Human Reasoning and Cognitive Science

Keith Stenning and Michiel van Lambalgen

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in latex by the authors. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Stenning, Keith.
Human reasoning and cognitive science / Keith Stenning and Michiel van Lambalgen.
p. cm. – (A Bradford book)
Includes bibliographical references and index.
ISBN 978-0-262-19583-6 (hardcover : alk. paper) 1. Cognitive science. 2. Reasoning. 3. Logic. I. Lambalgen, Michiel van, 1954- II. Title.
BF311.S67773 2008
153.4–dc22

2007046686

10 9 8 7 6 5 4 3 2 1

1 Introduction: Logic and Psychology

The purpose of this book is twofold. Our first aim is to see to what extent the psychology of reasoning and logic (more generally, semantics) are relevant to each other. After all, the psychology of reasoning and logic are in a sense about the same subject, even though in the past century a rift has opened up between them. Very superficially speaking, logic appears to be normative, whereas the psychology of reasoning is descriptive and concerned with processing. The first question then is: what is the relation between these two fields of inquiry?

The psychology of reasoning as a field currently adopts a particular view of this relation: we propose a quite different one. The book therefore should be relevant to students of logic who are interested in applications of logic to cognition. But the book should also be relevant to any psychologist who is interested in reasoning or communication, or any other cognitive capacity where a cognitive account has to be founded on an informational analysis of a cognitive capacity. These two groups come to the topic with very different methodological equipment. The logic student interested in cognition comes with an understanding of the level of abstraction that modern logical theories operate at, but with possibly sparse knowledge of psychological observations of reasoning. Students of psychology come with knowledge of the experimental literatures, but those literatures are strongly formed by a different conception of logic – a conception, current in the nineteenth century,[1] in which logic is thought of as a *mechanism* for reasoning, and a universal, normatively valid mechanism at that.

This presents us with an educational dilemma. The logical analyses presented here are couched at a level which is intended to be comprehensible to nonlogicians with sufficient patience to digest logical formulas.[2] From experience, the problems encountered are not so much problems about the technicalities of the systems (which are often not the main point here) but background assumptions about what logic is about and what such systems do and do not attempt to provide. So our message to the psychology student venturing here would

1. Although already at that time more refined conceptions existed; compare section 1.3 on Husserl.
2. Chapter 2 does duty as an introduction to those aspects of logic most important for our purposes.

be that we promise to show that these analyses can make a real difference to how empirical investigations are designed, so venturing is well worthwhile. But understanding requires that many routine assumptions about logic are left at the door. Modern logical theories provide a conceptual and mathematical framework for analyzing information systems such as people's reasoning and communication. They do not settle mechanisms or processes of reasoning, but without their conceptualization, it is impossible to know what are empirical and what are conceptual questions.

To the student venturing from logic our message is that we obviously cannot provide more than the very bare outlines of a few empirical results, along with some pointers to the literature. So we need to warn against assuming that the empirical phenomena are as simple or separable as they are bound to appear. The immense contribution that psychology has made to understanding the mind largely consists of bodies of empirical knowledge about what phenomena are replicable under what range of conditions. Much of this knowledge is implicit in the literature.

This educational dilemma leads naturally to our second wider aim, to discuss some of the theories offered in the literature from the point of view of the philosophy and methodology of science. For instance, both mental models theory and evolutionary psychology, which take their starting points in observations about the psychology of reasoning, have become hugely popular explanatory paradigms in psychology. We will see that the experiments claimed to support these theories are marred by conceptual confusions and attendant methodological errors, and that the theories themselves show little awareness of the subtleties of logic. Part of our purpose is therefore to propose a different methodology for this field, which takes Marr's idea of "levels of analysis" seriously.

1.1 Forms of Rationality

Traditionally, rationality is taken to be a defining characteristic of human nature: "man is a rational animal," apparently capable of deliberate thought, planning, problemsolving, scientific theorizing and prediction, moral reasoning, and so forth. If we ask what "rational" means here, we can read such things as: "In rational discourse one strives to arrive at justified true belief," a definition of rationality from an era oriented toward theory. Our more pragmatically oriented age has extended this concept of rationality to actions. For instance, in the MIT *Encyclopedia of Cognitive Science*, "rational agency" is defined as a coherence requirement:

> [T]he agent must have a means-end competence to fit its actions or decisions, according to its beliefs or knowledge representations, to its desires or goal-structure.

Without such coherence there is no agent. The onus here is on the term *fit* which seems to have a logical component. If an action is performed which is not part of a plan derived to achieve a given goal, there is no fit. In this sense checking my horoscope before mounting the bike to go to work is not rational, and neither is first puncturing the tires.

Philosophy, then, studies the question: are there optimal rules for conducting such activities? Various logics, scientific methodology, heuristics, probability, decision theory, all have claims to normative status here, where normativity means that everybody should obey the rules of these systems in all circumstances. As a consequence, there exists an absolute distinction between valid arguments and fallacies. Judged by these standards, human reasoning in the laboratory is very poor indeed (as shown by the seminal experiments of Wason [295] for logic and Kahneman and Tversky [150] for probability), and it has therefore been said that humans are actually not rational in the sense defined above.

It is usually assumed that the results obtained in the psychology of reasoning tell us something about the *rationality*, or rather the absence thereof, of human reasoning. The following extended quotation from Peter Wason, one of the founding fathers of the field whose "selection task" will serve as our entrypoint below, exemplifies this attitude to perfection. He writes, concluding an overview of his selection task paradigm for *The Oxford Companion to the Mind*,

> Our basic paradigm has the enormous advantage of being artificial and novel; in these studies we are not interested in everyday thought, but in the kind of thinking which occurs when there is minimal meaning in the things around us. On a much smaller scale, what do our students' remarks remind us of in real life? They are like saying "Of course, the earth is flat," "Of course, we are descended from Adam and Eve," "Of course, space has nothing to do with time." The old ways of seeing things now look like absurd prejudices, but our highly intelligent student volunteers display analogous miniature prejudices when their premature conclusions are challenged by the facts. As Kuhn has shown, old paradigms do not die in the face of a few counterexamples. In the same way, our volunteers do not often accommodate their thought to new observations, even those governed by logical necessity, in a deceptive problem situation. They will frequently deny the facts, or contradict themselves, rather than shift their frame of reference.
>
> Other treatments and interpretations of problem solving could have been cited. For instance, most problems studied by psychologists create a sense of perplexity rather than a specious answer. But the present interpretation, in terms of the development of dogma and its resistance to truth, reveals the interest and excitement generated by research in this area. (Wason [300,p. 644])

What lies behind remarks such as Wason's is the view that reasoning, whether logical or probabilistic, can be judged to be rational if certain reasoning *rules* from a fixed, given set are followed. If these rules are not followed, dire consequences may result. A good example of this attitude is furnished by Stanovich's book *Who is Rational?* [254]. The following quotation gives some idea of the

passions that infuse this approach. Stanovich considers irrationality to lead to the occurrence of

> wars, economic busts, technological accidents, pyramid sales schemes, telemarketing fraud, religious fanaticism, psychic scams, environmental degradation, broken marriages, and savings and loan scandals [254,p. 9]

and believes teaching good reasoning, that is, normatively correct rules, will go some way toward improving this distressing situation.

Stanovich's discussion of rules governing reasoning introduces a distinction between *normative*, *descriptive*, and *prescriptive* rules. We give brief characterizations of the three kinds, followed by representative examples.

- *Normative* rules: reasoning as it should be, ideally
  - Modus tollens: $\neg q, p \rightarrow q / \neg p$,
  - Bayes' theorem: $P(D \mid S) = \frac{P(S|D)P(D)}{P(S)}$.
- *Descriptive* rules: reasoning as it is actually practiced
  - Many people do not endorse modus tollens and believe that from $\neg q, p \rightarrow q$ nothing can be derived.
  - In doing probabilistic calculations of the probability of a disease given a cluster of symptoms, even experts sometimes neglect the "base rate" and put $P(D \mid S) = P(S \mid D)$.
- *Prescriptive*[3] rules: these are norms that result from taking into account our bounded rationality, i.e., computational limitations (due to the computational complexity of classical logic, and the even higher complexity of probability theory) and storage limitations (the impossibility of simultaneously representing all factors relevant to a computation, say, of a plan to achieve a given goal).
  - The classically invalid principle $\neg q, p \wedge r \rightarrow q / \neg p \wedge \neg r$ is correct according to *closed–world reasoning*, which is computationally much less complex than classical propositional logic, and ameliorates storage problems.
  - Chater and Oaksford's "heuristic rules" for solving syllogisms. [36]

In terms of these three kinds of rules, Stanovich then distinguishes the following positions on the relationship between reasoning and rationality [254,pp.4–9]:

- *Panglossian*. Human reasoning competence and performance is actually normatively correct. What appears to be incorrect reasoning can be explained

3. The term is not very apt, but we will stick to Stanovich's terminology.

by such maneuvers as different task construal, a different interpretation of logical terms, etc. (A famous defense of his point of view can be found in Henlé [122].) As a consequence, no education in "critical thinking" is necessary.

- *Apologist*. Actual human performance follows prescriptive rules, but the latter are in general (and necessarily) subnormal, because of the heavy computational demands of normatively correct reasoning. This point of view was defended by Oaksford and Chater [205, 36]. As a consequence, education in "critical thinking" is unlikely to be helpful.

- *Meliorist*. Actual human reasoning falls short of prescriptive standards, which are themselves subnormal; there is therefore much room for improvement by suitable education (Stanovich's own position).

- *Eliminativist*.[4] Reasoning rarely happens in real life, and mainly in institutional contexts such as schools. By contrast, true rationality is adaptiveness: we have developed "fast and frugal algorithms" which allow us to take quick decisions which are optimal given constraints of time and energy. This position is defended by evolutionary psychologists such as Cosmides and Tooby[47, 48] and, in more constructive detail, by Gigerenzer [95].

It will be helpful for the reader if we situate our own position with respect to this scheme. We are definitely not in the eliminativist camp, since we take the view that reasoning is everywhere, most prominently in discourse comprehension. This prime example is often overlooked because of the association of reasoning with conscious processing, but this association is wrong: some reasoning is automatic.[5] The same example leads us to think that human reasoning may not be so flawed after all, since it operates rather competently in this domain. We are not Panglossians either, although we emphasize that interpretation is of paramount importance in reasoning. But even if interpretation is important, and interpretations may differ, people may reason in ways which are inconsistent with their chosen interpretation. From a methodological point of view this means that if one uses a particular interpretation to explain performance, one must have evidence for the interpretation which is independent of the performance. The apologist and meliorist positions introduce the distinction between normative and prescriptive rules. Here it becomes clear that Stanovich's scheme is predicated on the assumption that reasoning is about following rules from a fixed, given set, say classical logic, rules which should apply always and everywhere. For if there is no given set of rules which constitutes the norm, and the norm is instead relative to a "domain," then the domain may well include the

4. Actually, this category does not occur in [254], but we have added it due to its increased prominence.
5. Chapter 5 has more on this.

cognitive constraints that gave rise to the notion of prescriptive rules, thus promoting the latter to the rank of norm. This is what we will argue for in several places in the book, in particular in chapter 11.

In the next section we briefly look at the role logic once played in cognitive science, and the reasons for its demise.

1.2 How Logic and Cognition Got Divorced

The cognitive sciences really got off the ground after they adopted the information-processing metaphor (Craik [52]):

1. Cognitive explanations must refer to models, conceived of as representational mechanisms
2. which function "in the same way" as the phenomena being represented
3. and which are capable of generating behavior and thoughts of various kinds.

The role of logic in this scheme was twofold: on the one hand as a formal, symbolic, representation language (which is very expressive!), on the other hand as an inference mechanism generating behavior and thoughts. It was furthermore believed that these inference mechanisms are continuous with overt reasoning; that is, the same processes can be applied both reflectively and automatically.

An extreme form of this attitude is of course Piaget's "logicism" [216], which maintains that the acquisition of formal-deductive operations is the crown of cognitive development. Piaget did the first studies to show that preschool children do not yet master simple classical predicate logic; but he also assumed that everyone gets there in the end. This proved to be the Achilles heel of this form of logicism. Indeed, Wason's selection task was inter alia directed against this assumption, and its apparent outcome – a striking deviation from classical logical reasoning – seemingly undermined the role of logic as an inference mechanism. A further criticism concerned the alleged slowness of logical inference mechanisms, especially when search is involved, for example when backtracking from a given goal. Thus, Newell and Simon style "production systems"[199], of which Anderson and Lebiere's ACT-R [2], is the most famous example, keep only the inference rule of modus ponens, allowing fast forward processing, at the cost of considerable complications elsewhere.

Lastly, the advent of neural network theory brought to the fore criticisms of the symbolic representational format given by logic: it would be tied to brittle, all-or-none representations, uncharacteristic of actual cognitive representations with their inherently fuzzy boundaries. As a further consequence of this brittleness, learning symbolic representations would be unrealistically hard. As a result, from the position of being absolutely central in the cognitive revolution, which was founded on conceptions of reasoning, computation, and the analysis

of language, the psychology of deduction has gone to being the deadbeat of cognitive psychology, pursued in a ghetto, surrounded by widespread skepticism as to whether human reasoning really happens outside the academy. "Isn't what we *really* do decision?" we increasingly often hear. Many eminent psychology departments do not teach courses on reasoning. Imagine such a psychology department (or indeed any psychology department) not teaching any courses on perception. Even where they do teach reasoning they are more likely to be focused on analogical reasoning, thought of as a kind of reasoning at the opposite end from deduction on some dimension of certainty.

We will argue that logic and reasoning have ended up in this ghetto because of a series of unwarranted assumptions. One of the tasks of this book is to examine these assumptions, and show that they do not bear scrutiny. As a prelude, we consider the vexed issue of the normative status of logic, through some of the history of present day conceptualizations.

1.3 Two Philosophers on the Certainty of Logic: Frege and Husserl

Famously, Aristotle provided the first rules for reasoning with quantifiers of the form "All A are B," "Some A are B," "No A are B" and "Some A are not B," starting centuries of work on how to provide principled explanations for the validity of some syllogisms, and the invalidity of others. This search for an explanation turned to the notion of validity itself (*überhaupt*, we are tempted to say), and Kant opined in the *Critique of Pure Reason* that logical laws constitute the very fabric of thought: thinking which does not proceed according to these laws is not properly thinking.[6]

In the nineteenth century, this "transcendental" doctrine of logic was watered down to a naturalistic version called *psychologism*, which holds that all of thinking and knowledge are psychological phenomena and that *therefore* logical laws are psychological laws. To take an example from John Stuart Mill, the law of noncontradiction $\neg(A \wedge \neg A)$ represents the impossibility of thinking contradictory thoughts at the same time. Thus, normative and descriptive rules coincide. What came after, a strong emphasis on normativity, can to a large extent be seen as a reaction to this view. Gottlob Frege was the driving force behind the reaction, and his views still exert their influence on the textbooks.

6. It is impossible to do justice here to Kant's thinking on logic. It is still common to think of Kant's logic as primitive, and its role in the *Critique of Pure Reason* as an instance of Kant's architectonic mania. This is very far from the truth, and Kant's thinking on, for instance, logical consequence remains relevant to this day. In fact, our concluding chapter 11 has many affinities with Kant, although it would require another book to explain why. Béatrice Longuenesse's *Kant and the Capacity to Judge* [175] is an excellent guide to the wider significance of Kant's logic. The reader may also consult Patricia Kitcher's *Kant's Transcendental Psychology* [159] for an exposition of Kant's relevance to cognitive science. Kitcher's remarks on the similarities between Kant's first *Critique* and Marr's program in cognitive science [183] have influenced chapter 11. We thank Theodora Achourioti for pointing out these connections.

1.3.1 Frege's Idealism in Logic

Frege did not hesitate to point out the weak empirical basis of psychologism: is it really true that we cannot simultaneously think contradictory thoughts? Wish it were so! His chief argument, however, was theoretical, and consisted of two main reservations about a naturalistic treatment of logic (and mathematics):

1. Psychologism makes logic pertain to ideas only, and as a consequence it lacks resources to explain why logic is applicable to the real world.
2. Logical and mathematical knowledge are objective, and this objectivity cannot be safeguarded if logical laws are properties of individual minds.

We now present a few extracts from Frege's writings to illustrate his views on psychologism.

As regards 1 we read:

> Psychological treatments of logic ... lead then necessarily to psychological idealism. Since all knowledge is judgmental, every bridge to the objective is now broken off. (G. Frege, *Nachgelassene Schriften*; see [85])

> Neither logic nor mathematics has the task of investigating minds and the contents of consciousness whose bearer is an individual person. (G. Frege, *Kleine Schriften*; see [85])

> The logicians ... are too much caught up in psychology ... Logic is in no way a part of psychology. The Pythagorean theorem expresses the same thought for all men, while each person has its own representations, feelings and resolutions that are different from those of every other person. Thoughts are not psychic structures, and thinking is not an inner producing and forming, but an apprehension of thoughts which are already objectively given. (G. Frege, letter to Husserl; see 6, p. 113 of [135])

The last sentence is especially interesting: if "thinking is not an inner producing and forming," cognitive science has no business investigating thinking, logical reasoning in particular. What the psychologist finds interesting in reasoning is precisely what steps the mind executes in drawing an inference, i.e., in the process more than the result. The quotation just given suggests that logic itself has little to contribute to this inquiry, and indeed psychologists have generally heeded Frege's message, either by designing logics which are supposedly cognitively more relevant (e.g., Johnson-Laird's 'mental models' [145]. See Chapter 10, especially 10.6.3), or by ignoring the contributions that formal logic can make to theories of processing. This is a pity, since, as will be shown in the body of the book, the technical apparatus of logic has much to offer to the psychologist.

Here is an excerpt relevant to 2:

> If we could grasp nothing but what is in ourselves, then a [genuine] conflict of opinions, [as well as] a reciprocity of understanding, would be impossible, since there would be

> no common ground, and no idea in the psychological sense can be such a ground. There would be no logic that can be appealed to as an arbiter in the conflict of opinions. (G. Frege, *Grundgesetze der Arithmetik*; the relevant part is reprinted in [85])

From the last quotation we gather that it is apparently highly desirable that logic "can be appealed to as an arbiter in the conflict of opinions," and that therefore there must be a single, objectively valid logic. The second quotation (from the letter to Husserl) provides a reason to believe there is one: logic is as it were the physics of the realm of thought, since it studies the structure of the "objectively given" thoughts. Psychologism is a threat to this normative character of logic, and since a logic worthy of the name must give rise to norms for thinking, psychologism is not a possible theory of logical validity. But, as we have seen, Frege must invoke an objectively given realm of thought to buttress the normative pretensions of logic, and this assumption seems hard to justify. However, if one is skeptical about this objective realm of thought, the specter of relativism rises again. At first sight, the normativity of logic seems to be bound up with the uniqueness of logic; and what better way to safeguard that uniqueness than by positing some underlying reality which the logical laws describe?

This is indeed a serious problem, and to solve it requires rethinking the sense in which logic can be considered to be normative. In a nutshell, our answer will be that norms apply to instances of reasoning only after the interpretation of the (logical and nonlogical) expressions in the argument has been fixed, and, furthermore, that there are in general multiple natural options for such interpretations, even for interpreting the logical expressions. Thus, the reasoning process inevitably involves also steps aimed at fixing an interpretation; once this has been achieved, the norms governing logical reasoning are also determined. It will not have escaped the reader's attention that the view of reasoning just outlined[7] is very different from the one implicitly assumed by the standard paradigms in the psychology of reasoning. As this book goes to press, a special issue on logical views of psychologism covering a range of contemporary positions is published [171].

One aim of this book is to present a view of reasoning as consisting of reasoning *to* and reasoning *from* an interpretation, and to apply this view to experimental studies on reasoning. In philosophy, our precursor here is Husserl, who is playing Aristotle (the metaphysician, not the logician) to Frege's Plato. Husserl's views on logic never made the logic textbooks (at least explicitly), but we nonetheless believe that one can find in him the germs of a semantic conception of logic, which comes much closer than Frege's to how logic functions in actual reasoning.

7. The outline is very rough indeed. For instance, the phrase "once this has been achieved" suggests that the two stages are successive. As we will see in the experimental chapters, however, it is much closer to the truth to view these stages as interactive.

1.3.2 Husserl as a Forerunner of Semantics

It was Frege who converted Husserl to antipsychologism. When we read Husserl's criticism of psychologism in *Logische Untersuchungen* [134], we at first seem to be on familiar ground. If logical laws were empirical laws about psychological events, they would have to be approximative and provisional, like all empirical laws. But logical laws are exact and unassailable, hence they cannot be empirical.[8] Psychologism about logical laws also leads to skeptical relativism: it is in principle possible that different people reason according to different logical laws,[9] so that what is true for one person may not be true for another – truth, however, is absolute, not indexed to a person.

So far these arguments are question begging: we may have a strong desire for logical laws to be exact and unassailable and objective, but we need a justification for assuming that they are. In trying to provide one, Husserl develops a strikingly modern view of logic, and one that is much more conducive to playing a role in cognitive science than Frege's. Husserl's *Logische Untersuchungen* brings the important innovation that logic must be viewed, not as a normative, but as a theoretical, or as we would now say, mathematical, discipline. Logic as a theoretical discipline is concerned with "truth," "judgment," and similar concepts. Husserl grounds the normative status of logic via a combination of the theoretical statement "only such and such arguments preserve truth" and the normative statement "truth is good," to conclude: "only such and such arguments are good." Splitting off the normative from the mathematical component of logic is potentially beneficial, since it focuses attention on what exactly justifies the normative statement "truth is good," and thus opens up space for a relativized version such as "(this kind of) truth is good (for that purpose)."

In slightly more detail,[10] Husserl introduces an essentially modern division of logic as concerned with

1. "the pure forms of judgments" (i.e., the syntax of a formal language, but here implying a Kantian delineation of what can be said at all);

2. "the formal categories of meaning" (i.e., the semantic study of concepts such as "variable," "reference," "truth," "proposition," "consequence"[11]);

8. Husserl remarks correctly that even if logical laws are considered empirical, psychologism is under the obligation to explain how we can acquire them, and that no account of how logical laws are learned has been forthcoming.

9. In modern times this is occasionally cheerfully accepted. The logician Dov Gabbay once said in an interview: "Everybody his own logic!"

10. Here we are much indebted to David Bell's *Husserl* [17], in particular pp. 85 –100. The quotes from Husserl are taken from Bell's monograph.

11. It is of some interest to observe here that for Frege, semantics, although intuitively given, was not a proper field of scientific study, since it involves stepping outside the system which is given a semantic interpretation. See also footnote 12. We agree with Husserl that it is both possible and necessary to reflect on semantic interpretation.

3. "the formal categories of objects" – that is, what is known as "formal ontology," the study of such concepts as "object," "state of affairs," "continuum," "moment." This can be read as saying that part of logic must be the characterization of the structures on which the chosen formal language is interpreted; the next quotation calls these structures "possible fields of knowledge":

   > The objective correlate of the concept of a possible theory, determined exclusively in terms of its form, is the concept of a possible field of knowledge over which a theory of this form will preside. [This field] is characterized by the fact that its objects are capable of certain relations that fall under certain basic laws of such-and-such a determinate form . . . the objects remain entirely indeterminate as regards their matter. . . (*[17,p. 90-1]*)

4. Lastly, rational thinking also involves *systematization*, and therefore pure logic must also comprise a study of formal theories, not only of propositions and their inferential relationships; in Husserl's words

   > The earlier level of logic had taken for its theme the pure forms of all significant formations that can occur within a science. Now however, judgment systems in their entirety become the theme (*Formale und transzendentale Logik* [17,p. 90-1]).

Modern logic has followed this last injunction, and studies what is known as "metaproperties" of a logical system such as consistency, the impossibility of deriving a contradiction in the system. Among the most important metaproperties are metatheorems of the form "only such-and-such argument patterns preserve truth," which depend on a preliminary characterization of the notion of truth in the "possible field of knowledge" studied. Normativity comes in only via a principle of the form "in this particular field of knowledge, truth of such-and-such a form is good, therefore only such-and-such arguments are good." This means that logical laws are unassailable in the sense that they are mathematical *consequences* of the structure of the domain studied, but by the same token these laws are relative to that domain. The reader will see in the body of the book that this view of logic, as not providing absolutely valid norms but as giving norms valid relative to a particular domain, sheds new light on results in the psychology of reasoning which have traditionally been taken to show the incompatibility of logic and actual reasoning.[12]

12. The preceding paragraphs are not intended as an exegesis of Husserl's thought; our intention is only to identify some strands in Husserl which we consider to be fruitful for thinking about logic. The contrast drawn here between Frege and Husserl is a particular case of the more general distinction, first proposed by Jean van Heijenoort [281], between "logic as a universal language" and "logic as a calculus." On the former conception of logic, whose main champion is Frege, logic is concerned with a single universe of discourse, and the semantic relation between logical language and that universe is ineffable. On the latter conception (which ultimately gave rise to the modern "model–theoretic logics" [15]), there are many possible universes of discourse, logical languages are reinterpretable to fit these universes, and semantics is a legitimate object of scientific study. Van Heijenoort's contrast has been called "a fundamental opposition in twentieth century philosophy" by Hintikka [125] and has been applied to Frege and Husserl in Kusch [165].

For the mathematically inclined reader we include an example of Husserl's views as applied to the domain of arithmetic. If this domain is given a classical, Platonistic interpretation, that is, as concerned with objects which exist independently of the human mind, then the following is *not* a logical law:

(1) (†) if A *or* B is provable (in system S), then A is provable (in system S) or B is provable (in system S)

because on the one hand Gödel's incompleteness theorem has shown that there is a sentence A such that neither A nor $\neg A$ is provable in classical arithmetic, whereas on the other hand the "law of excluded middle" $A \vee \neg A$ is a logical law in classical arithmetic. If, however, the domain of arithmetic is conceptualized as being about particular mental constructions, as mathematicians of the intuitionistic persuasion claim, then it is a mathematical fact that (†) is a logical law (and that therefore the law of excluded middle is not). Normative issues arise, not at the level of logical laws (e.g., the law of excluded middle), but at the level of what description to choose for the domain of interest. Changing one's logical laws then becomes tantamount to changing the description of the domain.

Husserl's view has the value of focusing attention on the relation between mathematics and empirical phenomena in general, as one source of difficulty in understanding the relation between logic and human reasoning. The relation between mathematics and empirical phenomena is problematical in any domain, but it may be more problematical in this domain than most. Appreciating the continuity of these problems, and identifying their source is one way forward. Seeing logic as the mathematics of information systems, of which people are one kind, is quite a good first approximation to the view we develop here. This view helps in that it makes clear from the start that one's choice is never between "doing psychology" and "doing logic." Understanding reasoning is always going to require doing both, simply because science does not proceed far without mathematical, or at least conceptual apparatus.

So we see history turning circle, though not full circle. Like Mill and Husserl, we see logic and psychology as very closely linked. Frege rejected this view. Husserl developed a much more sophisticated view of the relation, which foreshadows our own in its emphasis on semantics. Later, Frege's view of logic foundered on Russell's paradoxes which showed that logic couldn't be universal and homogeneous. In response logic developed the possibility of explicitly studying semantics, and still later, developed a multiplicity of logics. Much of the technical development necessary for studying semantics took place in the context of the foundations of mathematics which took logic very far from psychology. In mid-twentieth century, Montague reapplied the much transformed technical apparatus of logical semantics to the descriptive analysis of natural languages. We now apply the availability of a multiplicity of logics back onto

the subject matter of discourse and psychology. Of course psychology too has changed out of all recognition since Mill – the whole apparatus of psychological experiment postdates Mill's view. So our "psychologism" is very different from Mill's, but the closeness of psychology and logic is something shared. Our psychologism clearly requires an account of how logic in its modern guise as mathematical system is related to psychology in its modern guise as experimental science.

1.4 What the Reader May Expect

The remainder of the book is structured as follows. Chapter 2 is a somewhat unorthodox introduction to logic, which tries to break the hold of classical logic by showing it results from contingent assumptions on syntax and semantics. Systematic variation of these assumptions gives rise to several logics that have applications in actual human reasoning. This chapter in particular introduces closed–world reasoning, a form of reasoning that will be very important in part II of the book. In chapters 3 and 4 we study the Wason selection task from the vantage point developed in this introduction and in chapter 2: as a task in which subjects are mostly struggling to impose an interpretation on the experimental materials instead of engaging with the materials as the experimenter intends them to do. Chapter 3 contains many examples of tutorial dialogues with subjects which show what interpretational difficulties they experience, and chapter 4 reports on experiments establishing that alleviating these difficulties by modifying the task instructions leads to a vast increase in correct performance, when measured against the classical competence model.

The selection task has played a major role in debates on evolutionary psychology, with Cosmides [47] claiming that her results on facilitation of the task with social contract materials show that the only abilities humans have in logical reasoning are due to an innate module for "cheater detection." We believe that this highly influential point of view is mistaken, for two reasons: a faulty view of logic, and a faulty view of evolution. chapter 5 considers the influence this faulty logical paradigm has had on the psychology of reasoning, and chapter 6 is a lengthy discussion of the evolution of human cognition. The latter chapter introduces a hypothesis that will play an important role in part II: that the origin of the human ability for logical reasoning must be sought in the planning capacity, and that closed–world reasoning (which governs planning) is therefore a very fundamental form of reasoning, stretching across many domains.

Chapter 7 applies this idea to the analysis of the suppression task (Byrne [28]), which is standardly interpreted as providing evidence for "mental models" and against "mental rules." We show that the data from this task can be explained on the assumption that subjects assimilate the task to a discourse–processing task, using closed–world reasoning, by presenting a rigorously for-

mal model of subjects' reasoning in a logical system called "logic programming," which we consider to be the most appealing form of closed–world reasoning. We also present data from tutorial dialogues corroborating our interpretation of what subjects are doing. In chapter 8, it is shown that closed–world reasoning has a revealing neural implementation and that there need not be an opposition between logical and connectionist modeling. Chapter 9 applies the ideas of chapters 7 and 8 to autism. We analyse several tasks on which autistic people are known to fail, such as the false belief task and the box task, and find that these tasks have a common logical structure which is identical to that of the suppression task discussed in chapter 7. This leads to a prediction for autistic people's behavior on the suppression task, which has been verified. This latter result is analysed in terms of the neural implementation developed in chapter 8, which then allows us to make a connection to the genetics of autism. Chapter 10 discusses syllogisms. These are, of course, the first reasoning patterns studied in psychology, but for us their interest lies in the necessity to apply substantial interpretational theories from linguistics and philosophy to explain the data. This explains why syllogisms only occur near the end of the book: the reader must first be familiar with both interpretation processes of reasoning to an interpretation and derivational processes of reasoning from the interpretation imposed.

Lastly, chapter 11 makes explicit our view of logic and its relevance to actual reasoning. The role of logic is to aid in "going beyond the information given" when processing information. Just as in visual information processing, mathematical structure (edges etc.) must be *imposed* upon the retinal array, because this structure is not literally present in the data, so some logical form must first be imposed on a problem requiring reasoning before the actual reasoning can take place.

In the end, we aim to convince the reader that using the formal machinery of modern logic leads to a much more insightful explanation of existing data, and a much more promising research agenda for generating further data. If we succeed, then there are general morals to be drawn about what philosophy of science is appropriate for the psychology of "higher" cognitive functions.

At present, experimental psychology is much influenced by a Popperian philosophy which sees hypothesis testing as the central activity in science. We will see that Wason himself was much influenced by this account of science in explaining his subjects' responses. But Popper's account, important as it is, in the intervening years has been shown to be a very partial account. If science is hypothesis testing, where do hypotheses come from? Why test this one rather than that? A great deal of science is exploration and observation which don't fall easily under the umbrella of hypothesis testing. Highly developed sciences such as physics, which are overwhelmingly the cases studied by philosophers of science, have powerful abstract bodies of theory which guide exploration

and observation, and through them the selection of hypotheses, when the time comes to test hypotheses. Psychology lacks such bodies of abstract theory, and so one sees implicit theories playing important roles in choosing what hypotheses to test. "Surprisingness" of a phenomenon as compared to some implicit theory of that phenomenon is a crucial quality indicating when hypothesis testing is worth the effort. If Wason's observations of failures of rationality in the selection task hadn't been so counterintuitive, then we (and you) would never have heard of them. We will expend considerable effort in chapter 3 in making explicit Wason's (implicit) background theory against which the results are so surprising, and in showing how logic can provide explicit background abstractions which change the way hypotheses can be chosen and experiments designed.

A corollary of the lack of background abstract theories in psychology is the use of direct operationalization of abstract concepts in experimental procedures: data categories are assumed to be very closely related to their theoretical categories. As a consequence, the data observed are supposed to have a direct bearing on the theoretical hypotheses. In mature sciences this doesn't happen. There are always 'bridging' inferences required between observation and theory, and an apparent falsification may direct attention to unwarranted auxiliary assumptions. Especially in chapter 10, though also throughout, we will illustrate how logic can open up space for observation and exploration between data and theory. Young sciences like psychology require lots of observation and exploration, so a methodology which opens up space for these activities is vital.

In the study of human cognition, this space between data and theory or hypothesis is particularly broad because of our human capacity for multiple interpretation. Indeed, some resistance to taking interpretation seriously in studying reasoning comes from the belief that this space is too broad to be bridged. Once we acknowledge the possibility of full human interpretive capacities, then, so the argument goes, the possibility of science goes out the window. In particular it is often felt that the scientific study of reasoning is impossible once it is allowed that the logical expressions are subject to the possibility of multiple interpretations. Rejecting the possibility of multiple interpretation because it is held to make science impossible is truly the logic of the drunk searching beneath the lamppost who prefers the illuminated circle to the dark space where he knows he lost his keys. We take the general human capacity for multiple interpretation to be as close to fact as it is possible to get in cognitive science, and prefer to follow it where it leads in choosing our methods of investigation. In fact, one sees this battle fought repeatedly in each area of human cognition. It remains under control to some extent in perception, because there the experimenter has "stimulus control" – she can twiddle the display, and observe what the subject reports seeing.[13] But in, for example, memory, the problem has been

13. In fact this is something of an illusion, because the subject brings preexisting knowledge to bear in

recognized ever since Ebbinghaus [68] invented nonsense syllables in order to eliminate the interpretive effects of subjects' long-term knowledge which is beyond experimental control. In memory, the problem leads to a split between the Ebbinghaus tradition of laboratory experimentation on abstracted materials, and the Bartlettian tradition of studying, for example, autobiographical memory [12]. Both have contributions to make, but both require an understanding of their distinctive approaches to assimilate those contributions.

The problem of treating "content" or "general knowledge" or "experience prior to the experiment" is endemic in psychology, and the life of the psychological researcher is much taken up with getting around the barriers it throws up. The psychology of reasoning's adoption of "abstract tasks" can be seen as following in the Ebbinghausian tradition, and reaction against those tasks as Bartlettian rejection of that tradition. Logic itself is interpreted by psychologists as the most extreme form of the Ebbinghausian approach in which content is banished entirely. However, our argument is that this is no longer true in modern logic. The default logics we present here are actually interpretable as modeling the relation between a working memory holding the experimental input materials and a long-term memory holding "general knowledge." So these logics present a formalization of "content." They thus attack psychology's central problem head-on. Of course, they do not offer a model of the long-term memory of some actual adult human being (nor even an idealized adult human being at standard temperature and pressure). But they do offer precise formal models of how large databases of default conditionals can control the interpretation of richly meaningful input texts. Essentially similar computer architectures are the basis for implementing real-world useful databases of general knowledge in practical applications. Of course the philosophical problems of "symbol grounding" remain. But nevertheless, here is the first plausible head-on approach to the formal modeling of content which offers to reconcile Ebbinghaus with Bartlett.

It should be evident that these issues are close to the heart of problems about the relations between the humanities and the sciences, and it is entirely fitting that they should come up when trying to do scientific research into the nature of the human mind. It is interesting that representatives of more conservative approaches to both sciences and the humanities have felt it important to try to defeat the very possibility of cognitive science's computational model of the mind. We hope to explain just why that model is felt to be so threatening, to defuse some of the concerns arising, and to show that one can avoid both misplaced scientific reductionism and postmodern hyperrelativism by engaging in a logically based and experimentally informed study of human interpretive capacities.

---

making perceptual interpretations, but it is an illusion persistent enough to allow progress.