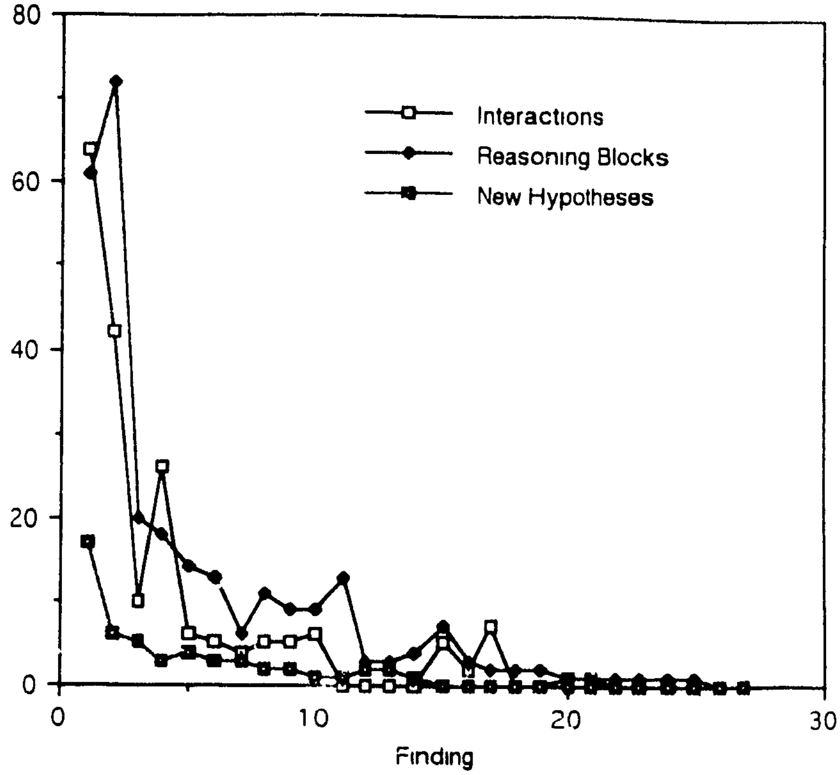
Macro file

CONDITION	NEW.HYPOTHS	TYPE.OF.NH
29 10.00 00·04 (MAR.23,B ,2,b,u,c)	8 10.00·00 04 10 00 00 04 ENTRY 6 53 Production patterns allow entry into brain	8 10 00 00 04 10 0 (PRESENTER,NEW,NEW, POS)
30 10.00 00 06 (MAR.23,B ,2,b,u,c)	9 10 00 00.06 10 00 00·06 (Artifact) ENTRY 6 56 Animal mutation allows cells in	9 10 00 00 06 10 0 (OTHER,ARTIFACT, ARTIFACT,NEG)
31 10 00 00.08 (MAR.23,B ,2,b,u,c)	10 10 00 00.08 10:00 00:08 ENTRY 7 02 Entry is related to expression	10 10 00 00 08 10 0 (PRESENTER,NEW,NEW, POS)
32 10.00.00:10 (MAR.23,B ,2,b,u,c)	11 10 00.00.10 10:00.00:10 ENTRY ADHESION MOLECULE VLA4 7.14 In addition to expression, X2 molecule is key to entry	11 10 00 00 10 10 0 (PI,NEW,DETAIL,POS)
33 10 00 00.12 (MAR.23,B ,2,b,u,c)	12 10 00.00 12 10.00 00.12 ??? ENTRY ADHESION MOLECULE 7 20	12 10 00 00 12 10 0 (PI,NEW,DETAIL,POS)

File Edit Spreadsheet Query Reports Import Windows Sys

Macro file

CONDITION	NUM.NEW.HYP	NUM.REASON	NUM.INTE
51 10.03 01.00 (MAR.23,A,3,b,i,<e>)	23 10 03.01 00 (0,0,0)	23 10 03 01 00 (4,0,4)	23 10 0
52 10 03 02 00 (MAR.23,A,3,a,ii,<e>)	24 10 03.02.00 (0,0,0)	24 10.03 02 00 (1,0,1)	24 10 0
53 10 03 03 00 (MAR 23,A,3,b,ii,<e>)	25 10 03 03:00 (0,0,0)	25 10 03 03 00 (2,0,2)	25 10 0
54 10·04·00:00 (MAR.23,B,2,a,ii,a)	26 10:04·00:00 (0,0,0)	26 10:04·00·00 (7,0,7)	26 10 5
55 10.04·01·00 (MAR.23,B,2,a,ii,b)	27 10:04·01:00 (0,0,0)	27 10:04·01·00 (1,0,1)	27 10 0
56 10 04 02·00 (MAR.23,B,2,a,ii,c)	28 10 04 02.00 (0,0,0)	28 10 04 02 00 (2,2,4)	28 10 2
57 10 04 03 00 (MAR.23,B,2,b,ii,a)	29 10 04 03 00 (0,0,0)	29 10 04 03 00 (5,0,5)	29 10 3
58 10 04 04·00 (MAR.23,B,2,b,ii,b)	30 10 04.04 00 (0,0,0)	30 10.04 04 00 (2,0,2)	30 10 0
59 10 04 05 00 (MAR.23,B,2,b,ii,c)	31 10 04 05 00 (6,11,17)	31 10 04 05 00 (23,38,61)	31 10 64


80
60
40
20
0
0
10
20
30
Finding
Interactions
Reasoning Blocks
New Hypotheses