Understanding Psychologists' Understanding

Eigner, K.

2010

UNDERSTANDING PSYCHOLOGISTS' UNDERSTANDING

Cover image: A subject performing a clock-watching task called the "speed and load test," in: R. Conrad, Speed and Load Stress in a Sensory-Motor Skill, *British Journal of Industrial Medicine* 8 (1951), 3.

Typeset by TAT Zetwerk, Utrecht
Printed by Wöhrmann Print Service, Zutphen

VRIJE UNIVERSITEIT

Understanding Psychologists' Understanding

The Application of Intelligible Models to Phenomena

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Wijsbegeerte
op woensdag 15 december 2010 om 9.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Gerrit Cornelis Eigner

geboren te Utrecht

promotor: prof.dr. J.A. Radder
copromotor: dr. H.W. de Regt

Contents

Preface

Writing this dissertation in philosophy of science was both an enormous challenge and an intellectual adventure. Now that it is finished, I would like to thank my supervisors Henk de Regt and Hans Radder for their support. Their enthusiasm, concern, assistance, and expert advice were invaluable to the success of this undertaking.

This dissertation is the end result of a Ph.D. project that was part of the research program "Understanding Scientific Understanding" of the Faculty of Philosophy at VU University Amsterdam. The aim of this program, which was supported financially by the Netherlands Organization for Scientific Research (NWO), was to investigate the nature of scientific understanding in a variety of scientific disciplines. Henk, who started the program, focused mainly on understanding in the physical sciences. Sabina Leonelli, who carried out the other Ph.D. project, concentrated on understanding in the life sciences, and I studied understanding in psychology. I enjoyed working with both, and I would like to take this opportunity to thank them for the pleasant and constructive teamwork that led to many satisfactory results. For example, together we organized the successful conference "Philosophical Perspectives on Scientific Understanding" held in Amsterdam in August 2005. Another highlight was the volume *Scientific Understanding: Philosophical Perspectives* (2009), which we edited together and in which I published some of the results of my research project. Especially chapter 4 of this dissertation overlaps with my contribution to that volume.

In the process of carrying out my Ph.D. research, I have benefited greatly from the help of many. Next to my supervisors, I would like to thank, in particular, Uljana Feest for her constructive comments on chapter 4. Furthermore, I would like to acknowledge the feedback from the members of the research group "Philosophy of Science and Technology" at VU University Amsterdam. In addition, I am grateful to Ivo Geradts and Johannes Rustenburg, not only for their excellent typographical work in this book but also for their warm collegiality

and support. A special thanks is extended to the philosophers octet "Otto e Mezzo." The opportunity to rehearse and perform music regularly by Weill, Shostakovich, and others together with fellow philosophers and friends is an enrichment of my (academic) life.

Last but not least, my deepest appreciation goes to my family, especially my wife Astrid and my son Jonne. They supported me greatly through their encouragement, cheerfulness, and – most of all – their understanding.

CHAPTER 1

Introduction

Scientific explanations provide understanding of phenomena. Although it seems quite natural to regard this as one of the main advantages of science, there was, until recently, a tendency among scientists and philosophers of science to downplay the value of scientific understanding. Illustrative for this is the ironic tone used by the psychologist Edward C. Tolman in a discussion on the development of behavioral theories. This leading figure of the psychological school of neo-behaviorism, which had its heyday between 1930 and 1950, is one of the subjects in chapter 4 of this study. In his view, theories of behavior are necessary primarily to gather the results of behavioral experiments and to use these results to generate new predictions. Furthermore, he suspected that some scientists have an additional motive for developing scientific theories:

> Some of us, psychologically, just demand theories. Even if we had all the million and one concrete facts, we would still want theories to, as we would say, "explain" those facts. Theories just seem to be necessary to some of us to relieve our inner tensions.
>
> (Tolman 1938/1966, 150)

Apparently, some scientists suffer from "inner tension" if they are confronted with unexplained empirical facts, and scientific theories can be helpful in reducing this psychological discomfort of lacking understanding. But, according to Tolman, this feature should not be taken seriously: he even put the word "explain" in quotation marks because he regarded it as nothing more than a psychological benefit. The view that is advocated with statements like this is that the psychological aspect of explanation – which was usually equated with understanding – does not belong to the genuine aims of science. This idea is in line with the view of science advocated by logical positivism.

Although it is currently more widely acknowledged that understanding is a central aim of science, in addition to accurate description and prediction, the influence of traditional positivist philosophy of

science can still be felt in philosophical debates on scientific understanding – for instance, in claims that understanding is merely a psychological by-product of scientific explanations (e.g. Trout 2002). In recent years, however, scientific understanding has received more serious attention in philosophy of science, especially in philosophical accounts of scientific explanation. Often, these accounts are justified by pointing to the capacity of an explanation to provide understanding. Typically, they make use of particular notions of understanding, such as the view that understanding is achieved through unification (e.g. Friedman 1974; Kitcher 1981; 1989; Schurz 1999) or through knowledge of causal mechanisms (e.g. Salmon 1984; 1990; 1998; Humphreys 1989; Dowe 2000). Although the topic of scientific understanding has acquired more interest due to the developments in the ideas about scientific explanation, the different models of explanation in philosophy of science still have no satisfactory underpinning for their accounts of understanding. For instance, as I discuss in chapter 3, the unificationist model lacks an adequate argument for the claim that understanding is achieved by unification, as does the causal-mechanical model for its claim that understanding is achieved by knowledge of causal mechanisms.

This present study is part of the research program *Understanding Scientific Understanding* developed by Henk W. de Regt. The aim of this program, in which Sabina Leonelli also participated, is to elaborate and defend a new approach to scientific understanding. This project breaks with the idea of universal standards for scientific understanding. The reason for this is the observation, made by historically-minded philosophers such as Thomas S. Kuhn (1962), that science as it is actually done displays a historical, social, and disciplinary variation that is bound to falsify any universal account of science. Therefore, our approach aspires to account for the contextual variation of standards for scientific understanding actually employed by scientists. A key notion in this approach is that of *intelligibility*. We assume that the scientific understanding of phenomena requires intelligible models or theories. Intelligibility is a pragmatic notion, which means that intelligibility is not an intrinsic property of the model or theory but depends on the scientists who use it. That intelligibility is a pragmatic notion implies the possibility of contextual variation. The intelligibility of models or theories may change over time. For instance, when

first proposed in the late 17th century, Isaac Newton's theory of gravitation was considered to be unintelligible because it involved action at a distance. However, in the 18th century, due to the successes of Newtonian theory, the theory became an exemplar of intelligibility. Because of intelligibility's dependence on context, standards for scientific understanding may vary over time and in different disciplines. To account for this contextual variation, the program is intended to explore a broad spectrum of scientific disciplines, including physics (De Regt 2005; 2009), the life sciences (Leonelli 2007; 2009), and the field investigated in the present study, psychology (Eigner 2009).

In chapter 2 I will discuss the aims and methods of philosophy of science in general and of this study in particular. Broadly speaking, an important aim of philosophy of science is to reveal the nature of science as an epistemic activity. The aim of this study is thus to develop an account of scientific understanding and to show that this notion is indispensable in a philosophical characterization of science. In line with the historicist approach in the philosophy of science, the use of historical case studies is an important feature of this study. In these case studies, the tentative account of intelligibility and understanding that I will develop in chapter 3 on the basis of recent philosophical literature is confronted with scientific practice. The inevitable tension between the philosophical characterization of science and actual scientific practices demands a critical approach to both the characterization and the practice. I will use the case studies of scientific practices to elaborate the philosophical account of scientific understanding in this study, which in turn can be used in the explanation and normative appraisal of scientific practices.

In chapter 3 I will develop my basic ideas about intelligibility and understanding. These ideas constitute the framework for the philosophical account of understanding that I will elaborate further via the case studies. The point of departure here is De Regt's work (2004; 2009; De Regt and Dieks 2005) on scientific understanding, which is the basis for the research program *Understanding Scientific Understanding*. One of the focal ideas here is that the scientific understanding of a phenomenon requires the *ability to use* the relevant scientific theories in particular ways. In developing this philosophical framework, I will concentrate on this ability and argue that it requires intelligible models. In chapters 4 and 5 I will discuss this framework in more detail by

means of two case studies of scientific practice. The first concerns the school of neo-behaviorism in psychology and the second its successor, cognitive psychology.

The main purpose of the first case study presented in chapter 4 is to support the claim that scientific understanding has epistemic significance by refuting a putative counterexample. Neo-behaviorism was strongly influenced by logical positivism, and Tolman's statement above exemplifies its rejection of any epistemic role of understanding in science. The big impact of neo-behaviorism on the practice of research in psychology, which lasted for several decades, may seem to disprove the idea that understanding is necessary for achieving the epistemic aims of science – at least in psychology. My analysis of the methodology of neo-behaviorism, however, invalidates this conclusion. The development of behavioral models led to the introduction of terms such as 'hunger' that were considered to be theoretical. This did not necessarily conflict with positivism as long as the meaning of these terms was captured in an objective way. For instance, in the case of rats in mazes, 'hunger' could be defined operationally in terms of the time passed since the rats last received food. However, in practice, the meaning of the theoretical terms turned out to exceed their objective definition. They acquired so-called "surplus meaning" (Reichenbach 1938). I will demonstrate that this surplus meaning of theoretical terms facilitates scientific understanding and that both the methodology and the resulting scientific knowledge crucially depend on this understanding.

The main purpose of the second case study presented in chapter 5 differs from the first. While the first case study demonstrates the epistemic significance of scientific understanding, I will use the case study of cognitive psychology to flesh out the notions of intelligibility and scientific understanding in more detail. In the 1950s, the rise of cognitive psychology was preceded by a more liberal approach to the introduction of explanatory constructs. At that time, neo-behaviorists gradually came to acknowledge that the surplus meaning of these constructs was relevant for their scientific endeavors. Thus, early cognitive psychologists, such as Donald E. Broadbent, advocated the use of a particular kind of explanatory concept to facilitate understanding of psychological phenomena. Here the contextual nature of scientific understanding becomes apparent: the early cognitive psychologists

believed that the concepts of information theory were especially highly suitable for understanding cognitive phenomena. Communication engineers were developing information theory at that time, which had a major impact in post-war technological developments. Early cognitive psychologists adopted ideas from this new discipline, and assumed that familiarity with the technique of providing information flow analyses made it easy to conceptualize psychological phenomena. I will examine the skills that are required for this technique, by means of which early cognitive psychologists applied information-theoretical models to understand psychological phenomena.

In sum, scientific understanding is an important epistemic aim of science. In this study I will show that scientific understanding requires intelligible models and that the intelligibility of a model for scientists depends on, among other things, their skills. I will use this account of understanding to analyze and explain scientific practices. For instance, I will use it to explain the transition from neo-behaviorism to cognitive psychology as motivated by the (implicit) recognition that the intelligibility of models is an important scientific value. With this account of understanding I hope to contribute to a characterization of science that is useful for the philosophical analysis of scientific practices and results.

CHAPTER 2

Aims and Method of this Study

2.1. *Introduction*

The main objective of this study is to develop an account of scientific understanding and to demonstrate that understanding is an important aim of science that has epistemic significance. As briefly mentioned in the first chapter, scientific understanding is a topic that was not traditionally treated by philosophy of science. Due to its psychological connotations, it was considered to be irrelevant for the main aim of philosophy of science, namely, the justification of scientific knowledge. In contrast, the present study argues that the development of an account of scientific understanding is a legitimate contribution to philosophy of science. In this chapter I will explain the aims and method of this study. In section 2.2 I will discuss the main aims of philosophy of science in a general way. In section 2.3 I will deal briefly with the general methodological implications of these ideas for philosophy of science studies. Subsequently, in section 2.4, I will explain the concrete method I will be using in this study. An important aspect of this method is the use of case studies, and in this study the two cases are taken from the field of psychology. The first concerns neobehaviorism, and the second cognitive psychology. I will explain how I use these case studies in the development of philosophical ideas about scientific understanding.

2.2. *Aims of Philosophy of Science*

Traditionally, a major aim of philosophy of science is to give a normative appraisal of science in general and of specific claims made by scientists in particular. This focus on the normative task of philosophy of science can be found in ideas on the discipline that were formulated by the logical positivists in the first half of the 20th century. In their characterization of philosophy of science a key role was assigned to

the distinction between the "context of discovery," in which the processes of the generation of knowledge in science are analyzed, and the "context of justification," in which the justification of this knowledge is analyzed. Philosophy of science was viewed as dealing with the context of justification, and the context distinction was meant to guarantee the autonomy of philosophy of science as an enterprise independent of other, empirical studies of science (Hoyningen-Huene 1987, 501).

Although the context distinction was traditionally an important subject in philosophy of science, in recent decades it has become less urgent (Schickore and Steinle 2006, vii). The reason for this decline of interest, which went hand in hand with a growing interest in the history and sociology of science, was not that philosophers of science had come to agreement on this topic. On the contrary, "agreement on the distinction between the contexts has never been reached, at least not publicly and explicitly. Rather, the interest in the distinction seems somehow to have faded away, without a real solution of the earlier disagreement" (Hoyningen-Huene 1987, 502). The context distinction fell out of favor because it was increasingly felt that that distinction was no longer needed for an accurate characterization of philosophy of science. In addition to justification, description and explanation also became valued as aims of philosophy of science. I will reflect on the aims of justification, explanation, and description, the common denominator of which is to provide a more or less general characterization of science. Following a proposal by Kuhn I will argue that such a characterization can be given by means of a list of epistemic values. I will start this reflection by looking more closely at the context distinction.

2.2.1. *The Distinction between Discovery and Justification*

During the Enlightenment, the idea flourished that the methods and processes of discovery in science were important for the justification of knowledge claims. This idea was based on the doctrines of philosophers and scientists, such as Francis Bacon, René Descartes, and Newton, who regarded the method of inquiry or discovery as a "straight and narrow path toward empirical or intellectual truth" (Nickles 2001, 86). They held that the method of discovery is also

a method of justification, and considered a claim to be justified if it had been arrived at by the right method. During the Romantic period, when intuition and creativity were highly valued, this view was abandoned because it became recognized that science could produce genuinely novel and interesting results only if scientists had flashes of genius based on intuition. Therefore, discovery and justification were considered to be different processes. After the Romantic period, the separation between discovery and justification culminated in the logical-positivist distinction between them (Nickles 2001, 86–87).

The core of the context distinction as intended by its proponents is a distinction between the factual and the normative (Hoyningen-Huene 1987, 511). When the well-known logical positivist Hans Reichenbach introduced the context distinction in 1938, he argued that analyses of science in the context of discovery concern the processes of thinking in their actual occurrence, whereas analyses of science in the context of justification, which is the realm of philosophy of science, concern the reconstruction of thought processes as they ought to occur if they are to be arranged in a consistent system of logical reasoning. According to Reichenbach (1938, 5), philosophy of science "does not regard the processes of thinking in their actual occurrence; this is entirely left to psychology." Instead, philosophy of science looks at a logical substitute: it regards thinking processes reconstructed in a rational form. Instead of describing the actual thinking processes, philosophy of science offers a rational reconstruction of them in which they are transformed into a chain of logical steps. This rational reconstruction can be subjected to logical evaluation by examining if all chains of thought are justified.

Reichenbach's view of the task of philosophy of science, which is closely connected to his view of the context distinction, appears to be at variance with the conception that most philosophers of science have of their discipline. This becomes apparent already from the different interpretations of the context distinction that were put forward by philosophers of science after Reichenbach introduced it. The most influential interpretation that differs from Reichenbach's to some extent is that the two contexts refer to temporally distinct, or perhaps partially overlapping, successive phases in science (Hoyningen-Huene 1987, 507). The origin of this interpretation was Karl R. Popper's *Logik der Forschung* (1934), which became well known after its

English translation as *The Logic of Scientific Discovery* (1959). In this work Popper distinguishes between two temporally distinct and successive stages of scientific development, namely, the stage of conceiving or inventing a theory and that of testing it. According to Popper, the first stage is not susceptible to logical analysis, although it may be of interest to psychology. The second stage however, in which scientific knowledge is justified by means of critical tests, can be analyzed logically. Popper's distinction between temporally distinct stages, which he called the "conduct of discovery" and the "conduct of justification" (*Auffindungs- und Rechtfertigungsverfahren*, cf. Popper 1935) respectively, differs from Reichenbach's context distinction. However, Popper's distinction would become confused with the latter. Reichenbach had introduced the distinction as a logical one, rather than a chronological one, and he had no intention of distinguishing two different stages of scientific inquiry. In later discussions on discovery and justification this was not clear to most of the parties. It was almost universally accepted that the context distinction was a distinction between an initial developmental phase and a phase of testing what had been developed, which is actually the distinction put forward by Popper instead of Reichenbach (Gutting 1980, 30).

In addition to the philosophical question about the justification of knowledge, which is the focal point of Reichenbach's distinction, Popper's distinction entails the historical question of the existence of distinct phases of discovery and justification in scientific inquiry. Several answers to this question were proposed. Some denied the possibility of separating different phases, while others even argued for a separation into three or more phases. It became common to distinguish three phases: the phase of generating or conceptualizing a hypothesis, the phase of pursuit or preliminary evaluation of this hypothesis, and finally the phase of justification or acceptance of this hypothesis (Kirschenmann 2001, 8). The growing interest of philosophers of science in the stages of discovery did not imply a return to the focus on scientific methodology that was characteristic of the Enlightenment. Instead, it implied a growing awareness of actual scientific practice. The philosophical studies of the stages of inquiry in the development of scientific knowledge were not purely normative. In addition, they also aimed at describing and explaining these stages of scientific inquiry. In the course of time, especially after Kuhn's (1962) *Structure of*

Scientific Revolutions, description and explanation of science became appreciated as additional aims of philosophy of science supplementary to justification. The aims of philosophy of science go beyond what Reichenbach determined was the task of the discipline. His characterization of philosophy of science via the context distinction is too restrictive.

All these aims – describing, explaining, and giving a normative appraisal of (the results of) scientific practices – are objectives of my study of scientific understanding in psychology. I will first *describe* aspects of the scientific practices of neo-behaviorism and early cognitive psychology. Second, I will *explain* aspects of these developments in psychology. For instance, I will use philosophical ideas on scientific understanding and on the intelligibility of theories to explain that the transition from neo-behaviorism to cognitive psychology was motivated by the (implicit) recognition that intelligibility of theories is an important scientific value. Third, I will use philosophical ideas about scientific understanding and the intelligibility to give a *normative appraisal* of scientific practices and ideas, especially of neo-behaviorism. Normative appraisals can be evaluative and prescriptive. An example of a prescriptive assessment of neo-behaviorism is the well-known phenomenological critique that, due to its objective approach, its portrayal of human nature is inadequate. This judgment is prescriptive because it results in the prescription that instead of the objective approach of the natural sciences, which the hermeneutical tradition refers to as *Erklären*, psychology should employ a qualitative or hermeneutical approach, which is called *Verstehen*, because that would do more justice to the psychological subject matter (Feest 2005, 145). In my study, the normative appraisal of neo-behaviorism will primarily be evaluative rather than prescriptive. Instead of prescribing an alternative approach, I will use my philosophical ideas about intelligibility and scientific understanding to evaluate the scientific character of the neo-behaviorist approach. I will analyze the scientific approach of neo-behaviorists such as Tolman and Clark L. Hull and assess to what extent their assertions about science as well as their actual scientific activities accord with the epistemic requirement of intelligibility.

2.2.2. *Characterizing Science by its Values*

An important task for philosophers of science, no matter whether they are interested primarily in describing, explaining, or giving a normative appraisal of science, is to give a more or less general characterization of science. Traditionally, they tried to do so by searching for universal and objective epistemic norms. For instance, logical positivists tried to formulate such norms by means of the verifiability principle, which states that statements are meaningful only if they are either empirically verifiable or tautological. They argued that scientific knowledge should therefore be based strictly on empirical evidence and logic. This project, however, faced several difficulties: for instance, universal claims, such as laws of nature, do not allow for verification. Tackling these difficulties became the main business of the logical positivists. Famous here are the problems related to the confirmation of knowledge, the Duhem-Quine thesis, and Rudolf Carnap's attempt to construct an inductive logic to justify universal claims by means of a finite number of observations. Another well-known problem concerns the distinction between theoretical and observational knowledge claims (e.g. Hanson 1958).

The impossibility of solving the key problems of logical positivism, together with an increasing historical interest in the processes of knowledge generation in real science, led to doubts about the meaningfulness of the search for universal, objective epistemic norms. Historically-minded philosophers of science, of which Kuhn is the most famous representative, noted that the norms used in scientific practice are context-dependent: different scientific communities or disciplines use different norms.

In the case of norms for the intelligibility of theories, this context-dependence is nicely illustrated by the example of action at a distance. This concept was felt to be unintelligible when it was introduced by Newton in 1687 because it was in conflict with the prevailing norm, based on the corpuscular worldview of that time, that an intelligible mechanics should describe interactions between bodies that have direct contact. This was expressed in the dictum that "nothing can act where it is not." However, a century later, when Newtonian theory turned out to be very successful, scientists abandoned the metaphysical principles of philosophers like Descartes and adopted the canon

that "a thing can only act where it is not," and, accordingly, the concept of action at a distance became a necessary condition for intelligibility (Van Lunteren 1991, 126). This demonstrates the contextual nature of norms for intelligible theories.

According to historically-minded philosophers of science, scientific practice shows that no universal epistemic norm exists. Therefore, they argue, the logical-positivist ideal of universal norms does not do justice to the dynamics of science. This line of reasoning is currently influential in several trends in philosophy of science in which universal epistemic norms are being replaced by context-dependent norms (Kirschenmann 2001, 3–7).

The view that epistemic norms in science are context-dependent raises the question if it is feasible to give a more or less general characterization of science. Kuhn proposed a way to deal with this problem. Although his analysis of the history of science shows that attempts to formulate universal epistemic norms will always be refuted, in Kuhn's view this does not mean that philosophers of science should stop searching for characteristic features of science. Such features can still be formulated, although they have to be formulated in such a way that they allow for historically situated scientific practices and different scientific schools. In his characterization of science, Kuhn (1973/1977) did this by listing a number of general characteristics of a scientific theory. The first is empirical accuracy, which includes predictive accuracy; the second is consistency, both internal and with other relevant accepted theories; the third is scope or unifying power; the fourth is simplicity, which involves organizing otherwise isolated phenomena; and the fifth is fruitfulness, or fertility for further research (Kuhn 1973/1977, 321–322). Kuhn argued that this list of criteria contains necessary (but not sufficient) conditions for a discipline to be scientific: an enterprise may have different criteria for judging their theories, but then it would not be science (Kuhn 1973/1977, 331).

Kuhn's list of criteria is compatible with his dynamic view of science. He deliberately formulated the criteria in an "imprecise" way, meaning that "individuals may legitimately differ about their application to concrete cases" (Kuhn 1973/1977, 322), and left open the possibility of disagreement about how the criteria play off against one another in conflicting situations. According to Kuhn (1973/1977, 331; see also McMullin 1983, 16), the criteria in his list operate as

"values," which means that theory choice is basically a matter of value judgment. All disciplines that endorse these values are scientific disciplines, even if the translation of these values into concrete norms differs among these disciplines. Kuhn's way of characterizing science by means of a list of values was a source of inspiration to other philosophers of science (e.g. McMullin 1983; Longino 1990; Lacey 2005). Ernan McMullin adopted Kuhn's ideas of scientific values and called them "epistemic values," which is also the term I will use in this study:

> We can provide a list of criteria that have gradually been shaped over the experience of many centuries, the values that are implicit in contemporary scientific practice. Such characteristic values I will call epistemic, because they are presumed to promote the truth-like character of science, its character as the most secure knowledge available to us of the world we seek to understand. An epistemic value is one we have reason to believe will, if pursued, help toward the attainment of such knowledge. (McMullin 1983, 18)

McMullin distinguishes epistemic values from non-epistemic values that (in the long run) are not relevant for theory choice. This seems to suggest that there is a relatively firm distinction between epistemic values and other values, such as social, cultural, and personal ones. Helen E. Longino (1990, 4) questions this. Although she distinguishes in her work between scientific values, which she labels "constitutive values" because they are "generated from an understanding of the goal of science," and other, contextual values, she seems to suggest that in practice this distinction is not that sharp. This becomes apparent in, for instance, her idea that the origin of a constitutive value might be a contextual value (1990, 100). I agree with Longino that contextual value judgments play an important role in science. In my reading of Kuhn, the contextual influence on the choice of theory is already an important element of his conception of the list of epistemic values because, as he explains, in scientific practice the translation of epistemic values into concrete epistemic norms may differ among different scientists, depending on their disciplinary background and other contextual factors. Characterizing science by means of a list of epistemic values is a drastic deviation from the traditional view of philosophy of science. First, the attention to the context-dependence of epistemic norms deviates from the traditional ideal of universal epistemic norms. Further,

the list of epistemic values proposed by philosophers such as Kuhn, McMullin, Longino, and Hugh Lacey may include elements that were traditionally not considered as belonging to the context of justification.

For example, the list of epistemic values may contain pragmatic elements, which is not in accordance with traditional philosophy of science. In line with Reichenbach's idea of offering rational reconstructions, traditional philosophy of science was concerned with the justification of scientific knowledge, where the basic idea was that this justification relies only on empirical evidence and logic. Despite the difficulties mentioned above with specifying this basic idea, it was clear that the philosophical justification should not rely on pragmatic considerations. Pragmatic aspects of science were not considered relevant for philosophy of science (cf. Van Fraassen 1980, 87–96). A characterization of science by means of a list of epistemic values may contradict this view. One of the elements in Kuhn's list that can be regarded as pragmatic is the value of simplicity: whether a theory is simple or difficult depends on the experience and intelligence of the scientists who use it. However, Kuhn's notion of simplicity implies bringing order to otherwise isolated phenomena. Instead of simple as opposed to difficult, he meant simple as opposed to complex, which is not evidently a pragmatic notion. Yet, whether or not Kuhn's value of simplicity does indeed include a pragmatic element, characterizing science by means of a list of epistemic values has the advantage of facilitating the inclusion of aspects of science that would traditionally have been excluded from a philosophical characterization of science.

In this study I argue that intelligibility of theories is an epistemic value of science. In other words, like empirical accuracy, consistency, scope, simplicity, and fruitfulness, intelligibility is one of the values relevant for judging theories (cf. De Regt, Leonelli, and Eigner 2009). Intelligibility is a pragmatic notion: a theory's intelligibility does not only depend on its relation to the world but also on those who use the theory. The claim that intelligibility is one of the epistemic values that are constitutive of science therefore broadens the traditional domain of inquiry of philosophy of science. I will support this claim by means of historical case studies from psychology.

2.3. *Methodology in Philosophy of Science*

That philosophy of science is an undertaking with both normative and descriptive facets has repercussions for methodology. Two opposing methodologies that are advocated by philosophers of science are the aprioristic approach and the empirical approach. In the aprioristic approach, of which logical positivism is illustrative, the nature of science is analyzed *a priori*, for instance, by means of postulated epistemic norms that are explicated analytically and eventually used for giving an appraisal of concrete scientific claims. The aprioristic approach is therefore primarily normative. In contrast, in the empirical approach the nature of science is analyzed *a posteriori* by means of empirical investigation. For instance, since the "historical turn" in philosophy of science, which was initiated largely through Kuhn's (1962) ideas concerning scientific development, it became common to turn to history to learn about the nature of science. This empirical approach is primarily descriptive rather than normative.

The aprioristic approach is a *top-down* one. The epistemic norms, which are supported by or derived from strictly philosophical considerations that are usually epistemological (such as the idea of the unity of science) and sometimes metaphysical (such as the idea that reality is simple), are used to evaluate scientific practice. On the other hand, the empirical method, in which normative claims about science are derived from empirical case studies, is a *bottom-up* approach (Burian 2001, 386):

> The point of case study methods is to work up from an appreciation of the scientific work in its context. A case study does its job only if it yields improved understanding of how scientists solved (or failed to solve) problems, what methods they used or tried to use, how their various tools were made to interact, how they evaluated hypotheses and factual claims, and so on. If one is willing to count work as genuinely scientific only if it meets a pre-set criterion or general aim (such as truth seeking), then one is not honestly working bottom up and risks misunderstanding the case. (Burian 2001, 388)

Ideally, in the top-down approach, philosophical criteria and norms are developed independently of scientific practice, whereas in the bottom-up approach, the nature of science is examined by means of case studies of scientific practice that are not biased by philosophical

presuppositions. However, both methods are ideal types. As Hans Radder argues, merely using the top-down approach and thus completely ignoring scientific practice is impossible:

> After all, the normative philosophers would certainly be embarrassed if their criteria appeared to be generally violated. Therefore, most of them "have in mind" some exemplars of what they take to be good science and good scientists. These exemplars, which are supposed to illustrate the advocated philosophical criteria, get a paradigmatic function for all of science. Then, from a quick look at one or two other episodes, it is concluded that, although of course some irrational deviations occur, science "mostly" conforms the philosophical criteria in question. (Radder 1996, 176)

Using only the bottom-up approach, independently of philosophical presuppositions about science, is also impossible. The main reason for this is that "even to conceive of a history of science one has to decide what counts as science and this arguably requires the introduction of evaluative norms" (McAllister 1986, 318). Therefore, the problem of working with case studies is that in order to know that the chosen case is appropriate, "one would already have to know how to determine whether a case should count as an instance of good science" (Burian 2001, 386). Therefore, this approach has the drawback of circularity. The philosophical conclusions depend partly on the initial presuppositions about the character of science. Inferring characteristics of science from historical episodes requires making a selection from available historical data and arranging them in a particular way. This is necessary, for "among the vagaries and vicissitudes of history" one can find evidence to support almost any conceptual point (Nickles 1986, 256). Therefore, determining the characteristics of good science requires some philosophical presuppositions about good science beforehand. This does not imply that the circularity is vicious. For instance, Ronald N. Giere (1999a, 160) argues that the circularity objection can be countered by means of an evolutionary epistemology in which the concepts of biological evolution are applied to the growth of scientific knowledge.

The top-down approach is not possible without having in mind some exemplary cases of good scientific practice, and the bottom-up approach is not possible without some philosophical criteria for good scientific practice. Therefore, developing an account of the nature

of science has to involve both the top-down and the bottom-up approaches. Illustrative is Kuhn's list of the epistemic values of science. From a top-down perspective, Kuhn's list appears to be an *a priori* postulation of epistemic values. In this view, the items on that list, such as empirical accuracy or consistency, can be explicated analytically, and eventually they can be used in normative appraisals of concrete scientific practices. From a bottom-up perspective, the list is a hypothesis about the nature of science that is open to modification when it is put to the test in case studies. Cases in which the development of scientific practice is not in accordance with the scientific values listed require that philosophers take a stand towards science. Then, either the case is regarded as an instance of good scientific practice, which means that the list should be altered, or the list is regarded as a proper characterization of science, which means that the case should be rejected as an instance of good scientific practice. According to Radder (1997, 649–650), taking a stand on these cases is a normative and reflexive affair that takes into account "the situatedness of philosophers, philosophical communities, or philosophical positions." Because philosophers of science are "committed to particular interpretative and explanatory preconceptions," their interpretations of science cannot be inferred from the historical case studies alone:

> The plausibility of philosophical interpretations certainly does not depend exclusively on empirical evidence about the historical development of scientific theories or the historical practice of working scientists. (Radder 1997, 650)

Although philosophical interpretations of science do not have to conform fully to what scientists actually do and believe, they must be informed by the development of research programs and the beliefs of scientists (Vicedo 1993, 494–495).

The tension between philosophical interpretations of science on the one hand and actual scientific practices on the other is especially tangible in the case study of neo-behaviorism in chapter 4. In my interpretation of science, one of the main characteristics of science is that it aims at understanding. At first sight, neo-behaviorism seems to be a scientific practice that does not fit this characterization. Due to the plea for a positivist attitude and the resulting refusal to speculate about what happens inside an organism, neo-behaviorists were called "black-box" psychologists. Instead of aiming at insight, they claimed

that prediction and control were the sole aims of science. I will critically analyze the practice of neo-behaviorism, and investigate if it does indeed contradict my characterization of science. I will demonstrate that there is a discrepancy between the words and deeds of the neo-behaviorists. Despite their plea for a positivist attitude, in their scientific endeavor they did aim at understanding. I will show that this discrepancy was the underlying motivation for a debate in the 1950s among neo-behaviorists and logical positivists that heralded changes in the field, which eventually resulted in the rise of cognitive psychology.

In this study, the interplay between the bottom-up and top-down approach in the philosophical analysis of the case studies will have a twofold effect. On the one hand, it will help in elaborating the philosophical framework for scientific understanding, and, on the other, it will help in understanding aspects of scientific developments like the transition from neo-behaviorism to cognitive psychology.

2.4. *Reflections on the Specific Method Used in this Study*

The method in this study is both top-down and bottom-up. In chapter 3, I will develop a philosophical framework for an account of scientific understanding. In chapters 4 and 5, which consist of case studies, I will flesh out this framework by confronting it with scientific practice.

In chapter 4, the case study of neo-behaviorism, I will focus on the epistemic significance of scientific understanding. The main purpose of this case study is to argue for the claim that understanding is an epistemic aim of science and that the intelligibility of theories is an epistemic value. Because neo-behaviorism seems to be a counterexample to this claim, studying this practice is in line with Popper's (1962/1989, 52) idea of the critical attitude in which a researcher has to investigate "the most severe experimental tests which his theories and his ingenuity permits him to design." Demonstrating that my characterization of science obtains even in the case of neo-behaviorism would be a major corroboration of that characterization.

In chapter 5, which is on cognitive science, I will focus on intelligibility as an epistemic value. The main purpose of this case study is to elaborate the notion of intelligibility developed in chapter 3. I will investigate the conditions under which theories are intelligible

to scientists. Because cognitive psychologists, who reacted to aspects of neo-behaviorist methodology, explicitly explored methods to gain scientific insight into cognitive processes, the school of cognitive psychology is interesting as a case study for analyzing the conditions for the intelligibility of theories.

In both case studies I will focus on individual scientists. In the case of neo-behaviorism I will look in particular at the work of Tolman (1886–1959) and Hull (1884–1952). Both were major figures in their discipline and made major methodological and theoretical contributions to it. In the case of early cognitive psychology I focus on the work of Broadbent (1926–1993), who is recognized for his theoretical contributions that helped shape the field. Although the focus is on individual scientists, the case studies attempt to describe how their work was embedded in their discipline.

One reason for taking the disciplinary context into account is to invalidate a traditional motivation for excluding concepts such as understanding and intelligibility from the realm of philosophy of science, namely, the idea that these notions depend on the idiosyncrasies and changing tastes of individual scientists (cf. Friedman 1974, 14). My analysis of scientists in their context will show that this idea is a misconception. As De Regt and Dennis Dieks argue, although scientists can have quite different specific views about intelligibility, the most important variation occurs between scientific communities instead of between individual scientists. Even though "scientists in different historical periods or in different communities have quite different specific views about precisely how scientific understanding is to be achieved … within a particular community standards of intelligibility are usually shared" (De Regt and Dieks 2005, 140).

My reconstructions of neo-behaviorism and early cognitive psychology are based on primary and secondary sources, such as articles and books by the psychologists in question and historical textbooks on the relevant psychological approaches. Although the reconstructions will probably not yield new historiographic results, this does not constitute a problem because my primary aim is philosophical rather than historical. I will use the reconstructions to elaborate my philosophical account of scientific understanding – and at the same time will use my ideas about intelligibility and understanding to reconstruct aspects of the history of psychology. Although the linear order

of the chapters in this book may suggest the opposite, both the reconstructions of neo-behaviorism and early cognitive psychology and the philosophical account of intelligibility and scientific understanding are the result of an iterative process of moving back and forth between the top-down and bottom-up approaches.

CHAPTER 3

Understanding Scientific Understanding

3.1. *Introduction*

Although the topic of scientific understanding has been the object of increasing attention in recent years (De Regt, Leonelli, and Eigner 2009), this topic was traditionally not of interest to philosophy of science. For instance, Carl G. Hempel, who developed the deductive-nomological model of scientific explanation (Hempel and Oppenheim 1948; Hempel 1965), argued that the notion of understanding lies outside the scope of philosophy of science, because it refers to the "psychological and pragmatic aspects of explanation" (Hempel 1965, 413). In his view, the psychological and pragmatic aspects of explanation are not relevant for philosophy of science. The notion of understanding is pragmatic because it refers to the person(s) involved in the process of explaining (Hempel 1965, 425). Hempel put the psychological aspects of explanation, such as the feeling of familiarity that explanations can produce, and the pragmatic aspects of explanation in the same category. He argued that explanation has these aspects because "[w]hether a given argument Y proves (or explains) a certain item X to a person P will depend not only on X and Y, but quite importantly also on P's beliefs at the time as well as on his intelligence, his critical standards, his personal idiosyncrasies, and so forth" (Hempel 1965, 426).

Hempel contrasted the pragmatic aspects of scientific explanation with its logical aspects. The logical aspects concern the two-term relation between what needs to be explained (the explanandum) and that which contains the explanation (the explanans), whereas the pragmatic aspects concern the three-term relation involving explanans, explanandum, and a subject. According to Hempel, the logical aspects of explanation are a legitimate topic within philosophy of science. The deductive-nomological model, in which an explanation is described as a logically valid argument in which the explanandum is deduced

from an explanans containing at least one universal law and relevant initial and background conditions, is an attempt to specify this logical aspect. In contrast, he argued, the pragmatic aspects of explanation do not belong to the domain of philosophy of science. The reason for this is that philosophers of science should be interested in an objective notion of explanation, whereas the pragmatic aspects depend on the subject's "knowledge, interests, intentions, and so forth" (Hempel 1965, 424), a fact that, according to Hempel, makes them subjective and hence philosophically irrelevant.

Several philosophers contested this by pointing out that there are pragmatic notions – such as knowledge if it is interpreted as "justified true belief" – that are clearly objective (e.g. Scriven 1962/1988, 53; Friedman 1974, 7–8). Michael Friedman argued that like the notion of understanding, the notions of knowledge and rational belief are also pragmatic because they refer to "the thoughts, beliefs, attitudes, etc. of persons." Friedman concluded from this that the pragmatic nature of understanding does not imply that there cannot be an objective or rational sense of scientific understanding. Therefore, he did not see "how the philosopher of science can afford to ignore such concepts as 'understanding' and 'intelligibility' when giving a theory of the explanation relation" (Friedman 1974, 8). Others, however, followed Hempel in drawing a sharp distinction between the epistemic and pragmatic aspects of science and asserting that the pragmatic aspects are irrelevant for a philosophical account of the epistemic aspects of science. For instance, Bas C. van Fraassen argued that explanatory power is a pragmatic virtue of theories, which does not give an *extra* good reason for accepting the theory:

> [E]xplanation is not a special additional feature that can give you good reasons for belief in addition to evidence that the theory fits the observable phenomena. For 'what more there is to' explanation is something quite pragmatic, related to the concerns of the user of the theory and not something new about the correspondence between theory and fact. (Van Fraassen 1980, 100)

In my view, a sharp distinction between the epistemic and pragmatic aspects of science is untenable. An important aim of this study is to demonstrate that intelligibility and scientific understanding, which are pragmatic notions, have epistemic significance. In this chapter I will develop a philosophical account of scientific understanding. In

section 3.2 I will discuss the significance of the notion of scientific understanding for the development of a philosophical account of scientific explanation. In section 3.3 I will develop a framework for my account of scientific understanding that is inspired by the work of De Regt (De Regt 2004; 2009; De Regt and Dieks 2005). A central idea is that the scientific understanding of a phenomenon implies being able to use the relevant scientific theories. In section 3.4 I will elaborate this framework by integrating philosophical ideas concerning scientific theories, models, and phenomena. Here I will use the representational view of models developed by Giere (1999b; 2004). In section 3.5 I will explore the epistemic conditions for the application of the relevant scientific theories to phenomena and relate this topic to a key notion of this study, namely, the notion of intelligible models. A complete account of scientific understanding also needs input from a study of scientific practice. In chapters 4 and 5, I will use historical case studies to elaborate the framework and analyze the role of understanding in science. At the end of this chapter, I will formulate key questions of this study that will be answered after the analysis of the historical cases.

3.2. *Scientific Understanding and Explanation*

Due to the focus on the two-term relation between explanandum and explanans, Hempel excluded the role of understanding in his account of scientific explanation. Several philosophers have argued that, because of that, his deductive-nomological model neglects an essential aspect of scientific explanation. One of the first to argue for a more prominent role of understanding in the analysis of explanation was Michael Scriven. According to him, "whatever an explanation *actually* does, in order to be called an explanation at all it must be *capable* of making clear something not previously clear, that is, of increasing or producing understanding of something" (Scriven 1962/1988, 53). Friedman also considered explanation to be intrinsically related to understanding and argued that one of the requirements for a good philosophical theory of explanation is that it "should somehow connect explanation and understanding" (Friedman 1974, 14). In several accounts of scientific explanation developed after Hempel's deductive-nomological model, the leading notion was a particular

view of scientific understanding. For instance, Friedman based his unificationist model of scientific explanation on the idea that scientific explanations provide understanding because they present a unified picture of the world:

> [S]cience increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more.
>
> (Friedman 1974, 15)

Therefore, according to Friedman, the essence of a scientific explanation is that it reduces the number of independent phenomena. The unificationist model became an important conception of scientific explanation (e.g. Friedman 1974; Kitcher 1981; 1989; Schurz 1999).

Another view of scientific explanation was the idea that scientific explanations also provide understanding of phenomena because they reveal the underlying causal mechanisms. For instance, Wesley C. Salmon, a major advocate of the causal-mechanical model of explanation, argued that "underlying causal mechanisms hold the key to our understanding of the world" (Salmon 1984, 260). Causal-mechanical explanations have understanding-providing power because "causal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand why certain things happen, we need to see how they are produced by these mechanisms" (Salmon 1984, 132). The causal-mechanical model became an important conception of scientific explanation (e.g. Salmon 1984; 1990; 1998; Humphreys 1989; Dowe 2000).

Several other accounts of explanation have been developed in addition to the unificationist and the causal-mechanical models. Philosophy of science has not produced any general framework to account for this diversity. William H. Newton-Smith, who argues that philosophers should want "some deeper theory that explained what it was about each of these apparently diverse forms of explanation that makes them explanatory," calls the present situation, in which such a theory is lacking, "an embarrassment for the philosophy of science" (Newton-Smith 2000, 130–132). A first step in the development of such a general framework for scientific understanding is the search for common features of the different types of scientific explanation. According to De Regt (forthcoming), one feature that is

often mentioned in accounts of explanation developed after Hempel's deductive-nomological model is that scientific explanations provide understanding. Therefore, he argues, one way to account for the plurality of forms of explanation is to invoke the notion of scientific understanding: there exists a variety of explanatory strategies to reach a single aim, namely scientific understanding.

Although the development of accounts of explanation, such as the unificationist model and the causal-mechanical model, was motivated by intuitions about the relation between explanation and understanding, until now the different models of explanation in philosophy of science lack satisfactory analyses of the notion of scientific understanding. For instance, Eric Barnes argues that proponents of the unificationist model have not provided an adequate argument for the thesis that understanding is achieved by unification. Instead, they take the connection between unification and understanding to be straightforward (Barnes 1992, 6). Similarly, proponents of the causal-mechanical model, including Barnes himself, have not provided an adequate argument for the thesis that phenomena are understood in virtue of knowing their causal basis. Ironically, this can be illustrated by Barnes' own attempt to justify this thesis, in which he says no more than that to search for understanding of almost any empirical fact "is just to seek the knowledge of its causal basis" (Barnes 1992, 8). Typically, proponents of the causal-mechanical model take the connection between knowing the causal basis and understanding to be straightforward. A theory of scientific understanding that clarifies how the different types of explanation achieve understanding could provide the desired framework to account for the diversity of types of explanation and would, accordingly, be a substantial contribution to philosophy of science.

3.3. *Basic Ideas of My Account of Scientific Understanding*

To develop an account of scientific understanding, I will make use of the historical case studies in chapters 4 and 5. However, as I discussed in chapter 2, to analyze the material of the case studies and understand its consequences for my philosophical account, it is necessary to have some basic ideas about scientific understanding. In this chapter I will formulate a general framework for scientific understanding

based on ideas posited by De Regt (2004; 2009; De Regt and Dieks 2005) concerning scientific understanding and the intelligibility of theories and by Giere (1999b; 2004) concerning the use of models in science.

In his search for an account of scientific understanding De Regt draws on the intuition shared by several philosophers (e.g. Wittgenstein 1953, sections 151–155; Kitcher 1989, 437–438) that scientific understanding of a phenomenon not only implies *knowing* the relevant scientific theories, laws, and background conditions but also *being able to use* them in the case at hand (De Regt 2004, 101). This ability to use the relevant scientific theories, laws, and background knowledge (henceforth "theory") is specified by De Regt and Dieks (2005, 149) as the ability to apply the theory successfully to concrete situations. In their view, scientific understanding of a phenomenon implies knowing the relevant scientific theory and being able to apply it successfully to this phenomenon.

According to De Regt, a prerequisite for this application of a theory to a phenomenon is that the theory is *intelligible* to scientists. He defines the intelligibility of a theory as the positive value that scientists attribute to the cluster of theoretical virtues that facilitate their use of the theory (De Regt 2004, 103; 2009, 31). In this study I will adopt De Regt's terminology: intelligibility of theories (and models) is an epistemic *value*, and theories (and models) can have *virtues* that render them intelligible to their users. Whether a property of a theory is such a virtue depends on its users: for a theory to be intelligible, the skills of the scientists and the virtues of the theory should fit together like a key in a lock:

> [W]hether scientists are able to use a theory for explaining a phenomenon depends both on their skills and on the virtues of the theory. More precisely, it depends on whether the *right combination* of scientists' skills and theoretical virtues is realized. Particular virtues of theories, e.g., visualizability or simplicity, may be valued by scientists because they facilitate the use of the theory in constructing models and predicting or explaining phenomena; in this sense they are pragmatic virtues. But not all scientists value the same qualities: their preferences are related to their skills, acquired by training and experience, and to other contextual factors such as their background knowledge, metaphysical commitments, and the properties of already entrenched theories. (De Regt 2004, 103)

De Regt argues that his analysis of scientific understanding provides the desired framework that can account for the plurality of types of scientific explanation. For instance, the causal-mechanical nature of a particular theory could match someone's skill in causal reasoning and thereby enable the successful application of the theory to a phenomenon (De Regt and Dieks 2005, 153). In the causal-mechanical model of explanation, causality is promoted as *the* standard virtue. However, as De Regt argues, it is not necessary that theories be causal-mechanical for them to be applied. A theory may have other virtues that enable its application, such as visualizability or simplicity. Whatever virtues are preferred depends on the skills of the scientists in question and is thus a pragmatic and context-dependent issue.

This view of scientific understanding accommodates not only the causal-mechanical model of explanation but also the unificationist model – at least in the way Philip Kitcher advocates it (De Regt and Dieks 2005, 149). According to Kitcher (1989, 438), applying scientific theories to concrete situations requires a skillful cognitive effort, namely, that of using an internalized set of argument patterns associated with the theory. These internalized patterns, consisting of schematic arguments and corresponding filling instructions that specify what sorts of entities can be filled in for each term in the argument, are the *know-how* that scientists need to apply to theories. Kitcher (1989, 432) argues that the purpose of using argument patterns is to realize unification by reducing "the number of types of facts that we have to accept as ultimate (or brute)," since this reduction "advances our understanding of nature." A scientist's ability to apply a theory successfully to a phenomenon depends both on the virtues of the theory – namely which type of argument patterns can be associated with it – and on the skills of the scientist – namely the capacity to use these argument patterns.

In sum, one basic idea in the framework for scientific understanding, which will be elaborated further in this chapter, is that scientists understand a phenomenon if they are able to apply a theory successfully to it. A second basic idea is that a prerequisite for the applicability of a theory is that it be intelligible to the scientists. This is the case if the theory possesses virtues that, because they match the skills of the scientists, facilitate its successful application to the phenomenon.

As an illustration of this view of understanding, De Regt and Dieks (2005) mention the way in which Ludwig E. Boltzmann's kinetic

theory of gases provides insight into the behavior of gases. In his famous *Lectures on Gas Theory* (1896/1964), Boltzmann pictured gases as a collection of freely moving molecules in a container. This picture is generally considered to provide understanding of certain aspects of the behavior of gases, such as the fact that a gas exerts pressure on the walls of the container and that a decrease in the volume of the container results in an increase of pressure. The picture implies that if a gas molecule collides with a wall of the container, it gives a little push, and the total effect of the molecules pushing produces the pressure. A decrease in the volume of the container will cause an increase in the number of molecules per unit of volume, and this causes an increase in the number of impacts per unit of time on the wall surface and thus an increase in pressure. According to De Regt and Dieks, the kinetic picture of gases is intelligible to us due to the causal-mechanical nature of the kinetic theory, which is a theoretical virtue that matches our skill of causal reasoning. In their view (De Regt and Dieks 2005, 151; De Regt 2009, 33), a sufficient but not necessary criterion for the intelligibility of a theory for scientists is that they can recognize qualitatively characteristic consequences of the theory without performing calculations. The kinetic theory satisfies this criterion because we are able to apply the theory successfully to particular situations: "the general picture of the moving gas particles allows us to make qualitative predictions of macroscopic properties of gases in particular situations" (De Regt and Dieks 2005, 153).

In the remaining part of this chapter I will elaborate this framework for scientific understanding by focusing on what it means to be able to apply a theory successfully to a phenomenon. First, I will discuss philosophical views on the connection between scientific theories and phenomena. I will subsequently discuss what the conditions are for being able to apply a theory successfully, which – as illustrated in the example of the kinetic theory – include the prerequisite that the theory is intelligible to its user. This framework for scientific understanding will be used in chapters 4 and 5 to analyze the case studies of understanding in psychology. In the concluding chapter, findings from these case studies will be used for a further articulation of the framework and to formulate a substantiated view of the important role of intelligibility and understanding in science.

3.4. *Applying Theories to Phenomena*

An important aspect of scientific understanding of a phenomenon is the ability of scientists to explain the phenomenon by means of the relevant scientific theories. Therefore, a view of scientific theories and how they are connected to phenomena is of importance for an account of scientific understanding. In philosophy of science, the conception of theories has undergone considerable changes since logical positivists developed the traditional syntactic view of theories. I will discuss the transition from this syntactic view to the semantic view of theories, which brought the notion of 'models' into prominence. Since this transition, philosophers of science have generally agreed that the connection between theories and phenomena is accomplished through the use of models. Despite this agreement, the semantic view did not settle the question how models mediate in scientific practice between theories and phenomena. To answer this question, several philosophers pointed at the representational character of models. I will deal with this account of models and focus especially on the representational view developed by Giere (1999b; 2004).

3.4.1. *The Transition from the Syntactic to the Semantic View of Theories*

Models have not always played a prominent role in philosophy of science. Logical positivists, who developed the syntactic view in which a theory is seen as a body of theorems or statements (Van Fraassen 1980, 44), attributed a very minor role to models. For instance, Carnap (1939, 68) argued in relation to physical models that "[i]t is important to realize that the discovery of a model has no more than an aesthetic or didactic or at best a heuristic value, but it is not at all essential for a successful application of the physical theory." Unlike the semantic view, in which – as I will discuss below – models play an important role in the interpretation of theories, in the syntactic view theories are not interpreted by means of models. Instead, in this view the meaning of the theoretical terms in the theorems is specified in other ways, such as by means of correspondence rules or operational definitions. This view faced several difficulties, of which many were related to the division of the theory's vocabulary into observational terms and

theoretical terms (cf. Suppe 1977). In chapter 4 (on neo-behaviorism) I will discuss some of the difficulties that have to do with the meaning of theoretical terms.

The view of theories that was developed – at least partly – in response to these difficulties is the semantic view of theories. In this view, a theory is not characterized by a collection of statements but by a class of models (Van Fraassen 1980, 44). Influential proponents of this view are Patrick C. Suppes (1960; 1967), Frederick Suppe (1977), Van Fraassen (1980), and Giere (1988). In this view, the models of a theory are taken to be those interpretations of the theory on which all the basic assertions of the theory are true (Morrison and Morgan 1999a, 2). The notion of models in this view is derived from logic and meta-mathematics in which models denote composite set-theoretic entities consisting of abstract elements with relations between them (cf. Frigg 2006, 52). In the semantic view, the connection between the models of a theory and the world is established by means of "models of data" (Suppes 1962) that represent certain observable aspects of the world. The construction of these models from experimental data involves methodological procedures, such as the correction of data, the elimination of errors, and curve fitting techniques (cf. Suppes 1962, 261). The relation between models of a theory and models of data is characterized in terms of (partial) isomorphism (e.g. Van Fraassen 1980; Suppes 2002; da Costa and French 2003). For example, Van Fraassen (2000, 181) calls a theory empirically adequate if the data models are embeddable in some of the models of the theory, where he defines embedding as a relation between two models that relies on the isomorphism of the embedded model with parts of the other.

The semantic view can be illustrated by means of Newton's theoretical principles of mechanics and gravitation. One of the models that satisfies these formal principles is the model of planetary motion in which planets are described as mechanical bodies located in Absolute Space in which they have absolute motions. The model is one of the interpretations in which Newton's theoretical principles are true. The data models in this example – which Van Fraassen (1980, 64) calls the "appearances" or the "empirical structures" that can be described in experimental and measurement reports – are composed of empirical data from observations of planetary motions. The construction of

these data models is not elementary because, as Van Fraassen (1980, 45) argues, "it takes thought" to realize that a planet's motion looks like an ellipsis around a moving centre. Instead of absolute motions, the observed motions of the planets are relative motions, for instance relative to the sun or to the earth. In order to demonstrate that these data models are embeddable in his model of planetary motion, Newton argued that relative motions are measures of absolute motions defined with reference to some system of bodies. This enabled him to identify the observed motions with the motions described in the model. According to Van Fraassen, this illustrated the empirical adequacy of Newton's principles of mechanics and gravitation.

3.4.2. *Representational Views of Models Based on Scientific Practice*

The attention paid to models in philosophy of science, which increased as a result of the transition from the syntactic to the semantic view, increased even more due to the turn towards the study of scientific practice caused by the growing appreciation for topics such as theory construction, theory change, and scientific discovery. Philosophers of science became aware that, in scientific practice, scientists make intensive use of models (Bailer-Jones 1999). It can be questioned if the concept of a model in the semantic view, which was inspired by mathematical logic, coincides with the concept of a model used in actual scientific practice. Suppes, a major proponent of the semantic view, claimed in his article "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences" (1960) that these two concepts are compatible. However, subsequent analyses of the use of models in scientific practice have put this claim into question.

For example, in their influential book *Models as Mediators* (Morgan and Morrison 1999), Margaret Morrison and Mary S. Morgan present an account of models based on case studies of actual scientific practice that to some extent is difficult to reconcile with the semantic view. In their view, models "mediate" between theory and data. They are "autonomous agents" that are partially independent of theory and data:

> The crucial feature of partial independence is that models are *not* situated in the middle of an hierarchical structure between theory and the world. Because models typically include other elements, and

> model building proceeds in part independently of theory and data, we construe models as being outside the theory-world axis. It is this feature which enables them to mediate effectively between the two.
> (Morrison and Morgan 1999b, 17–18)

Both the relation between theories and models and the relation between models and data in the semantic view do not do justice to scientific practice. For instance, as Mauricio Suárez (1999, 147) argues, scientists carry out several methods of approximation in the process of constructing models, such as introducing corrections into the theoretical description and simplifying the problem situation. Because of these approximations, scientific models are generally neither truth-makers of scientific theories nor (partial) isomorphic to empirical structures. An example is Newton's model of the solar system in which planets move only as a result of the effects of the sun's gravity. For computational reasons, Newton was not able to take into account the mutual gravitational interactions between the planets, let alone their satellites. This is at variance with his theory of gravitation, in which every massive particle in the universe attracts every other massive particle. Further, in his model, planets are spherical and have a homogenous mass distribution, which is in conflict with the observed data.

One of the problems of the semantic view concerning the relation between models and data has to do with the representational character of models. It is quite generally accepted that, in scientific practice, the relation between scientific models and phenomena is one of representation (cf. Suárez 2003, 225). For instance, distances in the model of the solar system represent the measured distances between the celestial bodies concerned. Morrison and Morgan (1999b, 27) argue that the relation of representation cannot be described by means of the relation of isomorphism because representations abstract from the data and transform it into another form. This irreducibility of scientific representation to isomorphism is elucidated in a study by Suárez (2003) based on examples of successful representation in science. He demonstrates persuasively that isomorphism is neither a necessary nor a sufficient condition for the relation of representation in scientific practice. It is not a necessary condition for representation because isomorphism is symmetric (if A is isomorphic to B, then B is isomorphic to A) whereas, due to its essential directionality, representation is

generally not: the model of the solar system represents the real solar system, but not vice versa. Isomorphism is not a sufficient condition for representation because the relation of representation may fail to obtain even if the relation of isomorphism holds. For instance, consider the trajectory in phase space described by the state vector of a quantum particle. According to Suárez (2003, 236), "Unbeknownst to us this trajectory may well be isomorphic to the motion in physical space of a real classical particle. But unless the phase space model is intended for the particle's motion, the representational relation will fail to obtain." Hence, the relation of representation in science cannot be described by means of isomorphism. The semantic view is not in accordance with the representational character of scientific models. Consequently, accounting for the role of models in scientific practice, which is relevant for the view of scientific understanding in this study, requires reconsidering both the relation between theories and models and that between models and the world.

Giere (1999b; 2004) made a move in that direction with the development of his view of models as representational tools. Although Giere (1988) is often mentioned as a proponent of the semantic view due to his conception of a theory as a "population of models" (Giere 1988, 85), his position has always been slightly different from that of other major proponents of that view, especially because of his different reading of the relation between models and the world. As I will discuss below, instead of describing this relation in terms of (partial) isomorphism, Giere focuses on the pragmatic and cognitive aspects of this relation. In recent years, due to his efforts to develop an illuminating characterization of the connection between models and theories, his position moved even further away from the traditional semantic view. The result is a balanced view that upholds some basic intuitions of the semantic view and does justice to actual scientific practice. Because of these merits, I regard Giere's representational view of models suitable for enhancing my general framework for understanding. Below, I will first discuss Giere's view of the relation between theories and models in science and subsequently discuss Giere's view of the representational relation between models and the world.

3.4.3. *Giere's Representational View of Models*

The difference between the semantic view and Giere's representational view lies primarily in the relation between models and the world, rather than in the relation between theories and models. Giere's view of the relation between theories and models hardly deviates from that in the semantic view. Nevertheless, I find his view on this issue more illuminating than the semantic view due to his choice of terminology. According to Giere, both in scientific practice and in meta-level discussions about the sciences terms such as "theory" and "law" are used quite broadly and even ambiguously. For instance, Newton's laws refer to central principles of classical mechanics, whereas the law of the pendulum refers to an abstract model. Therefore, to facilitate a "sound meta-understanding of scientific practice" it is better to avoid using the terms "law" and "theory" and to use the term "theoretical principle" instead (Giere 2004, 746). The function of theoretical principles is to act as general templates for the construction of models. In this sense they are aids in the construction of models and, as such, fulfill the role that is generally assigned to theories in the semantic view. Examples of theoretical principles are Newton's laws of mechanics, the evolutionary principle of natural selection, and Hull's principles of behavior that will be discussed in chapter 4. Although such principles have often been interpreted as statements that tell us something about the world, Giere sees them as definitions of the abstract objects or terms that compose theoretical models:

> Newton's three laws of motion, for example, refer to quantities called force and mass, and relate these to quantities previously well-understood: position, velocity, and acceleration. But they do not themselves tell us in more specific terms what might count as a force or a mass.
>
> (Giere 2004, 745)

A consequence of this view is that scientific models that make use of theoretical principles (I will refer to these models as theoretical models) satisfy those principles by definition. For instance, in mechanical models Newton's laws of mechanics are satisfied simply because the theoretical entities in these models, such as mechanical bodies, are defined by means of these laws. This view of theoretical principles fits in nicely with the practice of the neo-behaviorists discussed in chapter 4, in which the principles of behavior are used as

definitions of the abstract objects of which the theoretical models of behavior are composed (cf. Hull 1943b).

Speaking of theoretical principles instead of theories (and laws) has the advantage that it removes some of the discrepancies between the view of the relation between theories and models in philosophy and scientific practice. For instance, a drawback of the semantic view is that often its conception of a theory does not correspond with what is considered in scientific practice to be a theory. By not using the term "theory," Giere's account has, in a way, overcome this drawback while holding on to the idea that models satisfy theoretical principles.

The semantic view and Giere's representational view differ significantly in their respective conceptions of the relation between models and phenomena. According to most proponents of the semantic view, this relation should be described in terms of a (partial) isomorphism. Giere's description of this relation is more liberal. It uses a concept of "representing" that, as I have discussed above, cannot be reduced to formal mapping. Somewhat surprisingly, the question of how the relation between models and phenomena can be specified in another way has not received much attention for quite some time (cf. Frigg and Hartmann 2006), although this has changed recently (e.g. Bailer-Jones 2003; Frigg 2006; Giere 2004; Suárez 2004; Van Fraassen 2004). In what follows I will focus on Giere's ideas regarding the relation of representation between scientific models and phenomena in order to develop an account of models and their relation to reality.

Giere (1999b, 44) characterizes models as "tools for *representing* the world." Although these tools seem to be a heterogeneous class that includes physical models, scale models, analogue models, and mathematical models, Giere's characterization encompasses much of this heterogeneity. In his view, all models are designed so that elements of the model refer to features of the real world, which means that models can be used to represent aspects of the world (Giere 2004, 746–747). In philosophy of science it is quite generally accepted that the main purpose of models in science is to use them to represent some aspects or parts of the world, although some argue that their primary function is their use as epistemic tools to, for example, draw inferences (Boon and Knuuttila 2009, 695; Knuuttila and Merz 2009, 147). Giere also admits that scientists use models for all sorts of purposes other than representing the world. However, in his view, "representing the

world is a very important function of models and is often presupposed in discussions of other roles for models" (Giere 2004, 749).

Giere (2004, 742) regards this activity as fundamentally pragmatic, which means that a description of this activity requires reference to the scientists involved. The most important way in which scientists use models to represent aspects of the world is to exploit *similarities* between model and reality. Giere notes that there are no objective rules for picking out *relevant* similarities. Moreover, he argues that the concept of similarity is context-dependent (Giere 1999b, 46) and that there is no objective measure of similarity between the model and the real system:

> Anything is similar to anything else in countless respects, but not anything represents anything else. It is not the model that is doing the representing; it is the scientist using the model who is doing the representing. One way scientists do this is by picking out some specific features of the model that are then claimed to be similar to features of the designated real system to some (perhaps fairly loosely indicated) degree of fit. It is the existence of the specified similarities that makes possible the use of the model to represent the real system in this way.
> (Giere 2004, 747–748)

As an illustration, Giere discusses the analogy between the use of models and the use of maps. A map is a physical object, not a linguistic entity, and therefore it does not make sense to ask if a map is true or false. Like a model, a map is not isomorphic to reality but is a tool that can be used to represent some aspects of the world. This representation is partial and of limited accuracy. What parts of the world are depicted on the map depends on the interest of makers and users of the map (Giere 1999b, 46). Like a model, a map does not represent by itself. Instead, someone can use it to represent aspects of reality. For example, someone who uses a map of a city uses features of the surface of the map to represent features of the surface of the city (Giere 1999b, 45). Seeing the map as similar to the region mapped is context-dependent. For instance, someone who is interested in linear distances will not consider a schematic map, such as a subway map, to be very accurate. However, a subway passenger who is interested in topological features, such as the order of stations on individual lines, will consider the map and the railway system to be quite similar (Giere 1999b, 46–47).

Not only does the selection of the relevant features of scientific models require pragmatic evaluations, but the selection of relevant aspects of the world does as well. The act of making these evaluations is similar to that of the construction of data models in the semantic view, which involves the selection and, to a certain extent, the interpretation of experimental data. Suppes (1962), who introduced the idea of data models, already realized that it raises fundamental questions about the evaluation of empirical data:

> It is precisely the fundamental problem of scientific method to state the principles of scientific methodology that are to be used to answer these questions – questions of measurement, of goodness of fit, of parameter estimation, of identifiability, and the like.
>
> (Suppes 1962, 261)

Giere (1999b, 54) associates the distinction between data and data models with the distinction between data and phenomena put forward by James Bogen and James F. Woodward (Bogen and Woodward 1988; Woodward 1989; 2000). Their view is that theories do not explain data but explain phenomena, such as the melting point of lead or the bias of a coin, the existence of which is inferred from the data. According to Woodward (2009), this inference is guided by empirical assumptions as well as by "evaluative considerations having to do with the investigator's choice of goals, interests, and attitudes." In other words, inferring the existence of phenomena from data involves pragmatic evaluations. For instance, in an investigation of the bias of a coin in which the data are repeated flips of the coin, the significance level that the researcher adopts in a significance test is a result of these evaluative considerations:

> [A] researcher who employs a significance test with a significance level of 0.05 has a different attitude toward the costs of a certain kind of mistake (she adopts a different noise or error level) than a researcher who employs a significance level of 0.001. This might lead the second researcher to reject the hypothesis that the coin is fair in circumstances in which the first researcher does not. (Woodward 2009)

Because of its focus on pragmatic aspects, Giere's account of scientific models resembles Nancy Cartwright's. In the latter view, there are no formal principles for getting from a theory to a description of a real system via a model. Instead, the use of models in science is a non-deductive process that requires evaluations and decisions based

on "rules of thumb" and "good sense" (Cartwright 1983, 133). In my view, her account shows that it is impossible to avoid pragmatic evaluations and decisions in this non-deductive process. Yet, there is a fundamental difference between her view and Giere's. Whereas Giere argues that both the *construction* of models and the *use* of models to *represent* phenomena are pragmatic activities, Cartwright regards only the construction of models as such. In her view, once the connection between scientific theory and the world is accomplished by means of models, *they* – and not the scientists – are "doing" the representing:

> There are only real things and the real ways they behave. And these are represented by models, models constructed with the aid of all knowledge and techniques and tricks and devices we have.
>
> (Cartwright, Shomar, and Suárez 1995, 140)

In contrast, Giere argues that the model does not represent by itself. It is not the case that, once the model is constructed, there is an automatic connection between theory and empirical data. As described in recent literature on models (e.g. Bailer-Jones 2003), such a connection requires an active involvement of the model users. In Giere's (2004, 747) terms, this connection becomes established only if someone is "doing the representing." To neglect this pragmatic aspect of representation, as Cartwright seems to do, may be a relic of the traditional idea that scientific knowledge should not rely on pragmatic considerations. However, the empirical content of scientific theories depends on the connection between theory (or theoretical principles) and empirical data, and this connection requires the active involvement of the scientist. As a consequence, the pragmatic activity of "doing the representing" has epistemic significance.

Both the construction of models and the use of them to represent phenomena involve the pragmatic activities of selecting relevant features of the model and the empirical data and evaluating the possible similarities between these features. These activities are not rule-governed and require scientific judgments. According to Harold I. Brown, the ability to make such judgments, which he describes as "the ability to evaluate a situation, assess evidence, and come to a reasonable decision without following rules" (Brown 1988, 137), is a skill:

> [T]he ability to exercise judgement is a learned ability that is not explicitly rule-governed. This combination is characteristic of skills, and I am maintaining that when we develop the ability to exercise judgement in a particular field, we are developing a skill.
>
> (Brown 1988, 156)

For Giere's representational view, this implies that establishing connections between models and the world is a skillful act that involves the inference of phenomena from empirical data and the use of models to represent these phenomena. In practice, this skillful act will be intertwined with the skillful act of constructing models. In the case studies in chapters 4 and 5, especially in the second one on cognitive psychology, I will analyze the required skills.

The framework for understanding can now be refined. Scientists understand a phenomenon if they are *able* to *apply* a model *successfully* to that phenomenon. I discussed above what it means to apply a model to a concrete phenomenon, namely that it involves the active and skillful process of using the model to represent a phenomenon. Applying a model to a phenomenon involves the skillful activity of selecting relevant features of the model and the empirical data from which the phenomenon is inferred, and the skillful activity of evaluating the relevant similarities between these features. In the following section, I will further elaborate the framework by discussing what the conditions are for being able to apply the model successfully.

3.5. *Intelligibility as an Epistemic Condition for the Successful Application of Models*

The idea that the scientific understanding of phenomena involves the successful application of models fits in nicely with the intuition shared by many philosophers of science that understanding and modeling are related (e.g. Hartmann 1999; Bailer-Jones 1999). Support for this intuition is the famous statement by Lord Kelvin (Thomson 1884/1987, 111) in which he claims that the test of "Do we or do we not understand a particular subject in physics?" is: "Can we make a mechanical model of it?" I share this intuition and suggest generalizing this test by replacing "mechanical models" with "intelligible models": scientists understand phenomena if they can provide intelligible models of

them (cf. De Regt and Dieks 2005, 150; De Regt 2009, 32). In line with De Regt's definition of the intelligibility of a theory, in this study the intelligibility of a model is taken to be the positive value that scientists attribute to the model's virtues that enables them to apply the model successfully to phenomena.

To explore the criteria for the successful application of models, I will briefly discuss two examples of the application of a model. The first example concerns the use of a model developed by Boltzmann to account for the behavior of specific gases such as oxygen or nitrogen. In this so-called "dumbbell model," the gas molecules are considered to be diatomic molecules that can be depicted as two rigidly connected elastic spheres. According to Boltzmann, using this dumbbell model to describe the behavior of gases such as oxygen or nitrogen is successful because it provides an illuminating picture of the behavior of these gases. James C. Maxwell, however, did not agree. In his analysis of this disagreement, De Regt (2005; 2009) points to the decisions concerning approximations and idealizations that are necessary in the process of modeling. An example of an approximation is that, while a rigid dumbbell has five degrees of freedom, experiments on the specific heat ratio of gases such as oxygen or nitrogen yielded values ranging from 4.75 to 4.9. An example of an idealization is that in the model the vibrational degrees of freedom of the gases that explain the emission of spectral lines do not contribute to their specific heats, whereas, according to the kinetic theory of gases, all degrees of freedom contribute to specific heats (De Regt 2009, 36). Unlike Boltzmann, Maxwell was not willing to make these approximations and idealizations. De Regt (2005, 221) argues that this was due to Maxwell's commitment to the mechanistic worldview, which made him a "full-blooded scientific realist." Because of the approximations and idealizations, the dumbbell model could not be seen as a truly faithful representation of reality. According to Maxwell, its idealizations make Boltzmann's model inconsistent with the kinetic theory and with accepted theoretical explanations for the spectral lines of gases. In addition, it can be questioned if Boltzmann's dumbbell model is empirically accurate because, due to the approximation concerning the degrees of freedom, the specific heat ratio of the model merely approximates the measured specific heat ration in experiments on the real gases. Further, the emission of spectral lines is neglected in the model, whereas experiments show

that the real gases do emit spectral lines. Therefore, Maxwell did not consider the application of this model to be successful.

The example shows that empirical accuracy and consistency with accepted scientific knowledge are requirements for the successful application of a model, both of which are items on Kuhn's list of epistemic values discussed in chapter 2. The difference in Maxwell's and Boltzmann's assessment of the consistency and empirical accuracy of the application of the dumbbell model can be explained by Kuhn's (1977, 322, 331) idea that the evaluation of consistency and empirical accuracy is a matter of value judgment and that, due to contextual factors, individuals may differ about the application of epistemic values (Kuhn 1973/1977, 322). In this example Maxwell's assessment differs from Boltzmann's as a result of his commitment to the mechanistic worldview.

In my view, the observation that scientists do not assign the value of intelligibility to a model if, in their assessment, the model is not in accordance with epistemic values like consistency and empirical accuracy is important for my claim in chapter 2 that intelligibility, like empirical accuracy, consistency, scope, simplicity, and fruitfulness, is one of the epistemic values of science. I suggest that there is interdependency between epistemic values: the value of intelligibility depends on the values of consistency and empirical accuracy (in the sense that only consistent and empirically accurate models can be assigned the value of intelligibility). In chapter 4, when I discuss the debate among neo-behaviorists and logical positivists in the 1950s, I will argue that the value of intelligibility and the value of fruitfulness are interdependent. Intelligible models have to be consistent, empirically accurate, and fruitful. It may very well be that the epistemic value of intelligibility is also related to other epistemic values mentioned by Kuhn. For instance, the value of scope concerns the unifying power of a model, which is exactly what, according to the unificationist account of explanation, can be associated with its intelligibility. This would mean that the epistemic values of science are united in the value of intelligibility. Scientists would not consider a model to be intelligible if they did not consider that model to be in accordance with the epistemic values of science.

The second example concerns connectionist models of cognitive skills such as face and word recognition. This example shows that the

value of intelligibility is not exhausted by the other epistemic values. The connectionist models are neural network models consisting of a network of units connected by weighted links. "Activity is passed around the network in a parallel … manner as some function of the current activity of a unit and the weights on the links from it to other units. Thus the activity of the individual units changes over time, and the weights of the links may also be modified over time" (Partridge 1991, 63). Connectionist models appear to conform to Kuhn's epistemic values. For instance, they are empirically accurate because, according to the assessments of cognitive scientists, the skills of neural networks resemble cognitive skills such as face and word recognition. The models are also fruitful because they find wide application in artificial intelligence. Nevertheless, cognitive scientists report that they do not consider the neural network models to be intelligible. They experience difficulties in understanding how the network performs tasks such as face and word recognition (e.g. Partridge 1987; 1991; Flexer 1995). For instance, Derek Partridge (1991, 17) argues that, although it is straightforward to visualize the basic building blocks for neural networks and their mutual connections, "piecing them together to explain observed behaviours represents a formidable problem." Because the primary mechanism for processing the network is a parallel transfer of activity values, it is hard to reason about the effects of certain inputs on the weights of the links and on the output. "Why a certain model is doing what it is observed to be doing, and what would be needed to make it do something differently, are both extremely difficult questions to answer" (Partridge 1991, 72). This difficulty, which is regarded as a limitation of connectionist models, has been termed the "explanation problem" (Partridge 1987). Partridge (1991, 17) compares the difficulty in reasoning about the neural network model with the difficulty in explaining the details of a weather pattern in terms of intermolecular interactions rather than in terms of temperature, pressure, prevailing winds, etc. The explanation of weather patterns in the latter terms is more comprehensible. Interestingly, in a similar vein De Regt and Dieks argue that an understanding of weather patterns requires intelligible meteorological theories and models:

> Weather predictions are obtained by means of computer calculations in which the Navier-Stokes equations are solved for very large systems, using many auxiliary theories to incorporate small-scale effects. If

> meteorologists merely were occupied with making correct predictions in this manner, they would fail to understand the weather. But this is not the case: meteorologists are concerned not only with 'brute force' computer calculations but also with formulating intelligible meteorological theories and models.
>
> (De Regt and Dieks 2005, 153)

In the same way, according to Partridge, the behavior of a neural network can only be understood if it can be explained "in psychologically meaningful terms, rather than in terms of the implementation structures." He argues that it is incomprehensible how neural networks perform their tasks because they do not allow for "reasoning by analogy with one's own supposed thought processes" (Partridge 1991, 17).

The examples illustrate that meeting Kuhn's epistemic values of science is not a sufficient condition for models to be intelligible. In addition, the models should allow for reasoning about them in a way that meets the cognitive requirements of the scientists. In my view, this aspect of the intelligibility of a model fits in nicely with the account of intelligibility developed by De Regt (2004, 103; 2009, 31), who argues that the virtues of the models should match the (cognitive) skills of the model user. Typical examples of the skills mentioned by De Regt are visualization and causal reasoning, which, because they enable the recognition of qualitatively characteristic consequences of the model without performing exact calculations (De Regt 2009, 33), are both skills that are useful for reasoning about the model. In his view, scientists attribute the value of intelligibility to a model if it has virtues that facilitate the use of these cognitive skills.

In sum, my discussion of Giere's representational view of models and De Regt's view of scientific understanding gives an idea of what conditions must obtain if a model is to be intelligible to scientists. First, the *scientists* have to possess the skills to recognize phenomena in raw empirical data, to recognize relevant similarities between models and phenomena, and to recognize qualitatively characteristic consequences of the model. Second, the *model* has to possess virtues that match the skills of the scientists such that the combination of skills and virtues facilitates the active and skillful process of using the model to represent a phenomenon and reason about it. In the case studies in chapters 4 and 5, I will investigate the required virtues of the model

and skills of the scientists in more detail by analyzing specific instances of skills and virtues that enable the successful application of models to phenomena.

3.6. *Key Notions and Key Questions*

The philosophical framework developed in this chapter involves key notions such as the intelligibility of models, the skills of scientists, and the virtues of models. This framework serves as a point of departure for the analysis of the historical case studies in chapters 4 and 5 about neo-behaviorism and cognitive psychology. In turn, this analysis provides material for the development of a more articulated account of scientific understanding. In chapter 2 I discussed this process of going back and forth between the top-down and bottom-up approaches. On the basis of this iterative process the case studies are analyzed and the philosophical framework is articulated. In chapter 6, the concluding chapter, I will present the results of this process by answering the following key questions concerning the key notions of this study:

- Is the *intelligibility* of models an *epistemic value*, and how does it function in scientific practice?
- What kinds of *skills* are required for the successful application of a scientific model to a phenomenon?
- Which kind of *virtues* can render a model intelligible to its users?
- On what kind of *pragmatic and contextual factors* does intelligibility depend?
- Is the characterization of science advocated in this study useful for the *explanatory and normative tasks* of philosophy of science?

The answers to these key questions will provide a picture of scientific understanding that is useful for the philosophical analysis of scientific practices and thereby contributes to the philosophical characterization of science.

CHAPTER 4

The Virtue of Surplus Meaning: Neo-Behaviorism

4.1. *Introduction*

Since Kuhn (1973/1977, 322) characterized science by means of a list of epistemic values that provide "*the* shared basis for theory choice," there is a debate in philosophy of science about what the epistemic values of science are (e.g. McMullin 1983; Longino 1990; Lacey 2005). Kuhn's list comprised such values as empirical accuracy, consistency, scope, simplicity, and fruitfulness. In chapter 2 I suggested that the lists of epistemic values proposed by philosophers such as Kuhn, McMullin, Longino, and Lacey lack a highly important element, namely *intelligibility*. In this chapter I will provide evidence for this suggestion by means of an empirical case study. I will show that the intelligibility of models has epistemic significance.

My argument is based on a case study of neo-behaviorism. This case is interesting for my purposes not only because of the huge influence of neo-behaviorism on psychology in particular and on science in general but also on account of its extraordinary positivist inclinations. Between 1930 and 1960, when psychology grew enormously as an academic enterprise and neo-behaviorism was the leading approach in American psychology, neo-behaviorists advocated a view of science that was heavily influenced by logical positivism. It included, among other things, the pursuit of objectivity in science, which was meant to prevent the results of scientific endeavor from being prejudiced as a result of metaphysical considerations or subjective influences. Because of their positivist attitude, neo-behaviorists were sometimes called "black-box psychologists." They viewed organisms as black boxes with observable stimuli as input and observable responses as output and rejected any allegedly metaphysical speculation about mechanisms inside these black boxes. Instead of "opening" the black boxes to acquire "insight," they declared that the aim of psychology is to find functional relationships that describe correlations between stimuli and

responses. This positivist tendency makes neo-behaviorism highly relevant to the study of intelligibility and understanding. Because of the alleged subjective connotations of understanding and intelligibility, neo-behaviorists did not consider the claim that intelligibility is an epistemic value of science to be consonant with their view that science should be objective. I will examine the scientific practice of neo-behaviorism and show that, despite their positivist ideas, even neo-behaviorists aimed at insight. They implicitly embraced intelligibility as an epistemic value. What might seem to be one of the strongest counterexamples to my claim proves in fact to corroborate it and, as such, provides a strong case for my claims concerning intelligibility and understanding.

Section 4.2 will start with an overview of the methodological tenets of neo-behaviorism, where I will concentrate especially on the meaning and use of theoretical terms in neo-behaviorist psychology. Subsequently, I will focus on two leading figures of neo-behaviorism who made major contributions to methodology in this field, namely Tolman (1886–1959) and Hull (1884–1952). I will show that in their psychological models of behavior these neo-behaviorists used theoretical terms whose meaning transcended their objective definitions. I will demonstrate that this "surplus meaning" – a notion derived from Reichenbach (1938) – has epistemic significance because it renders the models of behavior intelligible to their users. In section 4.3 I will use my notion of intelligibility to offer an analysis of a vivid debate about surplus meaning among theoretical psychologists and logical positivists, including Tolman (1949), Melvin H. Marx (1951), John R. Maze (1954), Gustav Bergmann (1953), David Krech (1950), and Gardner Lindzey (1953) that was initiated by Kenneth MacCorquodale and Paul E. Meehl (1948). In section 4.4 I will conclude that – even in neo-behaviorism – the intelligibility of models is an epistemic value of science.

4.2. *The Meaning and Use of Theoretical Terms in Neo-Behaviorism*

According to the reconstruction that neo-behaviorists gave of the history of their discipline, classical behaviorism, the precursor of neo-behaviorism, was founded at the beginning of the 20th century. It was a reaction to subjectivism and the generally accepted "introspective"

method practiced by major psychologists like Wilhelm M. Wundt and Edward B. Titchener. The American psychologist John B. Watson (1878–1958) is usually given the credit for founding this school in 1913 with the publication of a very influential manifest called "Psychology as the Behaviorist Views It." He rejected the use of introspection as a basic procedure and instead advocated the use of objective methods in psychology.

Because self-observed conscious inner thoughts, desires, and sensations might be useful for understanding the behavior of others – even of rats, as my analysis of Tolman's work will illustrate – in my view, the rejection of introspection is an example of the attempt to exclude understanding from the aims of psychology. However, the historical case in this study will show that there are other ways to obtain understanding of psychological phenomena. Therefore, Watson's rejection of introspection does not completely shut the door to understanding in psychology.

The basic ideas of classical behaviorism can already be found in the first lines of his manuscript. First, the (almost-exclusive) subject matter of psychology is overt behavior; second, the major goals of psychology are the prediction and control of behavior instead of the description and explanation of states of consciousness; and third, the study of animal behavior falls within the domain of psychology:

> Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness. The behaviorist, in his efforts to get a unitary scheme of animal response, recognizes no dividing line between man and brute. The behavior of man, with all of its refinement and complexity, forms only a part of the behaviorist's total scheme of investigation.
>
> (Watson 1913, 158)

These basic ideas have their historical roots in the functionalist psychology at Columbia University and the school of objective psychology in Russia. From the functionalist school classical behaviorism adopted the focus on the adaptive capabilities of organisms to their environment, the focus on learning, and the typical use of the terms "stimulus" and "response." An influential proponent of this school

was Edward L. Thorndike (1874–1949) who is known for his animal experiments on the formation of associations. He formulated laws of association, such as the *law of effect*, which states that responses to a situation that are followed by a rewarding state of affairs become associated with the situation and are hence more likely to recur when that situation is subsequently encountered. Thorndike's associationism made learning an important topic in psychology and was a source of inspiration for classical behaviorism (Hergenhahn 1997, 330).

From Russian objective psychology classical behaviorism adopted ideas about objective methodology in psychology and about reflexes between stimulus and response. At the end of the 19th century, Russian objective psychologists, such as Ivan M. Sechenov and Ivan P. Pavlov, argued that psychology should be a "positive" science. These psychologists worked in the tradition of the positivist physiologists in Berlin, such as Emil H. Du Bois-Reymond, Hermann L.F. von Helmholtz, and Ernst W. von Brücke, who tried to reduce physiology to applied physics and chemistry. At the beginning of the 20th century, Pavlov's animal experiments on digestion became known in the West, particularly through the writings of Watson. An important concept in Pavlov's work was the conditioned reflex, which would become a key concept in learning theories of classical behaviorism (Hergenhahn 1997, 344).

It is quite generally believed that in the 1930s, the influence of the logical-positivist movement in America on classical behaviorism gave rise to the neo-behaviorist school in psychology. The logical-positivist movement in the United States of America differed in some respects from its European counterpart (Sanders 1972, 129–131). For instance, instead of Carnap's suggestion that correspondence rules be used for the definition of theoretical terms (e.g. Carnap 1956, 52–59), in the United States the dominant view on this topic was Percy W. Bridgman's operationism, in which a theoretical notion is defined as synonymous to a corresponding operation. Among the first to mention the relation between logical positivism and neo-behaviorism were Sigmund Koch (1964, 10) and Brian D. Mackenzie (1977, 149). Their account of the history of neo-behaviorism was widely accepted (e.g. Leahey 1980a; Sanders 1972; Hempel 1969), and has been incorporated in major textbooks on the history of psychology, where it

is stressed that "logical positivism had a powerful influence on psychology" and that "neo-behaviorism resulted when behaviorism was combined with logical positivism" (Hergenhahn 1997, 377). Laurence D. Smith (1986) has questioned this textbook history and argued that logical positivism was not the primary influence on the philosophical views of early neo-behaviorists such as Tolman and Hull. Instead, Smith claims, their views were rooted in the naturalistic approach of Clarence I. Lewis, Ralph B. Perry, Edwin B. Holt, John Dewey, and others. Although Smith probably has a point that it was not the philosophy of logical positivism that shaped these views, it certainly reinforced them (cf. Kitchener 2004). There are several documented cases of close interactions between both schools, for instance, at the University of Iowa where positivist philosophers of science and neo-behaviorist psychologists founded a center for neo-behaviorist thought.

4.2.1. *Operational Definitions and the Meaning of Theoretical Terms*

Although logical positivism offered the neo-behaviorists a philosophical foundation for their objectivist views, it was not evident how this body of thought could be translated into recommendations for scientific practice. Despite the affinity between neo-behaviorists and logical positivists, they had very different interests. The positivists wanted to clarify the formal relationship between theory and data, and they did so in the tradition of the analysis of language. The neo-behaviorists wanted to formulate practical methodologies. However, there was also an overlap, namely their pursuit of objectivity. Leading figures of neo-behaviorism, such as Tolman and Hull, initiated a behaviorist methodology that was meant to give shape to an objective psychology. A major ingredient of this methodology was operationism, which is a doctrine commonly associated with the physicist Bridgman, who claimed that a concept can best be defined in terms of a set of operations:

> We evidently know what we mean by length if we can tell what the length of any and every object is, and for the physicist nothing more is required. To find the length of an object, we have to perform certain physical operations. The concept of length is therefore fixed when the operations by which length is measured are fixed: that is, the

> concept of length involves as much as and nothing more than the set of operations by which length is determined. In general, we mean by any concept nothing more than a set of operations; *the concept is synonymous with the corresponding set of operations*.
>
> (Bridgman 1927, 5)

In an attempt to eliminate possibly non-objective connotations that, as I will discuss below, originated for instance from pre-scientific understanding, neo-behaviorists provided the theoretical terms in psychology with operational definitions. For example, someone who uses theoretical concepts such as 'demand' or 'hunger' in an account of the behavior of rats might be inclined to supply these terms with subjective connotations based on personal experiences. By operationally defining these concepts in terms of 'time since feeding,' neo-behaviorists such as Tolman and Hull tried to remove these connotations but, as I will show, did not succeed.

The doctrine of operationism is in the same spirit as the logical-positivist project of reducing the meaning of statements containing theoretical terms to methods of measurement. The project of the logical positivists faced the difficulty that not all allegedly meaningful theoretical statements in science can be defined exhaustively in terms of operations and resulting observations (cf. Feest 2005, 134). Similarly, operationism's doctrine that theoretical terms should be completely defined in terms of operations and resulting observations was problematic.

A major problem for operationism was that, in scientific practice, it is quite common that several measurement operations correspond to the same concept. For example, the measuring operation that involves the use of measuring rods is not the only operation that corresponds to the concept of length, because "[i]f we want to be able to measure the length of bodies moving with higher velocities such as we find existing in nature (stars or cathode particles), we must adopt another definition and other operations for measuring length" (Bridgman 1927, 11). The different operations for measuring length enable the application of this concept in different domains. The domain of a concept consists of the kind of entities to which the concept applies (Radder 2006, 94). Because the operation involving the use of measuring rods is not applicable in the domain of stars and cathode particles, the application of the concept in that domain involves other operations. This is not

in line with Bridgman's original intentions; he argued that "[i]n *principle* the operations by which length is measured should be *uniquely* specified. If we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations" (Bridgman 1927, 10). The common practice of connecting different operations to the same concept therefore raised the difficulty of how to justify the application of the same concept in different domains.

Another example is the application of the concept of 'reinforcement' in different domains. As I will show below, in different situations neo-behaviorists related different operations to this concept. In one experimental situation, 'reinforcement' was defined operationally as the result of repeatedly performing the operation of administering a mild electric shock to human subjects immediately after presenting a noise. In another experimental situation, 'reinforcement' was defined as the result of repeatedly performing the operation of giving food to a rat if it depressed a bar. Justifying that both operations are measurements of 'reinforcement' in different domains is problematic because it would require a conception of 'reinforcement' that transcends these operational definitions. This would imply that the operations do not capture the complete meaning of 'reinforcement.'

Related to this difficulty is the critique that operational definitions do not function as complete definitions of concepts. For instance, neo-behaviorists applied the concept of 'reinforcement,' which in the example is defined operationally in two specific domains, as well in other domains by adding operations to the definition of this concept. This demonstrates that, in general, operational definitions are partial or "open" (Zuriff 1985, 60). Whereas this well-known critique of operationism rejects the idea that operational definitions can capture the complete meaning of concepts, Radder (2006, 127) goes one step further by criticizing the static conception of the *complete* meaning of concepts. With its attempt at capturing the complete meaning of concepts, operationism fails to acknowledge the extensibility of concepts to novel situations. According to Radder (2006, 103) a concept is extensible if it "is successfully applied to a certain domain and ... might be used in one or more other domains." Extensible concepts transcend the "local" meaning they have in concrete situations. In the example,

'reinforcement' is an extensible concept because it may transcend the operational definitions it has in the two experimental settings mentioned.

Using a notion that MacCorquodale and Meehl (1948) adopted from Reichenbach (1938), I will call the aspects of the meaning of theoretical terms that are not captured by its (operational) definition the "surplus meaning" of these terms. This surplus meaning partly determines possible extensions of the application of the term to new domains. I will not only show that – in spite of their efforts – the meaning of the theoretical terms that neo-behaviorists like Tolman and Hull used in their models of animal and human behavior unavoidably transcended the objective definitions of these terms. I will also show that the surplus meaning of the concepts has epistemic significance. By means of the account of understanding presented in chapter 3, I will analyze the function of this surplus meaning and conclude that these models are intelligible due to the surplus meaning of their theoretical terms.

4.2.2. *Edward C. Tolman and the Intervening Variable*

Edward Chase Tolman (1886–1959) studied electrical engineering at the Massachusetts Institute of Technology, but after reading William James' *Principles of Psychology* in his final year, he switched to psychology and philosophy at Harvard where he received his Ph.D. in 1915. His teachers Perry and Holt, who had both been students of James, introduced him to the functionalist school of psychology. During his study, the influence of Watson's classical behaviorism was growing, and Tolman became acquainted with it. His teachers tried to translate James' functionalist terminology into behaviorist jargon. Tolman came under the spell of the behaviorist ideas as well, and in 1918, when he started experimenting with rats in mazes at the University of California at Berkeley, he started calling himself a behaviorist.

Since Tolman expressed a great deal of interest in various schools of psychology, it has been a matter of debate whether Tolman really was a behaviorist (Innis 1999, 115). He stayed loyal to the functionalist roots of the ideas he had acquired from his teachers by assigning a prominent role in his behaviorist theories to the concepts of 'purpose' and 'cognition,' albeit under different names. He regarded 'purposes'

to be the motivating forces – or 'drives' – for behavior and 'cognitions' to concern 'hypotheses' about how environmental features can be used or manipulated in order to attain certain goals (Feest 2005, 141). Furthermore, he was interested in Gestalt psychology and became sympathetic to mentalists such as William McDougall, who argued that behavior was generally goal-oriented and purposive (Still 1997a, 576). The incorporation of terms from other schools, such as 'purpose' and 'cognition,' which could be seen as contravening the spirit of behaviorism, led to a twofold problem for Tolman. He had to convince fellow behaviorists on the one hand that his terms were scientific, and mentalists such as McDougall on the other that his behaviorist psychology was able to do justice to the full complexity of behavior (Still 1997a, 577). The solution to this problem, which he proposed in his magnum opus *Purposive Behavior in Animals and Men* (1932), was the introduction of a special category of theoretical terms in behaviorism, namely "intervening variables." According to Tolman (1932, 414), cognitive and purposive events, or mental processes in general, had to be treated as "intervening variables," that is, theoretical terms whose meaning had to "be inferred 'back' from behavior." His view of concepts such as 'purpose' and 'cognition,' which was a realist view initially, had shifted to a constructivist view in which these concepts were to be seen as constructs introduced for pragmatic reasons:

> They are to behavior as electrons, waves, or whatever it may be, are to the happenings in inorganic matter. They are pragmatically conceived, objective variables the concepts of which can be altered and changed as proves most useful. (Tolman 1932, 414)

Tolman argued that concepts such as 'purpose' and 'cognition' had to be introduced in an objective way. For Tolman, it was not immediately clear how this should be done. It is likely that his decision to contact the members of the Vienna Circle was motivated by the hope that logical positivism would provide a solution to this problem (Smith 1986, 130).

In 1933, Tolman spent a sabbatical of seven months in Vienna. He became acquainted with logical positivists and psychologists who were influenced by positivist ideas, such as Egon Brunswik (1903–1955). After his sabbatical Tolman kept in contact with the logical positivists and was invited to speak on several occasions – for example,

at conferences of the *Unity of Science Movement*. Furthermore, Otto Neurath invited him to write an article in the *International Encyclopedia of Unified Science*, an ambitious and never completed project by logical positivists in the Vienna Circle. Although Tolman did not contribute to this project, his name appeared on the membership list of the advisory committee and the organizing committee for the related international congress that was held at Harvard in 1939 (Smith 1986, 128). During his stay in Vienna, Tolman reflected on the distinction between immediate experience on the one hand and "logical constructs" of scientific theories on the other. He tried to incorporate mental terms into behaviorist psychology by introducing them as intervening variables that were formulated in a positivist fashion as a set of functional equations. This is comparable to Carnap's attempts at that time to formalize the concepts of Sigmund Freud's psychoanalysis (Smith 1986, 116–117). Whether or not Tolman came into contact with Carnap during his trip in Vienna, his statements on intervening variables were clearly reminiscent of Carnap's approach (cf. Carnap 1963, 58). After his sabbatical in Vienna, Tolman expressed the intention to apply the ideas of logical positivism to psychology:

> The "logical positivists," that is such men as Wittgenstein, Schlick, Carnap in Europe and Bridgman, C.I. Lewis, Feigl and Blumberg in this country, have already done the task for physics. But no one as yet, so it seems to me, has satisfactorily done it for psychology.
>
> (Tolman 1935/1966, 100)

Tolman developed a very influential research program in psychology that he called *operational behaviorism*. Its main assertions were inspired by Watson's manifest on behaviorism, by logical-positivist ideas on objectivity, and by Bridgman's operationism:

> It asserts that the ultimate interest of psychology is solely the prediction and control of behavior. . . . It asserts that psychological concepts, i.e., the mental capacities and the mental events – may be conceived as objectively defined intervening variables. And it asserts that these intervening variables are to be defined wholly operationally – that is, in terms of the actual experimental operations whereby their presences or absences and their relations to the controlling independent variables and to the final dependent behavior are determined.
>
> (Tolman 1936/1966, 129)

Tolman was elected president of the American Psychological Association in 1937, and presented his research program in his inaugural speech (Tolman 1938/1966). In the logical-positivist spirit he defined the aim of psychology as finding functional relationships between the behavior of an organism and its environmental "determiners," labeled by him as "dependent" and "independent variables" respectively. He schematized this as follows (Tolman 1938/1966, 151):

INDEPENDENT VARIABLES —— f_1 —— DEPENDENT VARIABLE

For complex behavior, finding f_1 is a difficult task, and Tolman therefore proposed to break up this complicated function into more manageable component functions, f_2 and f_3, which led to the introduction of theoretical terms or "intervening variables." Tolman (1938/1966, 153) schematized the new situation as follows:

INDEPENDENT VARIABLES — f_2 — INTERVENING VARIABLES — f_3 — DEPENDENT VARIABLE

A classical example of an intervening variable is a theoretical term labeled 'demand' or 'hunger,' which is introduced to relate an independent experimental variable (e.g. the time since the organism last received food) with a certain dependent variable (e.g. the rate of food-reinforced lever pressing). The introduction of intervening variables did not conflict with the positivist ideals of neo-behaviorism as long as the meaning of these terms was captured in an objective way by means of operational definitions. From that point of view, an intervening variable is nothing more than a label used in an observed functional relationship between behavior and environment. It does not stand for an entity, event, or process occurring in an unobserved region in an organism's body or mind, and it should not to be reified or assigned a causal role (Zuriff 1985, 67).

In accordance with the second scheme, Tolman's methodology can be divided into three steps: first, picking out the appropriate independent, dependent, and intervening variables; second, formulating the f_2 functions between the independent variables and the intervening variables by operationally defining the intervening variables; and third, formulating the f_3 functions between the intervening variables and the dependent variables. I will investigate Tolman's methodology by looking at a concrete example about rats in mazes that Tolman supplied in order to exemplify his methodology. By means of this example I will

examine the three steps in his methodology. It will turn out that in practice Tolman required more than his methodological ideas alone. At each step he relied on his pre-scientific understanding of rats, which he used as a heuristic guide in carrying out his methodology.

As an illustration of his methodology, Tolman (1938/1966) analyzed the behavior of rats in T-mazes. A T-maze consists of a very simple maze with only one choice point where, one by one, the rats choose between an alley to the left and an alley to the right. Only if a rat enters the alley to the right will it find food (see figure 4.2.2.a).

As a first step, I will analyze how Tolman picked out the appropriate independent, dependent, and intervening variables. In this example, he simply proposed looking at certain variables. As a dependent variable, he proposed investigating the "left-turning tendency" of the rats, which he defined as the percentage of the rats that enter the left alley. He gave no reason as to why he considered this aspect of behavior worthy of note, as opposed to other aspects, such as whether the rats moved quickly or slowly. One may wonder why he decided to measure the tendency of the rats to enter the left alley, that is, the alley where there was no food. Because it is reasonable to expect that the presence of food in the right alley is a relevant factor for the behavior of the rats, it seems more sensible to measure the tendency of the rats to enter that alley. However, if choices are guided by such considerations, then they might appear biased. Therefore, it might be that his

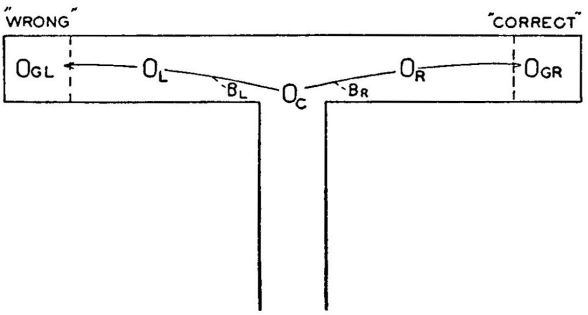

Figure 4.2.2.a. A T-maze with single choice point; O_C = *the point of choice;* $O_{L/R}$ = *the complex of stimulus objects met going down the left/right alley;* $O_{GL/R}$ = *the goal at the left/right;* $B_{L/R}$ = *the left/right-turning behavior*
(Tolman 1938/1966, 150)

intention with his choice of dependent variable was to avoid giving the impression that he was prejudiced. As independent variables, Tolman proposed the number of previous trials in the maze performed by the rats and the period of food deprivation. He gave no reasons for this choice either. Apparently, he considered this choice of variables to be obvious. That it was based on a pre-scientific understanding of the behavior of rats becomes visible from his choice of intervening variables.

In Tolman's methodology, intervening variables are introduced in the process of breaking down the function between the independent and dependent variables into component functions. In his example, Tolman introduced the intervening variable 'demand,' which intervenes between the period of food deprivation and the percentage of the rats that enter the left alley, as well as the intervening variable 'hypotheses,' which intervenes between the number of trials and the percentage of rats that enter the left alley. Although intervening variables were not supposed to represent entities or processes, it is obvious that this way of breaking down the function is based on the beliefs that the rats get hungry if they are deprived of food for a long time, that the rats develop hypotheses during the trials in the maze about where food can be found, and that hungry rats will use these hypotheses to find food. Apparently, Tolman used his pre-scientific understanding of the behavior of rats as a heuristic guide in his choice of dependent, independent, and intervening variables.

As a second step, I will analyze the formulation of the f_2 functions between the independent variables and the intervening variables, which was a central element in Tolman's methodology of operational behaviorism. The f_2 functions are the operational definitions of the intervening variables, and Tolman generated these definitions by means of what he called "standard experiments." The rationale behind the use of standard experiments was the idea that following a standard procedure for the formulation of the definitions of the intervening variables ensured the objectivity of these definitions. The underlying idea in traditional philosophy of science, that scientific endeavor should be a rule-governed activity following *the* methodological rules of science, has been criticized ever since Kuhn (1962) and Michael Polanyi (1967). For instance, Polanyi (1967, 4) argues that exploratory acts involve "tacit knowledge" that cannot be stated in

methodological rules, such as informed guesses, hunches, and imaginings. I want to reinforce this point and show that the skills that are required for such activities are related to scientific understanding.

In his standard experiments, Tolman measured the functional relation between a specific independent and a specific dependent variable while keeping the other independent variables at a "standard" value. Subsequently, he used these functional relations to formulate the operational definitions for the intervening variables he introduced. That the definitions that Tolman provided are not only based in practice on methodological rules follows already from their dependence on "standard values." These standard values are not specified in Tolman's methodology. In most experimental situations this is not a problem because the choice of these standard values is rather obvious. For instance, the rats used for training experiments in mazes should not be completely underfed or overstuffed. Determining the right (standard) level of nourishment is a judgment that is not entirely rule-governed.

In this example the operational definition of the intervening variable 'demand' was based on measurements of the relation between the period of food deprivation and the percentage of rats that entered the left alley, where the amount of training of all the rats was set to a standard value. The result of this standard experiment can be seen in figure 4.2.2.b.

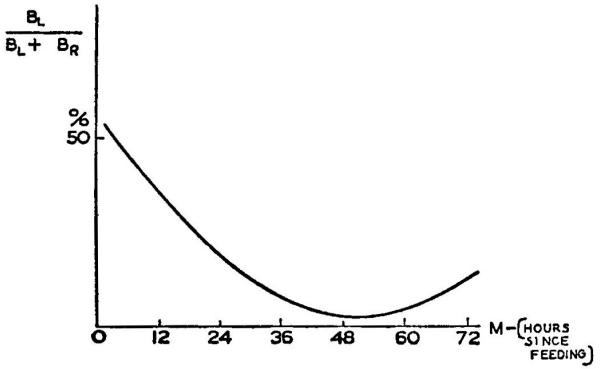

Figure 4.2.2.b. Percentage of rats entering the left alley as a function of the period of food deprivation (Tolman 1938/1966, 158)

The number of rats that entered the left alley decreased (and later increased) in relation to the increase in the number of hours passed since feeding. To turn this figure into a definition of the intervening variable 'demand,' Tolman (1938/1966, 158) argued, one modification was necessary, namely that 'demand' should really be defined as inversely related to this ratio. More precisely, the function that defines 'demand' operationally should be defined by the function that results from the operation of reflecting the original curve about the horizontal axis and translating it such that the curve starts at the origin of the coordinate system (see figure 4.2.2.c).

The inversion and translation of the graph, which is not rule-governed, indicates that the methodological rules of Tolman's operational behaviorism are not sufficient for formulating operational definitions. Apparently, Tolman regarded it to be essential that 'demand' increase with the time since feeding – at least for the first 50 hours. In addition, Tolman presumably regarded the decrease of 'demand' after 50 hours as an unwanted feature of this intervening variable. Although he did not comment on it, he did decide to stop the experiment after about 80 hours, which suggests that he regarded this last period to be irrelevant or uninteresting. It is plausible that he based this judgment on his pre-scientific understanding that, after about three days of food deprivation, the rats are so faint with hunger that their choice for the left or

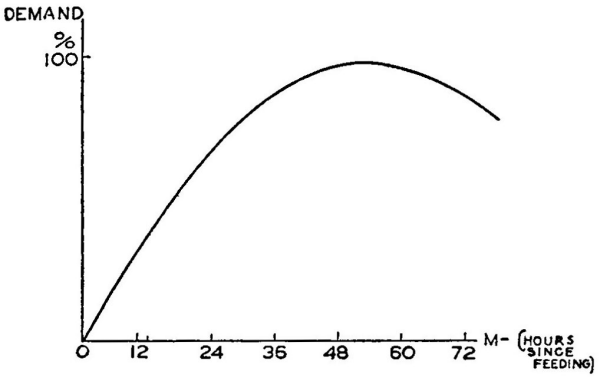

Figure 4.2.2.c. Demand as a function of the period of food deprivation (Tolman 1938/1966, 159)

right alley does not truly reflect their demand for food. Again, this is an example of a decision that is not rule-governed. Another is his choice to translate the curve such that the 'demand' varies between 0% (no hunger) and 100% (the maximum level of hunger). It shows that Tolman assumes that the demand is zero immediately after the rats are fed.

The notable act of transforming graphs without methodological motivations is not exceptional. Similar transformations can be found in the case of other intervening variables. For the formulation of the operational definition of the intervening variable 'hypotheses,' Tolman measured the function between the number of previous trials made by hungry rats and the percentage of the rats that entered the left alley, and it appeared that with the increase in the number of previous trials, the number of rats that entered the left alley decreased (see figure 4.2.2.d).

Tolman again transformed this function into the definition of 'hypotheses' by inverting and translating it. The result was, as he argued, "no more than our old friend, the learning curve" (Tolman 1938/1966, 145), which is a frequently used function in neo-behaviorism (see e.g. figure 4.2.3.a in section 4.2.3). In this way he provided 'hypotheses' with an objective, operational definition in terms of the number of

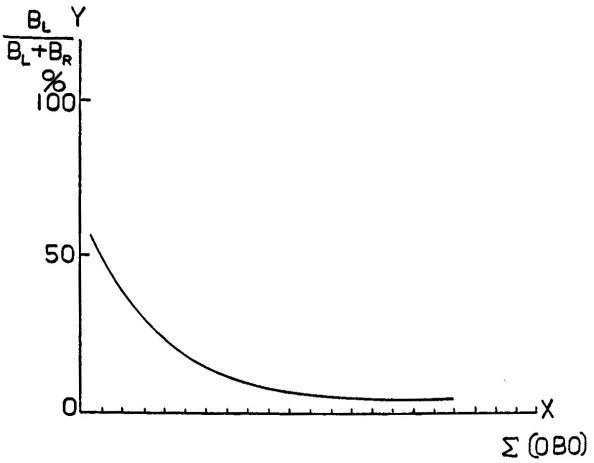

Figure 4.2.2.d. Percentage of rats entering the left alley as a function of the number of previous trials (Tolman 1938/1966, 148)

previous trials. Although it is clear that the transformations of the graphs that were performed to obtain the definitions of 'demand' and 'hypotheses' were based on Tolman's pre-understanding of the behavior of rats, he did not indicate any motivation for performing these transformations. As will become clear when I consider the next step, which concerns the formulation of the f_3 function, his reason for performing these transformations was that it resulted in a model of the behavior of rats that was intelligible to him. Instead of ensuring that the meaning of the theoretical terms in his model did not exceed their operational definition (as he ought to have done according to the philosophical doctrine he preached), Tolman in fact endowed the terms with surplus meaning.

As a third step, I will analyze the formulation of the f_3 function, about which Tolman (1938/1966, 160) wrote that "[i]t is by means of this f_3 function (if we but knew what it was) that we would be able to predict the final outcome for all possible values of the intervening variables." Finding the f_3 function is easier said than done. Tolman (1938/1966, 160) confessed that this is the feature of his doctrine about which he was "haziest." He proposed following a strategy of anthropomorphism:

> I am at present being openly and consciously just as anthropomorphic about it as I please. For, to be anthropomorphic is, as I see it, merely to cast one's concepts into a mold such that one can derive useful preliminary hunches from one's own human, everyday experience. These hunches may then, later, be translated into objective terms. … I … intend to go ahead imagining how, *if I were a rat*, I would behave as a result of such and such a demand combined with such and such [hypotheses] and so on. And then on the basis of such imaginings, I shall try to figure out some sort of f_3 rules or equations. And then eventually I shall try to state these latter in some kind of objective and respectable sounding terms. (Tolman 1938/1966, 163–164)

While Tolman called it the haziest part of his illustration of operational behaviorism, it is actually the clearest indication that in his scientific work he aimed at constructing intelligible models. In this example, the collection of intervening variables such as 'demand' and 'hypotheses' constitute the theoretical model of the behavior of rats in mazes. Tolman used his pre-scientific understanding of rats to construct this model. This understanding was based on certain skills, of

which imagining was the most important. He imagined how he would behave if he were a rat. This is imagining in the sense of empathizing. By means of this skill, he was able to give an anthropomorphic conceptualization of the events in the maze. It enabled him to cast concepts such as 'demand' and 'hypotheses' into a mold such that he could derive useful hunches from everyday experience about how he would behave if he were at the choice point. Obviously, the transformation of the graphs mentioned above was already part of this framing process. The theoretical terms used in his model of the rat in the maze were operationally defined in such a way that their definitions were in accordance with Tolman's anthropomorphic interpretation of these terms. This shows that the meaning of the theoretical terms exceeded their objective definition; these terms possessed surplus meaning. Conceptualizing the phenomenon in terms of 'demand' and 'hypotheses' is not trivial – at least not from the positivist point of view that Tolman advocated. It requires performing actions such as choosing variables and transforming graphs that are not strictly rule-governed. By using terms such as 'demand' and 'hypotheses' in his model, he ensured that the model had the virtue of being interpretable in the above-mentioned way, thus warranting its intelligibility (provided that one possesses the skill of imagining).

Tolman's model satisfies important criteria for intelligibility that are advocated in this study. First, Tolman was able to apply the model to the behavior of the rats, or in Giere's words (2004, 747), he was able to "do the representing." By putting himself in the position of a rat, he was able to judge the similarities between the model and the phenomenon of rats in mazes and pick out the relevant features of the model (e.g. from everyday life he knew that demand is a relevant feature). Second, he was able to use the model for developing qualitative insight into its consequences in concrete situations. His skill in imagining enabled him to derive useful preliminary hunches that he could use to see intuitively how the rats were expected to behave in concrete situations (such and such demands combined with such and such hypotheses). This skillful act of reasoning via the model can be explicated in more detail by invoking the view of "purposive behavior in animals and men" that Tolman (1932) had developed several years before he proposed his ideas about operational behaviorism. In this view, "environmental features are cognitively represented in terms of

how they can be used or manipulated in order to attain certain goals" (Feest 2005, 141). For instance, a chair may be represented as something on which, if placed against the wall, one can stand to reach a picture. The cognitive representations involve expectations as to the outcomes of hypothetical actions, and the purposive behavior of animals and men is the result of means-end reasoning with respect to these cognitive representations. Because Tolman's model of the rats in the maze is formulated in terms of expectations ('hypotheses') and goals ('demands'), reasoning via this model is also a matter of means-end reasoning. Therefore, the skill that is necessary to reason via this model can be described as instrumental reasoning, or means-end reasoning. In sum, using the model to represent the phenomenon requires the skill of imagining in the sense of empathizing and the skill of means-end reasoning. His emphatic abilities enabled Tolman to imagine the demands and hypotheses of the rats in the maze, and by means of means-end reasoning he was able to imagine how rats behave as a result of these demands and hypotheses. Therefore, the model was intelligible to Tolman and thus provided him with understanding of the phenomenon of rats in mazes.

However, one might ask if this understanding should be labeled scientific understanding. At first sight, the definitions of the theoretical terms in his model have no relation to theoretical principles. The model does not seem to be composed of abstract objects defined by theoretical principles as in Giere's representational view of scientific models. Therefore, one may wonder if the model does meet the epistemic values of science, such as scope or fruitfulness. For instance, the scope of the definition of 'hypotheses' is very narrow. Because this intervening variable is defined as a function of the number of trials of the *rats* in the maze, Tolman's model cannot be used to explain, for instance, the behavior of *mice* in mazes. Therefore, the model seems to be deficient in the possibility of fruitful application in other domains. Because in my account of scientific understanding the value of intelligibility encompasses the other epistemic values, including fruitfulness, Tolman's model does not seem to satisfy all the conditions for intelligibility advocated in this study. Therefore, Tolman's model does not seem to provide full scientific understanding.

Yet, his remark that the graph he used to formulate the definition of 'hypotheses' was actually the learning curve suggests that the

intervening variables in Tolman's model can, in fact, be linked to theoretical principles. According to the learning rule, which is an important principle in neo-behaviorism, the learning curve describes the function between the strength of a stimulus-response connection and the number of 'reinforcements' (rewards and punishments). By identifying the 'hypotheses' graph with the learning curve, Tolman apparently considered the 'hypotheses' to be a stimulus-response connection whose strength, according to the principle of the learning rule, is defined as a function of 'reinforcements.' In this conception, the meaning of 'hypotheses' is more comprehensive than the meaning of the term as specified by the operational definition. Instead of capturing the complete meaning of 'hypotheses,' the operational definition only partially specifies Tolman's usage of this concept, which has a potential domain of applicability that is more wide-ranging than the situation of the rats in the maze as described. 'Hypotheses' may, for instance, also be applicable to the situation of *mice* in mazes. With this conception of the theoretical terms in Tolman's model, the model does not have the shortcoming of being deficient in satisfying epistemic values such as scope and fruitfulness. Therefore, in this conception Tolman's model provides *scientific* understanding. In sum, my analysis of Tolman's neo-behaviorist methodology highlights the fact that his scientific models are intelligible due to the surplus meaning of their theoretical terms. The surplus meaning of the terms is valuable because it enables the scientist to understand the phenomena in the cases at hand scientifically.

Although this analysis is in line with the claim that intelligibility is an epistemic value of science, Tolman's assertion that he would eventually try to translate the outcome of his methodology into "objective and respectable sounding terms" – which, because of the name of his methodology, I take to be operational definitions – seems to indicate that his reliance on intelligible models is merely an intermediate step in the development of objective scientific claims. The final step would consist of a translation of his claims into objective terms without surplus meaning. If Tolman had succeeded in removing the surplus meaning of the terms, the model would no longer be intelligible in the way described above. In the example, however, Tolman did not perform the last step. In my view, his assumptions about the feasibility of this step are mere speculation. Epistemic reasons do not

allow the surplus meaning of the theoretical terms to be removed. One of these reasons is related to the observation made by the historian Thomas H. Leahey that, in actual scientific practice, the establishment of operational definitions requires justification. A psychologist who formulates an operational definition "must persuade the psychological community that his 'definition' . . . is a good one" (Leahey 1980b, 138). For instance, in order to establish the operational definitions of 'demand' and 'hypotheses,' Tolman performed several acts that were not methodologically motivated, such as the inversion and translation of graphs. I think that the justification for these acts lies in the surplus meaning of 'demand' and 'hypotheses' that renders the models in which these concepts are used intelligible to their users.

The second part of my case study will focus on the epistemic reasons why the surplus meaning of theoretical terms should not be removed. I will examine some of the theoretical terms introduced in the work of Hull. As in the case of Tolman, I will show that Hull's theoretical terms possess surplus meaning. In addition to the case of Tolman, I will demonstrate that removing the surplus meaning of the terms would imply such a restriction of their domain of applicability that they would loose their epistemic significance. I will conclude that the intelligibility of scientific models has epistemic significance: it is an epistemic value of science.

4.2.3. *Clark L. Hull and the Application of Theoretical Terms in Different Domains*

Clark Leonard Hull (1884–1952) studied engineering but was unable to work as a mining engineer because he was plagued by ill health and poor eyesight. At 24 he contracted polio, which left him disabled in one leg and forced to wear an iron brace. This made it necessary to find a physically less demanding career. Like Tolman, Hull switched to psychology after reading James' *Principles of Psychology*. He majored in this field at the University of Michigan, where he was especially enthusiastic about the course in experimental psychology given by Walter B. Pillsbury and John F. Shepard. He then enrolled in graduate school at the University of Wisconsin, where he was strongly influenced by Daniel Starch and Vivian A.C. Henmon (Beach 1959, 127).

After receiving his Ph.D. in 1918 from the University of Wisconsin, he worked there for ten years. In 1929 he accepted a position as research professor at Yale University.

Although Hull was a psychologist, his scientific approach revealed his engineering background. For instance, he developed several machines that he used in his psychological research, such as a machine for calculating correlations that was useful for the development and use of psychological tests, and a logic machine that could display all the implications of any type of syllogism. His worldview was inspired by materialism, which was a central characteristic of his research program. In 1928, he considered calling his magnum opus "Psychology from the Standpoint of a Mechanist." Although the book appeared in 1943 under another title, the proposed title demonstrates the significance of his mechanistic worldview for this scientific work. He applied what he claimed to be the "Newtonian" view of the universe as a machine to living organisms. For instance, he developed ingenious stimulus-response models, mainly on paper but also as actual machines. They enabled him to give a mechanistic account of the seemingly non-mechanical aspects of behavior (Still 1997b, 285).

Like Tolman, Hull was very interested in the possibilities of transforming psychology into an objective science. He gave a seminar on Watson's behaviorism in 1925 and studied Pavlov's work when it appeared in translation in 1927. Pavlov's idea of the conditioned reflex became an important ingredient of his thinking. Together with collaborators, Hull even designed and implemented mechanical devices to simulate the conditioned reflex. They experimented with different types of mechanical devices until they settled on an electric circuit:

> The mechanism which has given the best results is a combination of polarizable cells and mercury-toluene regulators, which are sensitive to temperature changes. Ordinary electric switches serve as "receptors"; a flashlight bulb is the responding "organ," analogous to the salivary gland of the experimental animal. By a manipulation of the switches in a manner strictly analogous to the presentation of stimuli in Pavlov's conditioned reflex experiments, phenomena, paralleling fairly accurately a considerable number of the properties of the conditioned reflex, are obtained. (Hull and Baernstein 1929, 15)

After studying Thorndike's *Fundamentals of Learning* in 1935, Thorndike's law of effect replaced Pavlov's conditioned reflex as central to

Hull's thinking. In his ideas about the possibility of an objective psychology, Hull expected much from the use of the hypothetical-deductive method. Because this method is not mentioned in Tolman's description of operational behaviorism, one might think that the methodologies of Tolman and Hull differ essentially. However, I will argue below that Hull's methodology is in fact a continuation and elaboration of that of Tolman. Hull used Tolman's idea to define theoretical terms operationally and, like Tolman, he constructed ingenious theoretical models of the adaptive behavior (learning behavior) of organisms. Hull's methodology was very influential: in the 1940s, at the peak of his career, Hull was the most widely known behaviorist and his methodology dominated psychology (Smith 1986, 149).

Hull's ideas about science were in many respects very similar to those of the logical positivists. He shared their interest in the use of formal logic, and he was attracted to British empiricism. David Hume's *Treatise of Human Nature* especially fascinated him, and he regarded behaviorism as a direct descendant of British associationism, of which Hume's work was exemplary (Smith 1986, 152). In the papers he contributed to the international conferences of the *Unity of Science Movement* – for instance, in 1937 in Paris and in 1941 in Chicago – he described his approach in psychology as the methodology of logical positivism. Here, and also in his later work (e.g. Hull 1938, 159–160), he emphasized the kinship between logical positivism and behaviorism:

> There is a striking and significant similarity between the physicalism doctrine of the logical positivists (Vienna Circle) and the approach characteristic of the American behaviorism originating in the work of J.B. Watson. Intimately related to both of the above movements are the pragmatism of Peirce, James, Dewey on the one hand, and the operationism of Bridgman, Boring and Stevens, on the other. These several methodological movements, together with the pioneering experimental work of Pavlov and the other Russian reflexologists, are, I believe, uniting to produce in America a behavioral discipline which will be a full-blown natural science; this means it may be expected to possess not only the basic empirical component of natural science, but a genuinely scientific theoretical component as well.
>
> (Hull 1943a, 273)

However, the idea of introducing a methodology that followed logical positivism is not unproblematic. The logical positivists dealt with the logical reconstruction of theories and concepts, and not primarily

with methodologies that could guide scientists in their construction of theories (Mackenzie 1977, 115). For Hull, who shared Hume's admiration for Newton, it was apparent that the best candidate for a logical-positivist methodology was Newton's hypothetical-deductive methodology. In Hull's interpretation of this methodology (which is more a reflection of his own view of the scientific method than of Newton's view) theories are formulated as deductive systems consisting of definitions and postulates from which empirical phenomena can be deduced. Hull had read Newton's *Principia* and had spent the summer of 1929 at Harvard (Smith 1986, 165), where he discussed *Principia Mathematica* (1927) by Alfred N. Whitehead and Bertrand A.W. Russell with Lewis and Whitehead. This work on the foundations of mathematics, which consisted of a formal system of axioms from which mathematical theorems were deduced, was a stimulus for the ideas in his *Principles of Behavior* (1943b).

Hull was very explicit about his view on methodology. As Koch (1954, 11) observes, "[i]n virtually every one of his theoretical publications, Hull felt compelled to include general discussion of the nature of scientific theory, and to lay down corresponding prescriptions for the construction of adequate psychological theory." For instance, in 1936, when he was elected president of the *American Psychological Association* (one year before Tolman), he devoted his inaugural speech to the explanation of the hypothetical-deductive method. Hull argued that theories are hypotheses that should be adjusted when the theorems deduced from them do not agree with empirical facts. The adjustment of a theory to empirical data is a matter of trial and error. Whenever a theorem deduced from a theory fails to agree with empirical facts, the postulates of that theory must be revised. If agreement cannot be attained, the system must be abandoned (Hull 1937, 8). How a scientist comes up with a theory does not really matter because eventually the theory is assessed and eventually modified on an empirical basis:

> The history of scientific practice so far shows that, in the main, the credentials of scientific postulates have consisted in what the postulates can do, rather than in some metaphysical quibble about where they came from. If a set of postulates is really bad it will sooner or later get its user into trouble with experimental results. On the other hand, no matter how bad it looks at first, if a set of postulates consistently

> yields valid deductions of laboratory results, it must be good. In a word, a complete laissez-faire policy should obtain in regard to postulates. Let the psychological theorist begin with neurological postulates, or stimulus-response postulates, or structural postulates, or functional postulates, or factor postulates, or organismic postulates, or Gestalt postulates, or sign-Gestalt postulates, or hormonic postulates, or mechanistic postulates, or dynamic postulates, or postulates concerned with the nature of consciousness, or the postulates of dialectical materialism, and no questions should be asked about his beginning save those of consistency and the principle of parsimony.
>
> (Hull 1935, 511)

A problem for Hull was that, in general, the hypothetical-deductive method relies heavily on the use of theoretical terms in the postulates of the theories. The use of these terms was condemned in classical behaviorism as metaphysical. Hull (1943b, 31) wrote that it was to Tolman's credit that he had resolved this problem: Tolman's objective behaviorism had introduced the use of theoretical terms in behaviorism and, as such, paved the way for the hypothetical-deductive method. In this section, I will examine the theoretical terms in Hull's theories of behavior to see if they possess surplus meaning that has epistemic significance.

According to Hull (1943b, 21), scientists frequently and usefully employ theoretical constructs such as "electrons, protons, positrons, etc." to facilitate their thinking. The use of these constructs has the advantage of being economical: "The use of logical constructs … comes down to a matter of convenience in thinking, i.e., an economy in the manipulation of symbols" (Hull 1943b, 111):

> In the case of hunger, for example, there must be an equation expressing the degree of drive or motivation as a function of the number of hours' food privation, say, and there must be a second equation expressing the vigor of organismic action as a function of the degree of drive (D) or motivation, combined in some sense with habit strength. … Now it is a relatively easy matter to find a single empirical equation expressing vigor of reaction as a function of the number of hours' food privation or the strength of an electric shock, but it is an exceedingly difficult task to break such an equation up into the two really meaningful component equations involving hunger drive (D) or motivation as an intervening variable. It may confidently be predicted that many writers with a positivistic or anti-theoretical inclination will reject such a procedure as both futile and unsound. From the point

> of view of systematic theory such a procedure, if successful, would present an immense economy. This statement is made on the assumption that motivation (D) as such, whether its origin be food, privation, electric shock, or whatever, bears a certain constant relationship to action intensity in combination with other factors, such as habit strength. If this fundamental relationship could be determined once and for all, the necessity for its determination for each special drive could not then exist, and so much useless labor would be avoided.
>
> (Hull 1943b, 67)

Although theoretical constructs are widely applied in natural science and have great advantages, Hull warned that their use "is attended with certain difficulties and even hazards" (Hull 1943b, 22). Because these constructs are hypothesized and claims about them are not directly verifiable, the use of theoretical constructs could result in "unverifiable theories." This should be avoided, and therefore the symbolic constructs have to be "anchored to observable and measurable conditions or events on both the antecedent and consequent sides" (Hull 1943a, 281):

> When a hypothetical dynamic entity, or even a chain of such entities each functionally related to the one logically preceding and following it, is thus securely anchored on both sides to observable and measurable conditions or events (A and B), the main theoretical danger vanishes. This at bottom is because under the assumed circumstances no ambiguity can exist as to when, and how much of, B should follow A.
>
> $A - f \rightarrow (X) - f \rightarrow B$ (Hull 1943b, 22)

Hull's schematization of the relation between antecedent events (A), theoretical terms (X), and consequent events (B), is similar to Tolman's schematization of the relation between independent variables, intervening variables, and dependent variables discussed above. The functional relations between A and X or between X and B (which are not meant to be identical, even if in the schematization they are indicated by the same symbol) can be used as operational definitions of X by means of which these theoretical terms can be securely anchored on the antecedent or consequent side. As in Tolman, Hull's rationale for expressing the theoretical terms as functions of these objectively observable conditions was to ensure that his scientific claims contained only objective terms that did not possess surplus meaning.

Although Hull was an advocate of positivist psychology, the question if his own scientific work was in agreement with the strict standards of logical positivism has been raised. For instance, in an article by MacCorquodale and Meehl (1948), which I will discuss in more detail below, it was argued that important theoretical terms in Hull's work had surplus meaning. A few years earlier, Kenneth W. Spence, one of Hull's former pupils, had already expressed similar worries about Hull's use of theoretical terms. In his writings, Hull often supplemented the formal definitions of hypothetical constructs with informal formulations in which the constructs were depicted as mechanisms mediating between stimuli and responses. Spence criticized this, and in 1942 and 1943 he wrote several letters in which he repeatedly warned Hull against his habitual practice:

> I have always been very unhappy about the fact that you have been inclined to throw in hypotheses as to the mediational mechanisms underlying the abstract mathematical concepts.
>
> (Spence to Hull, Sept. 8, 1943, quoted in Smith 1986, 223)

Illustrative is the comparison Spence made between Hull and the physicist Maxwell who regularly used mechanical analogies to illustrate his explanations of natural phenomena:

> What you are doing would be analogous to Maxwell insisting on the mediational mechanism of an ether-like medium to explain electromagnetic phenomena rather than depending upon the purely mathematical aspects of his theory.
>
> (Spence to Hull, March 8, 1942, quoted in Smith 1986, 223)

This habit of giving an informal, mechanistic interpretation of the theoretical terms was one of the sources of surplus meaning. Another source was the naming of the theoretical terms. Like Tolman, Hull used names that already possessed a meaning in everyday life, such as 'drive,' 'demand,' 'habit strength,' and 'reinforcement.' Although he formulated his scientific claims in objective terms with operational definitions, the naming practice could cause the meaning of these terms to exceed these definitions. I will show that Hull's reason for providing the theoretical terms with surplus meaning was the same as Tolman's: it enabled him to connect the theoretical principles of behavior to the concrete behavior of higher organisms (such as humans and other mammals) via the construction of intelligible theoretical models.

To support this claim, I will focus on Hull's use of the theoretical terms 'reinforcement' and 'habit strength,' which were important terms in the learning principle described in his *Principles of Behavior*. In his ideas about adaptation and learning of organisms, Hull was strongly influenced by Pavlov's theory of conditioned learning. He saw the reflex between receptors and effectors as a causal mechanism. He even built a mechanical apparatus to simulate the conditioned reflex. This shows that Hull was actually not as hesitant about "metaphysical" speculation as the logical positivists. He wrote that receptor-effector connections of an organism correspond roughly to what are known to common sense as habits (Hull 1943b, 102). He called the strength of such a connection "habit strength." Hull used the learning curve to define 'habit strength' (see figure 4.2.3.a). Because, as I discussed above, Tolman (1938/1966, 145) had argued that the graph that related 'hypotheses' to the number of previous trials of the rats in the maze was "no more than our old friend, the learning curve," it may seem that Hull's theoretical term 'habit strength' and Tolman's theoretical term 'hypotheses' have the same meaning.

However, Hull's 'habit strength' is more general than Tolman's 'hypotheses.' Whereas in Tolman's example, the intervening variable 'hypotheses' has a limited scope because it is operationally defined as a function of the number of previous trials made by hungry rats, in Hull's account 'habit strength' is defined by means of a theoretical principle, namely the learning principle depicted by the learning curve, in which

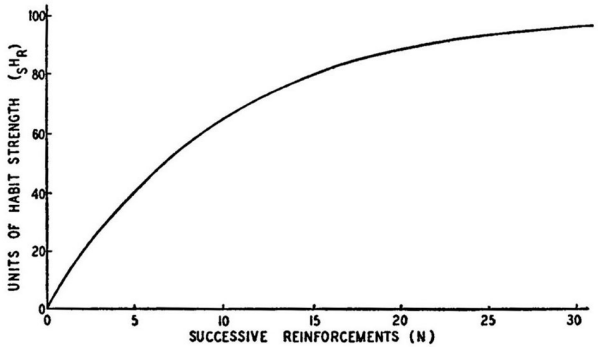

Figure 4.2.3.a. The learning curve (Hull 1943b, 117)

'habit strength' is described as a function of successive 'reinforcements.' In his attempt to be as general as possible, Hull (1943b, 71) did not give a specification of the meaning of 'reinforcement'. Hull's definition of 'habit strength' is, therefore, an exemplification of Giere's view that theoretical principles (such as the learning principle) function as definitions of the abstract "objects" (such as 'habit strength') that constitute theoretical models. For concrete situations, such as Tolman's T-maze, it has to be specified what counts as 'reinforcement.' I will show that Hull was able to connect his scientific models of behavior to concrete instances of the learning behavior of higher organisms by giving what Radder (2006, 119–120) calls a "local" interpretation of the meaning of the theoretical terms in these models.

I will investigate how Hull used scientific models to connect theoretical principles to concrete phenomena, and I will examine the importance of the surplus meaning of theoretical terms for applying these models. Although the theoretical terms in his learning models, such as 'habit strength' and 'reinforcement,' were defined by means of the theoretical principles of behavior, their meaning transcended these definitions. As in the case of Tolman, Hull's theoretical terms possessed surplus meaning that rendered his learning models intelligible. Therefore, these models could be used to represent the phenomena and to develop qualitative insight into the consequences of the models in concrete situations. I will demonstrate that, in contrast to the methodological ideas of Tolman described above, the surplus meaning of the theoretical terms should not be removed (if it is at all possible to do so) because it has epistemic significance.

In his *Principles of Behavior* (1943b), Hull discussed a number of experiments that he claimed were all measurements of 'habit strength' in an "indirect way" (Hull 1943b, 102). These measurements were indirect because the experiments only measured functional dependencies between independent and dependent variables, rather than dependencies between these variables and 'habit strength.' In that sense these indirect measurements are similar to Tolman's standard experiments, where only functional dependencies between independent and dependent variables were measured as well, and not between these variables and the intervening variables. It is remarkable that the experiments that Hull claimed were measurements of the same theoretical term 'habit strength' were completely different from each

other. They were performed with different organisms as subjects of investigation and with completely different techniques. Still, Hull considered all these experiments to be dealing with 'habit strength.' I will analyze how Hull managed to connect his learning principles, which were used to define 'habit strength,' to this wide range of different experiments.

In one of the experiments, a human subject repeatedly received a mild electric shock after hearing a noise. After the training period, only the noise was produced, and the strength of the galvanic skin reaction of the subject was measured as a function of the number of "reinforcement repetitions" received in the training period.

Hull recognized the learning curve in the graph that depicted the result (see figure 4.2.3.b) and argued that this meant that the result of the experiment was a manifestation of 'habit strength' (Hull 1943b, 102). Such a claim can only be made if it is reasonable to suppose that receiving a mild electric shock after hearing a noise can be seen as (negative) 'reinforcement,' and that it is reasonable to see the galvanic skin reaction after receiving a shock as a 'habit' whose amplitude corresponds to the 'habit strength.' Hull apparently considered this to be reasonable, and the similarities with Tolman's example suggest that this was due to the surplus meaning of theoretical terms, such as

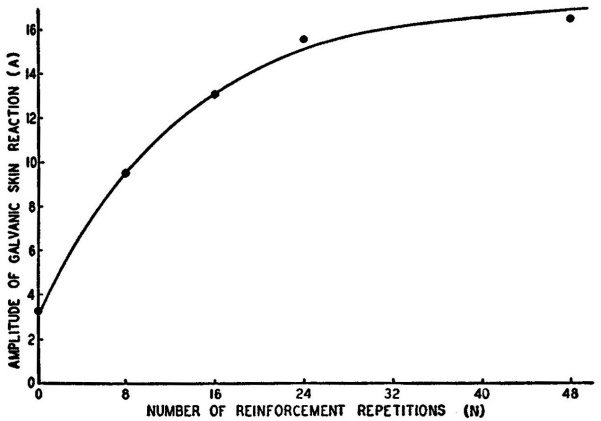

Figure 4.2.3.b. Manifestation of habit strength in an experiment with human subjects (Hull 1943b, 103)

'habit strength' and 'reinforcement.' By being empathic (like Tolman), Hull was able to understand that repeatedly receiving a mild electric shock is a form of negative 'reinforcement.' He could imagine that it is unpleasant to receive these shocks, and he could imagine that the experience that these shocks always succeeded the noise creates an expectation. The role played by the surplus meaning of these terms is similar to that played by surplus meaning in Tolman's case, where the surplus meaning of 'demand' made it possible to view this theoretical term in the light of everyday experiences, which enabled making the connection with concrete situations.

In another experiment, albino rats were trained to press a bar to obtain food. After the training period, bar pressing was no longer followed by a reward, and the number of times that the rats kept on pressing the bar without obtaining food was measured as a function of the number of "reinforcement repetitions" in the training period.

Again, Hull recognized the learning curve in the graph that depicted the result (see figure 4.2.3.c), and again he argued that this meant that the result of the experiment was a manifestation of 'habit strength' (Hull 1943b, 102).

In Hull's view, the theoretical term 'habit strength,' which can be used to account for the results of the experiment with the electric

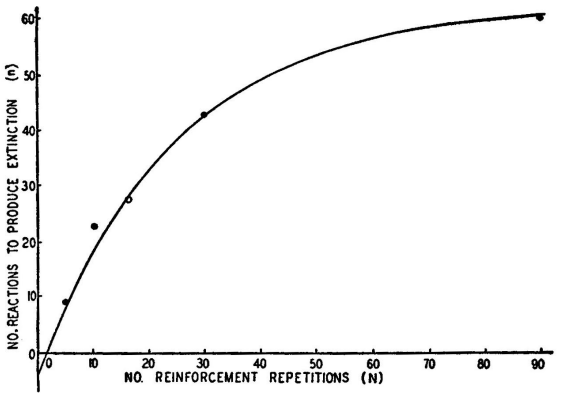

Figure 4.2.3.c. Manifestation of habit strength in an experiment with albino rats (Hull 1943b, 106)

shock, can also be used in other domains: he regarded his experiment with the rats a replication of the experiment with the electric shock – they are both experiments that measure the functional relationship between 'habit strength' and the number of 'reinforcement repetitions,' even though the "material realization" (Radder 2006, 103) of the experiments differs completely. This means that Hull regards 'habit strength' to be an extensible concept (Radder 2006, chapters 9 and 10). The scope of extensible concepts is not constrained to a specific domain but can be enlarged by successfully applying these terms in a materially different domain. To put it in Radder's terms, these terms have "nonlocal meaning."

To call experiments that produce similar graphs replications of one another requires certain judgments. For instance, one of these judgments concerns the question if it is reasonable to suppose that both receiving a shock after hearing a noise and obtaining food after depressing a bar can be seen as 'reinforcement.' Another judgment concerns the question if it is reasonable to see the number of times that rats keep pressing a bar without obtaining food as a measure of 'habit strength.' The view that the connection between models and phenomena depends on judgments by the model users goes beyond the neobehaviorist conception of this connection, in which the connection between models and phenomena should be stated in objective terms by means of operational definitions. In the case of Hull, if he would define 'reinforcement' by means of operations involving administering electric shocks to human subjects (in particular situations) and administering food to albino rats (in particular situations), the domain of his theoretical principles would be restricted to situations in which these operations are applicable. The principles would only denote relationships between specific variables in separate and limited domains, and, accordingly, have no epistemic relevance in other domains. This, however, is not in accordance with Hull's intention, which was to formulate a general theory of behavior.

Hull never specified how the theoretical principles should be applied in different domains; it seems that he regarded this as obvious. In his methodological writings he never explained how the terms in his models should be related to concrete situations, and in his discussion of the different experiments on 'habit strength' he did not explain how he made the judgment that all these experiments deal with

the same concepts. A detailed examination of Hull's work reveals that establishing the connection with concrete phenomena necessarily involves judgments that are not rule-governed. For instance, it involves judgments about the similarities of features of the models to features of the phenomena and about the relevance of these similarities. Hull seemed to make these judgments as a matter of course. Apart from the two experiments discussed above, Hull gave numerous other examples of indirect measurements of 'habit strength' in experiments with animals, such as rats, dogs, chickens, as well as human subjects, and in all of these different situations Hull recognized the same learning curve. This recognition of the learning model in the empirical data illustrates the theory-ladenness of observation. As for instance Hanson (1958), and Kuhn (1962) famously argued, observations are inherently imbued with theoretical preconceptions. In Hull's case, the observations were influenced by the theoretical concepts of his learning models. Due to this theory-ladenness of his observations, Hull was able to relate his experiments to his learning models. He simply knew what to call 'stimuli,' what to call 'responses,' and what to call 'reinforcement.' I submit that the reason for this is that the models were intelligible to him due to the surplus meaning of the theoretical terms.

Hull had no difficulty applying the same models to quite distinct behaviors of higher organisms. Like Tolman, he was able to "do the representing" (to use Giere's phrase again) due to his skill in imagining in the sense of empathizing. Because he was able to imagine himself in the position of the human subjects who received a mild electric shock after hearing a noise, or in the position of the rats that obtained food after pressing a bar, he was able to identify the shocks and the food as cases of 'reinforcement.' In other words, he was able to conceptualize these situations in the theoretical terms of his scientific models. Subsequently, he was able to develop qualitative insight into the consequences of the models, which is an ability that De Regt and Dieks (2005, 151) proposed as one of the criteria for intelligibility. For Hull, the models were intelligible, and he was able to apply them to the phenomena due to the right combination of his skills and the virtues of the models. Like Tolman's model, Hull's models possessed virtues, such as anthropomorphic interpretability, that originated from the surplus meaning of the theoretical terms in the models. In addition,

the theoretical terms possessed causal-mechanical surplus meaning. For instance, Hull regarded 'habit strength' as the outcome of a causal mechanism he had simulated by building a mechanical apparatus. Hull had attributed causal-mechanical surplus meaning to his theoretical terms, and it was precisely this attribution of surplus meaning that Spence criticized in his letters to Hull.

This critique, however, is misplaced. It was only because the terms possessed this surplus meaning and because Hull possessed the skills of empathic imagining and causal reasoning that he was able to perform the skillful activity of using his scientific models for representing the phenomena and for developing qualitative insight into their consequences in concrete situations.

In this chapter, the point of departure was Tolman's concrete example of rats in mazes that he used to exemplify his ideas of an objective psychological methodology. I used this example to illustrate the notion of intelligible models. Because Tolman used his pre-scientific understanding of rats as a heuristic guide in formulating abstract models about their behavior, the meaning of the terms in the model was initially based on this pre-scientific understanding. A detailed examination of his example revealed that this implied that the theoretical terms have surplus meaning. Their meaning transcended their objective definition. This surplus meaning rendered the model intelligible to him. The right combination of his skills (imagining in the sense of empathizing and means-end reasoning) and the virtues of the model (anthropomorphic interpretability) originating from the surplus meaning of the theoretical terms used in the model enabled him to use the model to represent the phenomena of rats in mazes and to develop qualitative insight into its consequences in concrete situations.

In Tolman's view on methodology, the use of intelligible models is merely an intermediate step. He believed that the surplus meaning of the theoretical terms that bring about the model's intelligibility should eventually be removed. I do not share this view. My study of Hull's work shows that the surplus meaning of theoretical terms in his learning models is significant for the application of these models to concrete phenomena.

The surplus meaning of the theoretical terms remains significant even if the terms are provided with specific operational definitions in concrete domains. First, the surplus meaning is significant for the

justification of the establishment of these specific operational definitions. I agree with Uljana Feest (2005, 134–135), who argues that in the scientific practice of neo-behaviorism an operational definition of a concept is "not intended to state necessary conditions for the applicability of the term." Instead, it has the methodological function of getting "an empirical handle on a phenomenon." Contrary to Bridgman's intentions, an operational definition "does not provide a necessary, but at best a sufficient condition for application" (Feest 2005, 137). The establishment of an operational definition therefore involves the judgment that the operational definition can be considered an adequate means for applying the theoretical concept it represents (cf. Leahey 1980b, p. 138). For instance, it has to be assessed if 'reinforcement' can be defined operationally as the result of repeatedly performing the operation of administering a mild electric shock to human subjects immediately after presenting a noise. The case study reveals that Hull was able to make these judgments due to the surplus meaning of his theoretical terms.

Second, the surplus meaning of the theoretical terms is significant for extending the domain of applicability of these terms, and thus of the models that contain them. For instance, in order to apply the term 'reinforcement' to the situation of the bar-pressing rats, it has to be established that, supplementary to the operational definition involving the electric shocks, 'reinforcement' can be defined operationally as the result of repeatedly performing the operation of giving food to a rat if it depresses a bar. This extension of the domain involves judgments that are facilitated by the surplus meaning of the theoretical concepts. The same judgments are required for the application of concepts in their original domain. Without the surplus meaning of the concepts, these judgments cannot be made.

Thus, in spite of Spence's critique in his letters to Hull, there are good epistemic reasons for providing terms with surplus meaning. The surplus meaning of the theoretical terms in theoretical models renders these models intelligible to their users. Removing the surplus meaning would make the models unintelligible and therefore useless. Scientific models need to be intelligible to be applicable to phenomena.

4.3. *The 1950s Dispute on Theoretical Terms and their Surplus Meaning*

In my discussion on the desirability of surplus meaning of theoretical terms I assumed that there are no major differences between Tolman's and Hull's methodologies. Hull regarded his hypothetical-deductive methodology as an elaboration of Tolman's operational behaviorism, and I have therefore assumed that the conclusion of my analysis that the surplus meaning could not be removed from Hull's theoretical terms without making his behavioral models useless also holds for Tolman. However, at the beginning of the 1940s, when psychologists became "increasingly aware of the methodological problems of their science," the affinity between Tolman's methodology, with its emphasis on the empirical component of scientific method, and Hull's, with its emphasis on the formal or theoretical component of scientific endeavor, was questioned (Bergmann and Spence 1941, 1). At that time, the University of Iowa had become a center for neobehaviorist thought where positivist philosophers of science and neobehaviorist psychologists collaborated. Bergmann (1906–1987), who studied mathematics, law, and philosophy in his native Vienna and was a member of the Vienna Circle, became attached to this university after his emigration to the United States in 1938. He worked in the departments of philosophy and psychology for forty years. Kenneth W. Spence (1907–1967), who would become Hull's best-known student, worked at the University of Iowa from 1938 to 1964. With their papers on theoretical psychology (Bergmann 1940a; Bergmann 1940b; Bergmann 1953; Bergmann and Spence 1941; Spence 1944; Spence 1948), Bergmann and Spence became leading figures in the alliance between behaviorism and logical positivism (Smith 1986, 209). In 1938 Koch (1917–1996), a talented student in philosophy and psychology, also came to Iowa, where he wrote his master's thesis on the "far-reaching methodological renaissance" in psychology (Koch 1941, 15). In 1941 Bergmann and Spence, as well as Koch, analyzed Tolman's and Hull's respective methodologies and concluded that these were "fully in line, indeed identical" (Bergmann and Spence 1941, 13).

According to Bergmann and Spence (1941, 13), Tolman's conception of intervening variables and Hull's conception of theoretical terms in his theoretical models were equal. Both Tolman and Hull

introduced theoretical terms to break the mathematical equation that describes the experimental laws relating the independent variables (external stimuli) and the dependent variables (types of behavior) down into component functions:

> [I]t is to Tolman's credit that he has been one of the first in psychology to outline this general methodological scheme. His actual theorizing however, has not gone beyond suggesting and cataloguing the various possible intervening variables and showing how they provide for the definition and use of mental terms (demands, hypotheses, etc.) in a behavioristic psychology. (Bergmann and Spence 1941, 13)

Hull's work can be seen as a continuation and elaboration of Tolman's program. Tolman and Hull used the same methodological procedure. Operationism and the hypothetical-deductive methodology supplement each other.

I agree with the analysis by Bergmann and Spence (1941) and Koch (1941) that there is no major difference between Tolman and Hull in their use of theoretical terms. Consequently, I hold that, as with Hull's theoretical terms, the surplus meaning cannot be removed from Tolman's either without making his behavioral models useless. It is because of the surplus meaning that the models are intelligible. Because of this surplus meaning, the models have virtues that, if they match the skills of the scientists, facilitate the successful application of the models. Therefore, removing this surplus meaning would make them inapplicable.

However, at the end of the 1940s and the beginning of the 1950s, a debate started among theoretical psychologists and logical positivists on the question if Tolman's and Hull's respective methodologies differed. The issue at stake was precisely if Tolman's theoretical terms, like Hull's, possessed surplus meaning and, if not, if they should possess it. Furthermore, the question was raised if terms with a particular kind of surplus meaning, such as physiologically inspired hypothetical concepts, were preferable to other theoretical terms. I will analyze this debate and relate the arguments that are put forward about the significance of surplus meaning to my account of intelligibility. At first sight, the contributions to the debate do not seem to be very coherent. They deal with a variety of issues such as operationism, realism, and reductionism. However, I submit that the debate among neobehaviorists and logical positivists around 1950 should be understood

as a coherent debate that was (implicitly) centered around one theme, namely the question of how the value of intelligibility should be incorporated into neo-behaviorist theories.

4.3.1. *Intervening Variables and Hypothetical Constructs*

The surplus meaning of theoretical terms became an issue in neo-behaviorism because of a classic paper by the psychologists MacCorquodale and Meehl (1948) entitled "On a Distinction between Hypothetical Constructs and Intervening Variables." In this paper they distinguished between "two subclasses of intervening variables, or we prefer to say, between 'intervening variables' and 'hypothetical constructs' which we feel is fundamental but is currently being neglected" (MacCorquodale and Meehl 1948, 95). Roughly speaking, this distinction is the difference in logical status between "constructs which involve the hypothesization of an *entity*, *process*, or *event* which is not itself observed, and constructs which do not involve such hypothesization." According to MacCorquodale and Meehl, the hypothetical entities that Hull postulated with his hypothetical-deductive methodology belonged to the first subclass of constructs, which they called hypothetical constructs, whereas Tolman's operationally defined theoretical terms exemplified the other subclass of constructs, which they called intervening variables. As an example of the difference between intervening variables and hypothetical constructs, they contrasted the notion of 'resistance' in electricity to the notion of 'electron':

> The resistance of a piece of wire is what Carnap has called a *dispositional concept*, and is defined by a special type of implication relation. When we say that the resistance of a wire is such-and-such, we mean that "so-and-so volts will give a current of so-and-so amperes." … Resistance, in other words, is 'operational' in a very direct and primitive sense. The electron, on the other hand, is supposedly an *entity* of some sort. Statements about the electron are, to be sure, supported by means of observational sentences. Nevertheless, it is no longer maintained even by the positivists that this set of supporting sentences exhaust the entire *meaning* of the sentences about the electron. Reichenbach, for example, distinguishes *abstracta* from *illata* (from Lat. *infero*). The latter are 'inferred things,' such as molecules, other people's minds, and so on. They are believed in on the basis of our impressions, but the

> sentences involving them, even those asserting their existence, are not reducible to sentences about impressions. This is the epistemological form, at rock bottom level, of the distinction we wish to make here.
> (MacCorquodale and Meehl 1948, 96)

This example may seem to suggest that intervening variables are theoretical terms without existential status, whereas hypothetical constructs are theoretical terms with existential status. However, this view does not capture the essence of the distinction that was discussed in the 1950s. Although the realistic interpretation of terms was one of the issues that played a role in the debate, the view that the surplus meaning of theoretical terms concerns their existential status is a specific interpretation of surplus meaning. Surplus meaning is not always existential. For instance, below I will discuss the ideas of Lindzey (1953, 28) who argued that hypothetical constructs do not have to be interpreted realistically.

MacCorquodale and Meehl, who discussed three ways of stating the distinction they had in mind, related the distinction between intervening variables and hypothetical constructs to what Reichenbach (1938) called "surplus meaning." First, the statement of a hypothetical construct, as distinguished from an intervening variable, contains terms "which are not explicitly defined by … empirical relations." In other words, intervening variables can be defined completely in terms of observables, whereas hypothetical constructs can only be partially defined in terms of observables. The meaning of hypothetical constructs cannot be wholly captured by means of operational definitions. They have additional meaning, that is, surplus meaning. Second, for sentences containing only intervening variables, the truth of the facts, that is, the observation sentences and empirical laws, constitutes "*both the necessary and sufficient conditions*" for the truth of these sentences. For sentences involving hypothetical concepts, this is "well known to be false." This is a direct consequence of the surplus meaning of the hypothetical constructs. Sentences involving hypothetical constructs cannot be reduced to empirical facts because the surplus meaning of the hypothetical constructs goes beyond empirical content. Third, in the case of intervening variables, "the quantitative form of the concept, *e.g.*, a measure of its 'amount,' can be derived directly from the empirical laws simply by grouping of terms. In the case of hypothetical concepts, mere grouping of terms is not sufficient." Intervening

variables are nothing more than what Spence (1950) called "calculational devices," which are abbreviations for certain groupings of terms. These calculational devices "have no factual content surplus to the empirical functions they serve to summarize." Because hypothetical constructs have surplus meaning, this does not obtain for them (MacCorquodale and Meehl 1948, 96–97).

As an example to illustrate the distinction between hypothetical constructs and intervening variables, MacCorquodale and Meehl mentioned the theoretical term 'habit strength,' which is defined in Hull's *Principles of Behavior* (1943b, 178–179) as ${}_{S}H_{R} = M\ (1\text{-}e^{-kw})\ e^{-jt}\ e^{-ut'}\ (1\text{-}e^{-iN})$. Habit strength (${}_{S}H_{R}$) is a certain joint function of four variables, namely the number of reinforcements, the delay in reinforcement, the amount of reinforcement, and the asynchronism between the discriminative stimuli and response (N, t, w, t' respectively). The other symbols in this formula (M, e, k, j, u, and i) are constants. If this definition is taken as capturing the whole meaning of 'habit strength,' then it is a convenient grouping of terms, and thus an intervening variable. Of course, Hull could also have chosen another way to group the terms. According to MacCorquodale and Meehl, if 'habit strength' is indeed an intervening variable instead of a hypothetical construct, then the question if it really *exists* is equal to the question "whether we have formulated a 'correct statement' concerning the relations of this intervening variable to the anchoring (empirical) variables." This is, in turn, equivalent to the question if the empirical variables are "related in such-and-such a way." To confirm or disconfirm this is a direct empirical matter (MacCorquodale and Meehl 1948, 98). This means that if theoretical terms such as 'habit strength' are interpreted as intervening variables, and thus merely as calculational devices, the proof of their existence boils down to the proof of the correctness of the empirical relations that are obtained from writing out these calculational devices. However, MacCorquodale and Meehl called into question if the meaning of 'habit strength' in Hull's work is indeed captured by the joint function of four variables or it means "something more of a neural or other physiological nature." In that case, "the theory could be false even if the empirical relations hold" (MacCorquodale and Meehl 1948, 99).

At first sight, the theoretical terms, such as 'habit strength,' with their formal definitions in Hull's *Principles of Behavior* seem to be

calculational devices. Therefore, it seems that Tolman and Hull agreed on the use of theoretical terms (MacCorquodale and Meehl 1948, 101). This observation is in line with the analysis by Bergmann and Spence that the methodologies of Tolman and Hull are identical. However, MacCorquodale and Meehl argued that Hull's formal definitions do not capture the whole meaning of the concepts he used. In Hull's scientific writings, the formal definitions of theoretical concepts were often supplemented with statements referring to mechanisms. As I discussed above, due to this practice, which Spence had criticized in his letters to Hull, the theoretical terms acquired surplus meaning. According to MacCorquodale and Meehl, this implies that, despite the alleged similarities, there is a fundamental difference between Tolman's and Hull's respective uses of theoretical concepts, namely that Hull hypothesized the existence of unobserved entities and the occurrence of unobserved events. Hull's constructs have a "surplus meaning that is existential" (MacCorquodale and Meehl 1948, 106):

> There are various places in Hull's *Principles* where the verbal accompaniment of a concept, which in its mathematical form is an intervening variable in the strict (Tolman) sense, makes it a hypothetical construct. (MacCorquodale and Meehl 1948, 101)

The article by MacCorquodale and Meehl (1948) caused a lively debate among theoretical psychologists and logical positivists about intervening variables and hypothetical constructs. The difference between these two subclasses of theoretical terms was seen to be important, and hundreds of papers were written on this topic (Furedy 1988, 74). In this debate the questions if Tolman's use of theoretical terms indeed differed from that of Hull, if the use of hypothetical constructs is to be favored over the use of intervening variables, and if some kinds of surplus meaning are preferable over others were explored. I will analyze some contributions to this debate and shed new light on it by arguing that the focal point in the debate is the value of intelligibility. Neo-behaviorists, who, following logical positivism, considered verifiability as the criterion for assessing scientific theories, implicitly came to value the intelligibility of models and theories. Thus, the debate about the surplus meaning of theoretical terms should be understood as dealing with the question how to combine the positivist attitude towards scientific theories with the demand for intelligible theories.

4.3.2. *The Merits of Using Theoretical Terms with Surplus Meaning*

MacCorquodale and Meehl believed that the use of hypothetical constructs was preferable to the use of intervening variables. They argued that statements with only intervening variables could be translated or reduced to statements that contain nothing but empirical terms and relations. This is "to be sure, what Tolman's original definition implied." But it excludes "extremely fruitful hypotheses . . . for which the strict reducibility does not exist" (MacCorquodale and Meehl 1948, 101). A scientist who wants to make progress should use hypothetical constructs. In addition to MacCorquodale and Meehl, others promoted the use of hypothetical constructs as well. Because MacCorquodale and Meehl had characterized Tolman's theoretical terms as exemplary intervening variables, it was surprising that in this debate Tolman also encouraged the use of hypothetical constructs instead of intervening variables. After the publication of MacCorquodale and Meehl's paper, Tolman was quick to make clear that he did not agree with the ideas about the intervening variables that were ascribed to him. He did not think that scientists should only use intervening variables. Instead, he recommended that psychologists develop physiologically inspired brain models containing hypothesized neurological entities. He admitted that he made "such a statement with surprise" because for many years he had objected to what he called "premature neurologizing" (Tolman 1949, 48):

> Nevertheless, . . . I have come to realize that a tentative, hypothetical brain model is in fact inevitable and to be desired. Further, I would now coin the term "pseudo-brain models" to cover such hypotheses, not to cast aspersions on them or to suggest that such models should not be as solid and as consonant with known neurology as possible, but rather to empathize that if they are to be comprehensive enough for *our* purposes they must at the present time be very speculative, very tentative, and often only pseudo neurological in character.
>
> (Tolman 1949, 48)

In Tolman's plea for physiologically inspired hypothesized entities he explicitly mentioned the use of models. To work with hypothesized entities amounts to building a model. He recommended the use of models in science because they can be fruitful in the search for new relations and predictions:

> But the important point, as I see it, is that a model (whether almost good physiology or almost wholly "pseudo") has certain specific intrinsic properties attributed to it by its authors. And it is by following out the consequences of these attributed properties that one is led to wider predictions … A model provides a conceptual substrate – a substrate which is endowed by its author with certain intrinsic properties of its own. And, if the model be a happy one, then we are led by it to expect new behavioral relations which we would probably otherwise never have thought of. (Tolman 1949, 48–49)

In the interpretation of MacCorquodale and Meehl, intervening variables can be used only for the description of empirical relations and not for the prediction and discovery of novel empirical relations. This changes when the intervening variable is integrated into a model. However, by doing this, the intervening variable acquires the interpretation of a hypothesized entity, and thus transforms into a hypothetical construct:

> Or to put the whole matter another way, I am now convinced that "intervening variables" to which we attempt to give merely operational meaning by tying them through empirically grounded functions either to the stimulus variables, on the one hand, or to the response variables, on the other, really can give us no help unless we can also imbed them in a model from whose attributed properties we can deduce new relationships to be looked for. That is, to use Meehl and MacCorquodale's distinction, I would now abandon what they call pure "intervening variables" for what they call "hypothetical constructs," and insist that hypothetical constructs be parts of a more general hypothesized model or substrate. (Tolman 1949, 49)

Tolman emphasized that scientific theories are more than merely a collection of empirical relations. If intervening variables are the only constructs used in a scientific system, then this system is not a theory but merely a collection of empirical relations – or an "unconscious" theory, as he wittily called it. A "conscious" theory has to contain hypothetical constructs, and, as such, it is a hypothesized model with certain specific intrinsic properties attributed to it. Only the use of such a "conscious" model is fruitful and can lead to new predictions:

> [T]his demand for models may … seem to be asking too much. Any model, we can suggest now, must inevitably be inadequate and but sketchily related to the meager sets of empirical evidence now at hand. Nevertheless, my final plea is that conscious, even though bad, theory

> is better than unconscious theory. We all do have models in the back of our heads when we collect systematic sets of data. Therefore, we should stick our necks out so that both we and others can see what these models are. For I have faith that the more explicit we can become about our theories, the more fruitful they will be, and also the more likely it is that we will be ready to drop or change them when they no longer prove useful. Or, to put it another way, the sin of "rigidity" is, I believe, more apt to function in the unconscious than in the conscious.
>
> (Tolman 1949, 50)

In my view, Tolman's plea for "conscious" models is, implicitly, a plea for intelligible models. Tolman's reason for rejecting "unconscious" models, which consist only of intervening variables without surplus meaning, is that they are not fruitful. In other words, Tolman regarded models as fruitful if they provide a conceptual substrate that suggests their application to (new) phenomena. In chapter 5, especially in section 5.3.4, I will discuss the importance of the conceptual structure of models for their intelligibility. I will argue that providing such a conceptual substrate is a characteristic feature of intelligible models. Therefore, Tolman's plea for "conscious," and thus fruitful models is in line with my view that fruitfulness is a requirement for the intelligibility of models (in the sense that only fruitful models can be assigned the value of intelligibility). In my account of intelligible models, Tolman's requirement that models should provide a conceptual substrate that suggests their application does not yet guarantee their intelligibility. In addition, the suggested application should be successful – which might be what Tolman means by the situation that the model is a "happy" one. In this interpretation, Tolman's plea for "conscious" models is, implicitly, a plea for intelligible models.

It is clear that Tolman's ideas about theories underwent an extraordinary development. As he wrote in his autobiography (Tolman 1952, 335), he changed his view from conceiving a theory as a set of 'intervening variables' to the contention that all theories use 'hypothetical constructs.' In light of my view of scientific understanding, this is a positive development. However, not all neo-behaviorists agreed with Tolman's new ideas. First, the question was raised if Tolman was right that hypothetical constructs should be used in science, and, second, if he was right that the surplus meaning had to be physiological in character. I will discuss both issues. I will start with the first issue, about

the merits of theoretical terms that possess surplus meaning, and in the subsequent section I will discuss the nature of the surplus meaning of psychological concepts.

A neo-behaviorist who argued against the use of hypothetical constructs was Marx. According to this psychologist, who worked at the University of Missouri, Columbia, it was "especially discouraging to find Tolman, whose introduction of the intervening variable contributed notably to the establishment of the recent operational trend, now apparently reversing his earlier position" (Marx 1951, 235). In his view, the "improved scientific sophistication ... evidenced by psychological theorists has been largely characterized by an increased sensitivity to the need for operational validity in the formation and use of logical constructs" (Marx 1951, 235). Despite this, he noticed a tendency by several neo-behaviorists to apply constructs of the hypothetical type and to minimize the value of the operational type. According to him, the reason for this was that physiological models, which include hypothetical constructs, were considered to be useful guides in the development of science (Marx 1951, 239–240). Although Marx admitted reluctantly that such guides were necessary, he argued that the benefits of hypothetical constructs should not be at the expense of objectivity. In other words, although he acknowledged the heuristic importance of the use of the surplus meaning of theoretical terms, he did not allow that to interfere with a positivist attitude towards scientific theories. As a solution to this dilemma, he proposed a special labeling technique for intervening variables:

> [I]n deciding which verbal label to give this intervening variable, we may draw upon our own informal observations or upon some particular theoretical framework. This use of the intervening variable technique thus makes it possible not only to give a purely operational meaning to the construct used, but also to relate them to some prior observations or theoretical systems in a way that should help to move these constructs in the direction of a more clear-cut operationism.
>
> (Marx 1951, 243)

His idea was to label the intervening variables in such a way that such labeling contributes to the semantic clarification of psychological language.

As an example, he discussed two situations. The first concerned the situation of animals in a cage in which they have access to food

and receive electric shocks from which they cannot escape. The second concerned the situation of animals in a cage in which they have access to food and receive electric shocks from which they can escape by jumping off a grid. It appears that there is a relation between the restraining conditions of the animals and the amount of food they consume: they consume more in the cage in which they can escape the shocks. Marx suggested that the amount of food that the animals consume could be described as a function of the "sense of helplessness," which is an intervening variable that he defined purely in terms of the restraining conditions of the animals. On the one hand, this intervening variable functions as a calculational device that is useful for the description of the functional relation between the restraining conditions and the amount of food consumed. On the other hand, because of its name, it suggests a hypothesis that might explain the different amount of food consumed in the two situations. Therefore, according to Marx, this naming technique "seems to combine the best features of both the hypothetical construct and the orthodox intervening variable. That is to say, it offers the experimenter an opportunity to draw upon the suggestions of a theoretic model and yet remains on a strictly operational level of discourse" (Marx 1951, 243–245).

On the one hand, Marx demanded that theoretical terms remain free of surplus meaning, which means that their meaning is completely covered by their operational definition. On the other hand, he recommended labeling them so that they are extensible to novel situations. This seems impossible because the applicability of the concepts can only be extendable to domains other than those in which the concepts are defined if the definition of the concepts is partial or "open." The heuristic advantage that Marx expected from the use of his labeling technique for theoretical terms is similar to the fruitfulness that Tolman expected from the use of theoretical terms that possess surplus meaning. Both expectations are about the extensibility of the domain of application of the models that contain these theoretical terms. My interpretation of Tolman's remarks about "conscious" theories is that this extensibility of the domain of application is facilitated by the surplus meaning of the theoretical terms in the models that renders them intelligible to their users. Therefore, I submit that the dilemma that Marx tries to solve with his labeling technique can be seen as the value conflict between the value of intelligibility and typical positivist

values such as objectivity and empirical adequacy. This value conflict is a crucial issue in the debate about intervening variables and hypothetical constructs. Although Marx seems to regard the positivist values as superior to other values, by promoting his labeling technique he indirectly admitted the importance of intelligibility, which is actually in line with Tolman's plea for "conscious" theories.

That Marx's labeling technique will not work follows, for instance, from the reflections on "verbal magic" in an article with the intriguing title "Do Intervening Variables Intervene?" by Maze (1954) that attracted widespread interest. Its author, a theoretical psychologist at the University of Sydney, was influenced by the realist views of the philosopher John Anderson who worked at the same university and whose theory that whatever exists is a spatiotemporal situation or occurrence flourished in several universities in Australia (Furedy 1988, 71). In this article, Maze focused on surplus meaning that is existential and showed that the use of a labeling technique already introduces such surplus meaning. He argued that, due to their use of the relation between intervening variables and empirical variables, MacCorquodale and Meehl were implicitly ascribing surplus meaning to the intervening variables:

> [T]o ask about "its" relation to anything is to treat it as being the sort of thing that can be a term of a relation – that is, as being qualitative, as being some state or condition or "stuff" that there can be quantities of. Such a notion must always have "surplus meaning"; that is, a term of a relation must have some nature, some collection of properties, other than its having that relation; otherwise it would be unintelligible to say that *it* had that relation. What would "it" refer to? Now, such a conclusion is precisely what MacCorquodale and Meehl want to avoid, but it is entailed by their speaking of "its relation to the empirical variables."
>
> (Maze 1954, 227)

According to Maze, MacCorquodale and Meehl were not alone in implicitly treating the intervening variable as being some state, condition, or entity. Hull did the same when he spoke of the requirement of anchoring variables such as 'habit strength' at both ends:

> We have to be clear, then, that the empirically found mathematical relations are not between (for example) ${}_{S}H_{R}$ and its antecedents on the one hand, and between ${}_{S}H_{R}$ and its consequents on the other, but just between the antecedent and consequent events – in fact, that that relationship is just what ${}_{S}H_{R}$ *is*. (Maze 1954, 228)

Hull's use of the term reveals that he regarded 'habit strength' as more than merely a calculational device that was convenient for manipulation. As discussed above, the definition of 'habit strength' consists of four factors. Maze argued that "[i]f convenience of manipulation is the only concern, it must be very difficult to show why that particular four out of all these factors should be taken together, and why their product should be given a special name." Hull did not group them for the sake of convenience but on the basis of the "material consideration" (Maze 1954, 229) that "it is hard to believe that an event such as a stimulation in a remote learning situation can be causally active long after it has ceased to act on the receptors" (Hull 1943a, 285). Although Hull (1943a, 285) claimed that "it is perfectly possible to put into a single equation the values of events which occur at different times," those past events cannot be causally active now, and 'habit strength' "is merely a quantitative representation of the preseverative after-effects of the no-longer-existent compound events" represented by four factors in the definition of 'habit strength':

> Hull, then, groups the variables in this regular way not merely as a matter of convenience in calculation (if indeed one way could be more convenient than another), but also because he regards the variables in any group as acting together to build up some specific condition in the animal, and the intervening variable based on that group would then be thought of, if not actually as a "measure" of that condition, at least as varying quantitatively in direct relation with it. (Maze 1954, 229)

The intervening variable "*becomes* a hypothetical construct in the ascription *to it* of qualitative content" (Maze 1954, 233). These hypothesized states are thought of as mediating between temporally remote stimuli and responses. Maze called this the fallacy of verbal magic, "i.e., giving a name to a certain kind of event and then using that name as if it accounted for the *occurrence* of that kind of event" (Maze 1954, 226). Marx's labeling technique is an example of this fallacy of verbal magic. Labeling a variable that is defined purely in terms of the restraining conditions of animals as "sense of helplessness" suggests the existence of an internal state of the animals that explains their behavior. Therefore, Marx is wrong in claiming that his labeling technique allows the theoretical terms remain free from surplus meaning. The name "intervening variable" already suggests causal interaction: "the words themselves inevitably suggest some state-like thing that *intervenes between*

stimulus and response" (Maze 1954, 233). In sum, verbal magic is a source of the surplus meaning of theoretical terms. Due to this verbal magic, it is not possible to use intervening variables in science without supplying them with surplus meaning.

The view that intervening variables acquire surplus meaning when they are used in science was also supported from a very different angle. Bergmann, who together with Spence more than a decade earlier, already claimed that there is no essential difference between Tolman's and Hull's use of theoretical terms (Bergmann and Spence 1941), reaffirmed this claim by objecting to the division made by MacCorquodale and Meehl between Tolman's intervening variables and Hull's hypothetical constructs (Bergmann 1953). According to Bergmann, in scientific practice all terms have surplus meaning (or "excess meaning" as he called it). He argued that intervening variables, which MacCorquodale and Meehl "discounted as trivial and arbitrary, in the sense in which mere abbreviations are arbitrary" (Bergmann 1953, 441–442), are brought into science as the components of theories, and theories always carry surplus meaning in that they predict laws quite different from those on which they are based. Therefore, all theoretical terms in science have excess meaning. In fact, for very strict neo-behaviorists such as Burrhus F. Skinner, the inevitability of theoretical terms having surplus meaning was a reason to condemn the use of theoretical terms in science because it leads to unwarranted predictions and explanations (Greenwood 1999, 5). According to Bergmann (1953, 446), it follows that "as soon as the intervening variables of behavior theory are put to use they acquire automatically excess meaning. Why, then, are they now dismissed as 'mere abbreviations'?" Bergmann (1953, 446–447) argues that because in practice surplus meaning cannot be a property that distinguishes hypothetical constructs from intervening variables, the distinction made by MacCorquodale and Meehl is a "pseudo-distinction," and, in addition, "the whole controversy of intervening variables versus hypothetical constructs is a pseudo-issue."

In sum, when theoretical terms are "put to use" (Bergmann 1953, 446) in scientific models, they acquire surplus meaning. In this respect there is no difference between the terms used by Tolman and Hull, which is in line with the analysis of Bergmann and Spence (1941). Inevitably, the theoretical concepts advocated by both Tolman and Hull

have surplus meaning. However, whereas Bergmann considers the surplus meaning of theoretical terms to be a result of their application in scientific models – the terms acquire it when they are "put to use" (Bergmann 1953, 446) – I suggest that it is often the other way around: the surplus meaning is often the driving force behind the application of models in science. The surplus meaning of the theoretical terms in the models renders the models intelligible, and this enables their users to apply them to phenomena.

Several participants in the debate about intervening variables and theoretical constructs openly appreciated the added value of the surplus meaning of theoretical terms. I submit that the underlying reason for this was that they implicitly came to appreciate intelligibility as an important scientific value. In chapter 3, intelligibility was described as a positive value that scientists attribute to a model's virtues that facilitate their use of the model. Tolman's plea for "conscious" models indicates that the appreciation for models containing theoretical terms with surplus meaning was based on the insight that this surplus meaning facilitates the successful application of these models. Containing theoretical terms that possess surplus meaning was regarded as a virtue of a model that facilitates its use. In other words, although the participants in the debate did not explicitly put it that way, models that had this virtue were appreciated for their intelligibility.

The debate in the 1950s illustrates that the use of theoretical terms with surplus meaning became appreciated among neo-behaviorists and logical positivists, as was the case with Bergmann. However, the participants in the 1950s debate disagreed about the type of surplus meaning that is appropriate for theoretical terms in psychology. For instance, some argued that their nature should be neuro-physiological, while others argued against this and supported other types of hypothetical constructs. Apparently, among the neo-behaviorists there was disagreement about how the value of intelligibility should be incorporated into the particular context of their discipline.

4.3.3. *The Nature of the Surplus Meaning of Psychological Concepts*

In Tolman's plea for "conscious" theories, he argued that the hypothetical constructs should be (tentatively) neurological in character. This was contested by, for instance, Lindzey (1953), a social psychologist

who became widely known as the editor of *The Handbook of Social Psychology* (1954). Although he supported the use of theoretical constructs that are more than mere shorthand designations of observed empirical relations, he questioned if such a construct "must possess existential status" (Lindzey 1953, 28). In particular, he asked if it was necessary to employ only constructs that seem to promise eventual linkage with physiology and neurology, as advocated by Tolman and also by Krech (1950), a professor of psychology at the University of California at Berkeley and a follower of Tolman:

> Much of this emphasis has come from quarters where one might least expect it and may be considered evidence of a strong swing in the direction of forcing integration between psychological theory and physiological theory and observation. In particular, both Tolman and Krech have pointed to physiology and neurology as a *necessary* area of consideration for the psychological theorist. (Lindzey 1953, 29)

Lindzey strongly disagreed with this position and objected to the notion that "all psychological theorists are bound to examine their concepts in the light of physiological evidence" (Lindzey 1953, 30). Although Lindzey admitted that physiology could be useful for psychology, he reduced its usefulness to a possible heuristic advantage for the formation of theories and stressed that physiological plausibility is not a reason to assess theories positively:

> The eventual test of whether the influence upon his theorizing has been beneficial or baleful, however, will be independent of physiological data. That is, the theory will be evaluated not in terms of how well it represents physiological data but rather in terms of how efficiently it is able to control behavioral data. I would suggest consistently that the person who wishes to, may turn to Dostoevsky, Shakespeare, Faulkner, or any other literary or graphic artist for suggestions or inspirations as to the most fruitful means of representing behavior. (Lindzey 1953, 31)

However, the possible heuristic advantage was not the only reason for psychologists such as Krech to plead for the use of physiologically inspired hypothetical constructs. They considered it essential that hypothetical constructs refer to real entities and processes, which they expected to be of a neurological or physiological kind. For instance, Krech (1950, 283–284) argued that it is obvious to ask where these hypothetical constructs are located, and he claimed that the only

legitimate answer is that these constructs should be viewed as neurological events. This is so because if it is said that they "are in the psychological field, are psychological processes and can be studied by psychological analysis," then "we are forced to place all would-be hypothetical constructs in a sort of never-never land – a domain which is forever inaccessible to scientific inquiry" (Krech 1950, 284). The reason for this is that hypothetical constructs are not shorthand terms for certain behavior, as intervening variables were considered to be, but determinants of behavior that cannot be directly inferred from observations. Merely studying the immediate data of psychology – stimulus and response – cannot reveal the intrinsic attributes of these determinants of behavior. And because "psychological psychologists" deny that the hypothetical constructs are names for neurological activities, the possibility of an objective physiological study of the intrinsic attributes of these hypothetical constructs is also rejected. Therefore, "such an answer does not even offer the slightest guess or hunch or even fantasy of how the scientist is *ever* to get beyond the study of correlations between … stimulus and response" (Krech 1950, 285).

Lindzey did not agree with this. He held that if the behavioral consequences of psychological models can be verified, the question if the concepts in the model have a neurological correlate becomes irrelevant. He even approved the use of concepts that "violate physiological data," such as the psychoanalytic notion of the libido. This notion, if given existential status, "implies a rough sort of plumbing system within the body for the transmission of nervous impulses," and such a system is contradictory to all evidence and belief in the field of neurology (Lindzey 1953, 32). Still, Lindzey (1953, 32) insisted upon retaining it as long as it assisted in producing more verifiable behavioral consequences than alternative concepts that might be more congruent with physiological data, because the only criterion for evaluating theories is if they agree with empirical data:

> It should not be a question of whether a theorist *must* turn toward physiology or *must* ignore it. He should be permitted to do either with equal respectability since the eventual evaluation of the fruitfulness of his approach should hinge on grounds quite independent of the agreement of his theory with physiological data.
>
> (Lindzey 1953, 30)

According to Lindzey, it is not necessary for hypothetical constructs in psychology to be physiologically inspired. Moreover, he claimed that it is not even necessary for them to be assigned an existential status. He clarified this claim by a rhetorical question: "All theoretical constructs must be related to reality *via* empirical propositions derivable from these constructions. The question here is whether the theorist sees some direct tie to reality other than these derivable empirical consequences" (Lindzey 1953, 28).

Lindzey's response is both open-minded and narrow-minded. It is open-minded because it is open to the use of hypothetical constructs that are not physiologically inspired. His arguments for not restricting psychology to the use of physiologically inspired hypothetical constructs and for disregarding the existential status of their constructs are to the point, although I would rephrase them in terms of intelligibility, rather than fruitfulness. A concept such as the libido may be useful in scientific accounts of certain behavior if it renders those accounts intelligible to its users, which means that they are able to apply that account in specific situations such that they understand the behavior. Here it does not seem necessary for the concepts to have a physiological nature, and their existential status may also be irrelevant.

At the same time, Lindzey's response is narrow-minded because it reduces the value of intelligibility to a heuristic advantage in the development of theories. In his view, for the assessment of theories only their empirical adequacy matters, and it seems that he regards the surplus meaning of the terms to be unrelated to the derivation of their empirical consequences. This view is similar to Marx's, who regarded the value of intelligibility to be subordinate to positivist values. However, it is incorrect. As I have shown, especially in my discussion of Hull's application of the term 'habit strength,' intelligibility – and thus the surplus value of theoretical terms – plays a significant role in accounting for the phenomena.

I contend that an underlying motivation for the debate among neo-behaviorists and logical positivists around 1950 was their implicit dissatisfaction with the logical-positivist view of science that failed to recognize the value of the intelligibility of scientific models. The debate can be understood as the search for an answer to the question how the value of intelligibility had to be incorporated into neo-behaviorist theories. An important issue in this debate was if the hypothetical

constructs used in psychology should be of a specific nature. An answer to this question lies partly in the transition from neo-behaviorism to cognitive psychology. This transition has been interpreted in several ways, for instance, as a Kuhnian paradigm shift, as a movement from an instrumentalist to a realist conception of psychological theory, or as a continuous evolution out of neo-behaviorism (Greenwood 1999, 1). I agree with John D. Greenwood that this transition is best represented in terms of the replacement of intervening variables by hypothetical constructs that possess cognitive surplus meaning. These hypothetical constructs consisted of theoretical terms from information theory, which was developed in the same period. Apparently, the early cognitive psychologists believed that the use of this kind of construct resulted in intelligible psychological models. In the next chapter I will examine the early days of cognitive psychology and investigate the merits of using theoretical terms from information theory. I will analyze the virtues of the models containing theoretical terms with cognitive surplus meaning, and I will explore the skills that were required to use these virtues and apply the informational models to psychological phenomena.

4.4. *The Epistemic Significance of Surplus Meaning*

The aim of this chapter has been to demonstrate, by means of a case study of scientific practice, that intelligibility is an epistemic value that is constitutive of science. The account of understanding used in this chapter is based on the idea that merely "possessing" relevant theoretical knowledge is not enough for the scientific understanding of phenomena. In addition, one should be able to *use* this knowledge, which implies applying it to concrete cases by means of intelligible models.

Both Tolman and Hull formulated theoretical models of behavior that were intelligible to them due to the surplus meaning of the theoretical terms in their models. For instance, Tolman supplied the theoretical terms such as 'demand' and 'hypotheses' with surplus meaning that originated in his pre-scientific understanding of rats. Although Tolman argued in the presentation of his methodology of operational behaviorism that this surplus meaning should be removed, he later recognized that the use of hypothetical constructs is "inevitable and

to be desired" (Tolman 1949, 48). The surplus meaning of the theoretical terms in the psychological models has epistemic significance because it renders the models intelligible to its users. Only because of the surplus meaning is it possible to apply these models to concrete phenomena. The epistemic content of the models therefore depends on their intelligibility.

In the 1950s the neo-behaviorists started openly to appreciate the use of theoretical terms that possess surplus meaning. In my view, an explanation of this development was the growing appreciation of intelligibility as an epistemic value of science. In a debate among neo-behaviorists and logical positivists several merits of using terms with surplus meaning were mentioned, such as its fruitfulness. Although these merits were not explicitly related to intelligibility, my analysis of the debate, in which I use the account of scientific understanding developed in chapter 3, shows that the underlying reason for their appreciation of the use of models containing theoretical terms with surplus meaning was that it rendered these models intelligible to them.

My analysis of the 1950s dispute about theoretical terms and their surplus meaning is mainly explanatory: it sheds light on a historical development in psychology. However, it is also evaluative. In chapter 2 I argued that one of the normative tasks of philosophy of science is to give evaluative appraisals. The different opinions about the use of surplus meaning in neo-behaviorism that were put forward by the participants of this debate can be assessed by asking if they meet the characterization of science developed in this study. According to this characterization, the intelligibility of models is an epistemic value. Because the case study in this chapter shows that the intelligibility of theoretical models depends on the surplus meaning of their theoretical terms, one of the norms for evaluating the different opinions about the use of surplus meaning is the extent to which they are in agreement with the view that the use of hypothetical constructs is to be preferred over the use of intervening variables.

For instance, in 1949 Tolman publicly declared his conversion to the use of hypothetical constructs. Not all of his colleagues appreciated this. Especially those who expressed an extremely positivist attitude towards science, like Marx, rejected the use of hypothetical constructs and proclaimed that intervening variables that do not possess surplus meaning are "the only kinds of constructs admissible in

sound scientific theory" (Marx 1951, 246). Evaluating the positions of Tolman and Marx using the characterization of science developed in this study results in an endorsement of Tolman's conversion and a rejection of Marx's reaction. Tolman's conversion showed that, unlike Marx, he was aware of the value of intelligibility in science. If it does not use the surplus meaning of theoretical terms, neo-behaviorism will only produce "unconscious" theories. To avoid this, psychologists have to develop intelligible models, which according to Tolman are models that provide a conceptual substrate that facilitates their application. His argument for "conscious" models is in line with the idea that intelligibility is one of epistemic values that are constitutive of science.

My primary motivation for investigating neo-behaviorism was its extraordinary positivist inclination. The neo-behaviorist pursuit of developing an objective science of psychology – which in the end would leave no room for intelligibility and understanding – can be seen as a test case of the philosophical thesis that understanding plays a fundamental role in science. The analysis of the scientific practice of the neo-behaviorists has revealed that, despite their positivist attitude, they came to realize that the use of theoretical terms that have surplus meaning is inevitable and even desirable. Implicitly, they began to appreciate the use of intelligible models. Their (implicit) adoption of intelligibility as an epistemic value corroborates the claim that understanding is more than a by-product of theorizing: it demonstrates that understanding has epistemic significance.

CHAPTER 5

Skills for Understanding: Cognitive Psychology

5.1. *Introduction*

In chapter 3 I presented a theoretical framework for my account of scientific understanding, based on the idea that understanding is the ability to apply a model successfully to a phenomenon. A prerequisite for the successful application of a model by a scientist is that the model be intelligible to this scientist. This requires that the scientist possess certain skills and the model certain virtues, such that the combination of skills and virtues facilitates the successful application of the model (cf. De Regt 2004; De Regt and Dieks 2005). The successful application of a model to a phenomenon is a pragmatic activity that involves the skillful act of making assessments concerning relevant similarities and the skillful act of reasoning via the model.

The main aim of the case studies is to articulate the proposed theoretical framework. In chapter 4 the focus of the case study of neo-behaviorism was on the epistemic significance of the intelligibility of models. In the present chapter, the focus of the case study of cognitive psychology will be on the skills that are required for the successful application of a model to a phenomenon. The reason for looking at cognitive psychology is not only a matter of chronological order, given that it is the successor of neo-behaviorism. Also – and more importantly – the case is illuminating because, compared with the neo-behaviorists, cognitive psychologists were much more explicit about the use of models in their scientific endeavor.

Cognitive psychology is one of the disciplines of cognitive science. It includes topics such as memory, language, imagery, and attention (Baars 1986, 147, 180). Other disciplines that are part of cognitive science are linguistics, artificial intelligence, and the neurosciences. In her voluminous book of more than 1600 pages on the history of cognitive science, Margaret A. Boden (2006, 12) describes cognitive science as "the study of mind as machine," for "the core assumption is

that the same type of scientific theory applies to minds and mindlike artefacts." This interdisciplinary study of mind was informed by theoretical concepts drawn from information theory, which was developed shortly after World War II. I will focus on the use of information-theoretical models to explain psychological phenomena and investigate the skills required for the successful application of these models. In section 5.2 I will briefly describe the rise of cognitive science. In section 5.3 I will investigate how early cognitive psychologists such as George A. Miller applied information-theoretical models to cognitive phenomena. I will show that the application of these models was a non-trivial achievement that involved the use of skills and metaphors. In section 5.4 I will look in detail at a specific example of the application of the information-theoretical approach in cognitive psychology, namely the work of Broadbent on the phenomenon of attention. In section 5.5 I will present the results of the analysis of the skills that are involved in understanding cognitive phenomena.

5.2. *A Brief Review of the Rise of Cognitive Psychology*

An event that marked the rise of cognitive science and the decline of behaviorism was the Hixon symposium on "Cerebral Mechanisms in Behavior" in September 1948 at the California Institute of Technology (Gardner 1985, 10). The scientists that gathered at this symposium were very distinguished scholars from several scientific disciplines. In the opening talk the mathematician John von Neumann discussed an analogy that would become one of the most inspiring ideas in cognitive science, namely the comparison of the human brain and a computer. In another talk the psychologist Karl S. Lashley criticized neobehaviorism, the dominant discipline at that time, by arguing that a linear succession of stimulus and response cannot explain complex human behavior that is planned and organized by hierarchical brain processes. Other speakers at the symposium who would become important representatives of the new field of cognitive science were the neurophysiologist Warren S. McCulloch and the logician Walter H. Pitts, who compared the working of the brain to a neural network, the logician Alan M. Turing, who pioneered in computer theory, the mathematician Norbert Wiener, who coined the term "cybernetics" as the science of communication and control in the animal and the machine,

and the electronic engineer and mathematician Claude E. Shannon, who was one of the founders of information theory.

Not long after this symposium, cognitive science became recognized as a separate discipline. This was due especially to a symposium on information theory in 1956 at the Massachusetts Institute of Technology and to the establishment of the *Center for Cognitive Studies* in 1960 at Harvard. At the symposium, computer scientist Allen Newell and economist Herbert A. Simon discussed their Logic Theory Machine, linguist Noam A. Chomsky presented his new approach to linguistics, and psychologist Miller delivered a famous paper in which he claimed that the capacity of human short-term memory is limited to approximately seven entries. Together with Jerome S. Bruner, Miller co-founded the research center at Harvard for the investigation of the human mind as the first institute to be dedicated to what is now called cognitive science (Schultz and Schultz 2000, 472):

> In using the word "cognition" we were setting ourselves off from behaviorism. We wanted something that was *mental* – but "mental psychology" seemed terribly redundant. "Commonsense psychology" would have suggested some sort of anthropological investigation, and "folk psychology" would have suggested Wundt's social psychology. What word do you use to label this set of views? We chose "cognition."
>
> (Miller, quoted in Baars 1986, 210)

The new discipline in psychology was very successful. After a decade it had so many adherents it could have its own journals, such as *Cognitive Psychology* (first published in 1970), *Cognition* (1971), *Memory and Cognition* (1973), *Journal of Mental Imagery* (1977), *Cognitive Therapy and Research* (1977), and *Cognitive Science* (1977) (Schultz and Schultz 2000, 482).

The cognitive shift is sometimes described as a radical turnaround (Sperry 1995, 35) and a revolt against the neo-behaviorist doctrine where positivist ideas about science were pushed too far. To put it in Ulric Neisser's (1967, 5) words, one of the main figures of cognitive science, cognitive psychology is incompatible with the behaviorist view that "man's actions should be explained only in terms of observable variables, without any inner vicissitudes at all." People joked that with behaviorism psychology had "lost consciousness" or "lost its mind" and that with the cognitive revolution it had regained it (Schultz and Schultz 2000, 468). Instead of focusing merely on

stimulus-response connections, cognitive psychologists shifted their attention to mental processes and events and used behavioral responses as sources for making inferences about mental processes. With the rise of cognitive psychology the emphasis shifted from behavior to mind. This may suggest a discontinuity in the history of psychology.

In my opinion, a more accurate view is articulated by Bernard J. Baars (1986), who in his book *The Cognitive Revolution in Psychology* describes the cognitive shift as a quiet revolution. The fact that most of the changes occurred silently is related to an important topic discussed in the previous chapter, namely the surplus meaning of theoretical terms. The meaning of theoretical terms used by behaviorists exceeded their operational definition, and this enabled them to use these terms outside the domain for which they were defined. According to Baars, this extension of the meaning of terms eventually resulted in the rise of cognitive psychology:

> Most of the milestone experiments in the cognitive revolution were published without fanfare, and changes were occurring even as their profundity was being denied. Many such changes seemed to proceed by a process of euphemism. Psychologists did not speak of "mental representation" at first, but of "memory"; not of "consciousness," but of "selective attention"; not of "the organization of meaning," but of "semantic features." In each case, the modest euphemistic term was defined operationally, by precise and reliable experiments, and the results were interpreted within narrow theoretical limits. But as the new ideas gained momentum, theoretical terms such as "mental representation," "meaning organization," and recently, even "consciousness" burst the boundaries of the experimental situations in which they were first defined. Soon, all the terms were extended far beyond any single experimental model or technique. (Baars 1986, 141–142)

In the 1950s the neo-behaviorists began to appreciate theoretical terms that had surplus meaning. In the previous chapter I argued that an underlying reason for this was that they (implicitly) came to appreciate intelligibility as an important scientific value. One of the issues in the debate among neo-behaviorists at that time was the question what kind of surplus meaning is most appropriate for psychology. Some argued for a reductionist program in psychology that used physiologically inspired hypothetical constructs. Another idea, one that would gain much support, was the use of theoretical terms from the newly

developed information theory. Psychologists came to view the brain as a kind of information processor: "The term 'cognition' refers to all the processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used" (Neisser 1967, 4). In the 1950s "it seemed natural for psychologists and neurophysiologists to investigate how information theory could be used to understand human beings" (Baars 1986, 153). In this chapter, I will describe how, in the early days of cognitive science, psychologists found inspiration in information theory in modeling the mind. By focusing on their information-theoretical accounts of behavior, I will investigate the skills that were required for the application of models to phenomena.

5.3. *Applying Information-Theoretical Models in Cognitive Psychology*

As I argued in chapter 3, the application of models to phenomena can be successful only if the scientists possess skills that match the virtues of the models. In this section I will examine the principles of information theory and investigate the skillful nature of the application of these principles in cognitive psychology.

5.3.1. *Information Theory*

Information theory has its roots in the Second World War, when communication engineers working on communication systems such as the telegraph and the telephone developed methods to measure their information capacity. The comparison of the of rival communication systems with respect to efficiency required measuring the amount of information. In 1948, Shannon, who was an engineer for the Bell Telephone Company, proposed a measure for the amount of information in his influential article "A Mathematical Theory of Communication." In this article he presented a schematic description of communication systems (see figure 5.3.1) and developed a mathematical theory about the transmission of messages over "channels."

According to Shannon, a communication system uses a *channel*, such as "a pair of wires, a coaxial cable, a band of radio frequencies, a beam of light," to transmit information from a *source* to a *destination*. As a rule, a *transmitter* transforms the messages from the information source into signals suitable for transmission over the channel.

For instance, in telephony this operation consists of converting sound pressure into an electrical current. Usually, the *receiver* on the other side of the channel performs the inverse operation in order to reconstruct the message from the signal (Shannon 1948, 380–381). If the channel is not noiseless, the signal is disrupted by noise during transmission and/or at the terminals. In that case, the received message is not the same as the initial message (Shannon 1948, 406).

Shannon argued that the meaning of the messages transmitted over the channels is irrelevant for the comparison of the channel capacity of different communication systems. The semantic aspects are irrelevant to the engineering problem. The essential question is if the receiver can use a transmitted signal to make a *choice* within a set of possible messages. The smallest unit of information is one that enables the receiver to discriminate between two alternative messages. This smallest amount of information is called a "bit." If the transmitted signals enable the choice of one message from a set of alternatives, then the number of alternatives, or any monotonic function of this number, can be regarded as a measure of the amount of information that is involved in this choice. Shannon proposed using a logarithmic function to measure information content because parameters of engineering importance, such as time and bandwidth, tend to vary linearly with the logarithm of the number of possibilities. Moreover, this logarithmic measure agrees with the intuitive feeling that, for example, "two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information" (Shannon 1948, 380). For example, in a

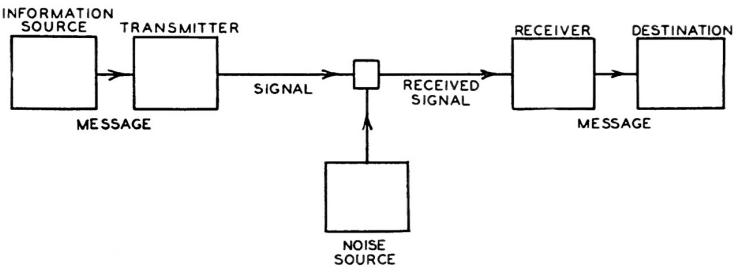

Figure 5.3.1. Schematic diagram of a general communication system (Shannon 1948, 381)

communication system in which the transmitter and receiver are used for processing of messages composed of letters from the Roman alphabet, the information content of a one-letter message, which enables the choice between the 26 letters of the alphabet, is $-\log_2 1/26 = 4.7$ bits, while the information content of a two-letter message is $-\log_2 1/26^2 = 9.4$ bits. A measure of the capacity of a communication channel, which is the maximum amount of information a channel can carry per second, is the number of bits per second that can be transmitted over the channel.

5.3.2. *Experiments on the Capacity of Human Information Transmission*

Soon after the appearance of Shannon's article, the idea emerged of using information theory in psychology. This new insight was nicely articulated at a *BBC Third Programme Talk* in 1950, in which a review was given of the *First International Symposium on Information Theory* in London. The speaker was Donald M. MacKay, who at that time was a doctoral student working on the measurement of information, and who would later change his primary field of research to the physiological organization of the brain:

> Psychologists want to know how much information could be carried by a nerve-fibre in a second, how much per second is received by the eye or the ear, and a host of other questions like that. It becomes quite precisely meaningful to ask, for example, whether a particular hypothetical model of the brain could contain as much information as we believe to be held by the real article. In fact, I believe that the ideas of information theory may make a number of real contributions to the study of the brain, which, if nothing else, is a remarkably efficient transformer of information. (MacKay 1969, 15)

At the symposium MacKay (1952) had spoken about "The Nomenclature of Information Theory." This lecture was an explanation of information (or communication) theory in which the definitions of the most important terms of information theory, such as information, channel, and capacity, were presented. This glossary would become a useful aid for the pioneering scientists who tried to apply information theory in psychology. As I will discuss below, one of them was Broadbent who, in his magnum opus *Perception and Communication* (1958) on the transformations of information in the mind, admitted that he

had not always introduced the theoretical terms properly. He anticipated critique of the "non-technical and therefore inexact" use of the terms by referring to MacKay's lecture:

> For precise definitions of this and other terms in communication theory, reference should be made to the excellent glossary given by MacKay (in von Foerster 1952). All such terms are used in this book with MacKay's definitions, though they are introduced in popular language. (Broadbent 1958, 5)

Psychologists like Broadbent, who modeled the human mind as a processor of information, explicitly claimed that in their modeling work they made use of the concepts and principles from information theory. A concrete example of how, according to the early cognitive psychologists, information theory can be used in psychology is Miller's "The Magical Number Seven, Plus or Minus Two" (1956). In this groundbreaking article, which promoted the use of information theory in psychology and would become one of the most significant landmarks in the "cognitive shift" (Baars 1986, 199), Miller introduced information theory as a basic theoretical tool for psychology and used it to investigate the structural properties and limitations built into the human mind. In his article Miller reviewed absolute judgment experiments, which are experiments in which it is tested how accurately people can distinguish differences between stimuli. Subjects are confronted with stimuli, such as sounds, that differ in loudness or lines that differ in length, and their task is to classify them accordingly, for instance by ranking them with numbers. According to Miller, these experiments, which were already quite common at the time in experimental psychology (cf. Garner and Hake 1951), are suitable candidates for an information-theoretic treatment:

> In the traditional language of psychology these would be called experiments in absolute judgment. Historical accident, however, has decreed that they should have another name. We now call them experiments on the capacity of people to transmit information. (Miller 1956, 81)

In Miller's view, the stimuli can be interpreted as signals that carry information that is processed by the subjects and transformed into a response. Therefore, "[i]n the experiments on absolute judgment, the observer is considered to be a communication channel" (Miller 1956, 82). Like all information channels, humans also have a limited

capacity. In fact it is this capacity that is measured in the experiments of absolute judgment:

> The experimental problem is to increase the amount of input information and to measure the amount of transmitted information. If the observer's absolute judgments are quite accurate, then nearly all of the input information will be transmitted and will be recoverable from his responses. If he makes errors, the transmitted information may be considerably less than the input. We expect that, as we increase the amount of input information, the observer will begin to make more and more errors; we can test the limits of accuracy of his absolute judgments. If the human observer is a reasonable kind of communication system, then when we increase the amount of input information the transmitted information will increase at first and will eventually level off at some asymptotic value. This asymptotic value we take to be the channel capacity of the observer: it represents the greatest amount of information that he can give us about the stimulus on the basis of an absolute judgment. The channel capacity is the upper limit on the extent to which the observer can match his responses to the stimuli we give him. (Miller 1956, 82)

An example of such an experiment of absolute judgment mentioned by Miller (1956, 84) is Irwin Pollack's (1952) measurement of the channel capacity of subjects that perform the task of classifying distinct tones according to their pitch. In an experimental session human subjects were presented tones selected from a series of tones whose frequencies were spaced equidistantly on a logarithmic scale. Their task was to rank the tones by assigning numbers to them. The selection of the tones was random with one statistical restriction, namely that in an experimental session the number of presentations of each tone of the series was equal. A tone was presented for about 2.5 seconds, and the interval between successive presentations was about 25 seconds. After a tone was presented, the subjects tried to assign the right number to it. They were subsequently informed about the correct identification of the tone. Then the next tone was presented. Pollack gave an information-theoretical description of this task using Shannon's model of a communication system (see figure 5.3.2.a):

> The outstanding characteristic is a defined *message-source*. The source is, in turn, made up of a defined class or *set* of possible messages and the rules for *selecting* messages from the set. The selected message is *transmitted* over a *channel* to a *receiver* and, thence, to a *destination*. In

> terms of the present study, the message-set is a defined number of sinusoidal voltages from the oscillator, the method of selection or choice is a table of random numbers, the oscillator voltages are transmitted by a loudspeaker over the air sound path to the receiver – the experimental subject – who attempts to identify the tone by assigning a numeral to it, and enters his response at the destination – the answer sheet. It is possible to obtain a quantitative measure of the amount of information presented to the experimental subject; and a quantitative measure of the amount of information lost by the subject. The difference between the information presented and the information lost is the information gained, received, or transferred by the subject.
>
> (Pollack 1952, 745)

Pollack varied the number of tones in the message set that he used in different experimental sessions. The experiment showed that when the tones were selected from a series of only two or three tones, the listeners never confused the tones. With four different tones confusions were quite rare, but with five or more tones confusions became more frequent. With fourteen different tones Pollack's research subjects – who, in the *Human Resources Research Laboratories U.S. Air Force* in Washington D.C. where he worked, were mostly military and civilian students (Lumsdaine 1953) – made many mistakes. Pollack used the results of his experiments to calculate the amount of information that the listeners were able to transmit under these conditions by taking the logarithm to the base 2 of the number of tones perfectly identified (see figure 5.3.2.b). According to Miller, the results of Pollack's experiment fit in nicely with the idea of the observer as a communication system:

> The amount of transmitted information behaves in much the way we would expect a communication channel to behave; the transmitted information increases linearly up to about 2 bits and then bends off

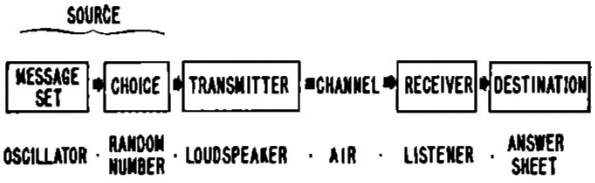

Figure 5.3.2.a. Pollack's specification of Shannon's model of a communication system (Pollack 1952, 745)

> toward an asymptote at about 2.5 bits. This value, 2.5 bits, therefore, is what we are calling the channel capacity of the listener for absolute judgments of pitch. . . . 2.5 bits corresponds to about six equally likely alternatives. The result means that we cannot pick more than six different pitches that the listener will never confuse. (Miller 1956, 84)

Miller's use of the concept of channel capacity differs to some extent from Shannon's. Instead of representing the maximum amount of information that can be transmitted over a channel in one second, it stands for the maximum amount of information that a subject can give about the stimulus on the basis of an absolute judgment (Miller 1956, 82). In Pollack's experiment, the maximum amount of information that the subjects could give about the pitch of the tones was about 2.5 bits. This result of 2.5 bits could easily be compared with the channel capacity found in other experiments of absolute judgment. Miller showed that, in general, the capacities found in these experiments fluctuate around the 2.8 bits. The "magical number seven" in the title of Miller's article refers to this phenomenon, because 2.8 bits stands for the possibility of identifying seven stimuli without making mistakes. The advantage of the information-theoretical account advocated by Miller is that it enables one to compare the results of completely different experiments of absolute judgment. This makes it possible to bring together a large amount of hitherto dispersed data and to argue that

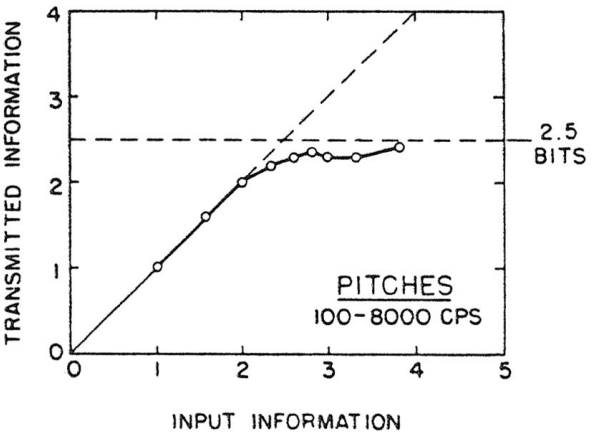

Figure 5.3.2.b. Pollack's results presented in bits (Miller 1956, 83)

they point to a common conclusion, namely, that there are limitations in human information-processing capacities. Because of these advantages, the information-theoretical approach promoted by Miller was received with enthusiasm. The hope was that these "rigorous new approaches of the engineering-oriented scientists" (Gardner 1985, 91) would allow a whole range of psychological phenomena to be understood.

5.3.3. *The Non-Trivial Application of Information-Theoretical Models*

The schematic illustration of Shannon's informational system (see figure 5.3.2.a) that Pollack used to specify the setup of his experiment gives the impression that, for him, filling in Shannon's model was simple and unambiguous. Obviously, the *transmitter* is the loudspeaker, the *receiver* the listener, and the *channel* the air in between. Scientists such as Pollack and Miller probably had no trouble at all with connecting model and phenomenon. Applying Shannon's model seems to be an easy exercise of finding a mapping between the model and the psychological phenomenon. The handwritten specification of the elements of Shannon's model in Pollack's schematic illustration of his experimental setup suggests that Pollack simply adopted Shannon's model and straightforwardly supplemented it with such a mapping. Because it seems to be a matter of course to account for the phenomenon of the 'magical number seven' by means of Shannon's model, it might seem exaggerated to argue that, from a philosophical point of view, the successful application of this model is a skillful act. One might object that it is an overstatement to argue that skills are required for the selection of relevant features of the model and the phenomenon, and for the evaluation of the possible similarities between these features. The same holds for the examples discussed in the previous chapter, where it might sound overstated to call it a skillful achievement to see a connection between, for instance, 'demand' in Tolman's model and some aspects of the behavior of rats in mazes. However, I will argue that, from a philosophical point of view, making these connections is not as straightforward as it may seem at first sight.

First, in the case of Pollack's use of Shannon's model, it can be questioned how straightforward it is to regard randomly chosen tones as signals that carry messages. In the normal use of this term a message is

produced by an intentional agent and has a meaningful content. Pollack's use stretches the meaning of what is normally understood by a message and a message source. Specifying the model therefore involves judgments about the applicability of theoretical terms. Another aspect of Pollack's use of Shannon's model that is not entirely straightforward is his choice to leave out the *noise source*. He apparently regarded the noise as negligible. However, it could be that his assessment was wrong. For instance, there could be a disturbing smell in the room that he incorrectly deemed to be irrelevant. This shows that in order to apply Shannon's model, Pollack had to make judgments concerning the applicability of terms and the relevancy of features of model and phenomenon that are not strictly based on methodological rules. In my view, making such judgments requires experience. In section 5.4.3 I will show that Broadbent, for instance, considered the application of information-theoretical models to psychological phenomena a technique that had to be mastered. Because of that, Pollack's application of Shannon's model matches Brown's (1988, 156) idea of a skillful activity. Therefore, although making a connection between Shannon's model and certain cognitive phenomena may be straightforward from a practical point of view, from a philosophical point of view it is far from trivial.

Another indication that connecting model and phenomenon is not straightforward is that in scientific practice there are differences in the way scientists carry out the exercise of specifying the model. For instance, Pollack described the listener as receiver (positioned on the right side of Shannon's model). Warren Weaver, who co-wrote a book with Shannon on the mathematical theory of communication, held a similar view. He considered the nervous system to be the receiver of information: "When I talk to you, my brain is the information source, yours the destination; my vocal system is the transmitter, and your ears and the associated eighth nerve is the receiver" (Shannon and Weaver 1949, 17). Miller, however, described the listener as the communication channel (positioned in the middle of Shannon's model). The sense organs receive information encoded in sensory perceptions, and the amount of information transmitted by the communication channels is the difference between the amount of information encoded in the stimuli and the amount of information recoverable from the responses. Broadbent held a similar view. He saw the nervous system

as a "channel for communication": "Information transmitted through the man … will be limited by the size of his nervous system" (Broadbent 1958, 41), and "a nervous system acts to some extent as a single communication channel" (Broadbent 1958, 297). Apparently, there is a divergence of views among scientists about the mapping relation between Shannon's model and the phenomenon under consideration, which indicates that the application of Shannon's model is not merely a matter of course. This divergence can have consequences for the way in which the scientists reason about the phenomena and for the way in which they design their experiments.

However, by making a distinction between the *articulation* and the *application* of a model and by introducing a distinction between *abstract* and *concrete* models, one might argue that the way in which models are specified is irrelevant for the philosophical question concerning the triviality of making a connection between a model and a phenomenon. An abstract model can be made more concrete by giving a more specific interpretation to its constituent parts by, for instance, identifying the signals in Shannon's model with randomly chosen tones and identifying the receiver with a listener. Different articulations of an abstract model result in different concrete models. If the articulation of an abstract model is seen as a preliminary task preceding the real application of a model, then a difference in how scientists articulate an abstract model does not mean that these scientists differ in the way they apply models. Even though Pollack and Miller differed in their identification of components of Shannon's model with parts of the experimental setup, this does not necessarily indicate that they differed in the way they applied models to phenomena. In this view, Pollack and Miller simply tried to apply different models. That Pollack conceived listeners as receivers whereas Miller took them to be communication channels does not indicate that connecting model and phenomenon is not a straightforward activity.

This line of thought can be continued by arguing that the way Pollack and Miller arrived at their more articulated models might be of interest for the historian or psychologist of science. But it is not of interest for a philosophical analysis of the application of models to phenomena. A further step would be to claim that, although the preliminary task of specifying the model might be a skillful act, skills no longer play a role once the model is ready for application. For instance,

although identifying randomly chosen tones with signals that carry messages is philosophically not trivial – it involves judgments concerning the extensibility of the concept of message – once the mapping relations between the model and a concrete phenomenon are specified, applying the model amounts to nothing more than straightforwardly working out its consequences in this concrete situation.

However, the view that Pollack and Miller applied different models does not do justice to the intentions of these scientists. They tried to unify the results of different experiments of absolute judgment by arguing that they were all examples of the phenomenon of the 'magical number seven.' Therefore, their intention was to apply the same model, namely Shannon's, and the idea that they applied different models is wrong: they both applied Shannon's model, although in slightly different ways. However, even if one assumes, for the sake of the argument, that they *did* apply different models, skills would still play a role not only in the articulation of the model but also in its application. Even the application of an articulated model involves philosophically non-trivial acts, such as the selection of relevant features of the model and the phenomenon and the evaluation of the possible similarities between these features.

For example, Shannon's model of a communication system has several features, one of which is that a channel has a limited capacity and another that a receiver has a prefabricated set of possible outcomes (for instance, all the letters of the Roman alphabet) that can be triggered by incoming signals. Scientists such as Pollack and Miller focused mainly on the feature of limited capacity. Shannon had developed his account of communication systems, including measuring of their limited capacity, precisely because of the need for comparing the efficiency of rival communication systems. Therefore, with respect to comparing humans with communication channels, it matters if human performance gives evidence of limitations regarding capacity similar to those of communication channels. Accordingly, both Pollack and Miller related the number of mistakes made by their subjects to the limited capacity of communication systems. They did not focus on the feature that information systems have a prefabricated set of possible outcomes because the question if humans have such a prefabricated set was not relevant to them either. More than that, they would probably even dismiss experimental sessions in which, instead of responding

in accordance with the supposed set of possible outcomes, the subject started, for instance, to question the limitations of the experimental setup. This shows that the application of models involves philosophically non-trivial judgments about relevant features.

The application also involves the evaluation of the possible similarities between the relevant features of the model and the phenomenon. By means of figure 5.3.2.b Miller showed that the behavior of the subjects in experiments of absolute judgment is similar to the behavior of communication systems with a limited capacity of 2.5 bits. The solid line represents the amount of information transmitted by listeners. The two intersecting dashed lines are drawn to compare the results with the behavior of a communication channel with a limited capacity of 2.5 bits. For an amount of input information less than 2.5 bits, the solid line should be compared with the diagonal dashed line representing the transmission of information without any capacity limitation, and for an amount of input information more than 2.5 bits, the solid line should be compared with the horizontal dashed line representing the transmission of 2.5 bits of information. The solid line does not coincide completely with the relevant parts of the dashed lines. For instance, the solid line descends slightly at 3 bits of input information. This could be a reason to argue that the relevant behavior of the subjects and the communication channel are dissimilar. However, Miller stated that the behaviors are similar. This statement was not substantiated by methodological or statistical rules of curve fitting. Apparently, he made the philosophically non-trivial judgment that deviations such as the slight decrease were irrelevant:

> The amount of transmitted information behaves in much the way we would expect a communication channel to behave; the transmitted information increases linearly up to about 2 bits and then bends off toward an asymptote at about 2.5 bits. This value, 2.5 bits, therefore, is what we are calling the channel capacity of the listener for absolute judgments of pitch. (Miller 1956, 84)

It appears that applying Shannon's model to Pollack's experimental results, which involves, for instance, posing connections between the limited capacity of a communication channel and the errors made by the listeners, requires philosophically non-trivial judgments about relevant features of the model and the experimental results and the evaluation of possible similarities between these features. That these

judgments are non-trivial becomes even more clear from critical remarks made by some cognitive psychologists a few years after the first attempts to apply Shannon's theory of information. For instance, MacKay, who at first considered these attempts as very promising and shared the hope that Shannon's information theory was of great value for psychology, later changed his view:

> Such were the hopes of 1950. It soon became clear that the biggest problem in applying Shannon's selective information measure to human information-processing was to establish meaningful probabilities to be attached to the different possible signals or brain-states concerned. After a flourish of 'applications of information theory' in psychology and biology which underrated the difficulty of this requirement, it has now come to be recognized that information theory has more to offer to the biologists in terms of its qualitative concepts than of its quantitative measures, though these can sometimes be useful in setting upper or lower limits to information-processing performance.
> (MacKay 1969, 17–18)

MacKay came to recognize that the use of Shannon's information theory in psychology poses difficulties because it is not evident how the concept of the amount of information, which is defined as a function of the number of alternatives of choice, can be applied in a psychological account of the human mind. The comparison of humans with communication systems presupposed that, like communication systems, the human mind also has a prefabricated set of possible outcomes. As MacKay (1969, 28) put it, "[t]he object of communication is to select some particular conditional readiness in the recipient from the range of states that are possible." It is questionable if it makes sense to describe the human mind as containing such an ingrained and prefabricated set of possible outcomes. To put it in Baars' words (1986, 155), "[t]he mathematical theory of information presupposed that the context of alternatives was already known; but this is precisely what psychologists needed to discover." Making a comparison between the functioning of communication systems and the functioning of the human mind, which seemed to be quite straightforward in the case of experiments such as that of Pollack, was later considered to be rather problematic. What scientists had regarded to be similar – such as the functioning of a receiver in a communication system and that of a listener in Pollack's experiment – was later seen to be dissimilar. The

primary reason for this was that they came to attach more importance to the feature of Shannon's model that a communication system has a number of prefabricated possible outcomes. It seemed unlikely that this is also the case for the human mind. Although subjects in experiments of absolute judgment could be instructed to use only a limited set of possible answers, such as a few numerals, this did not mean that their mind really functioned like a communication system with an ingrained set of possible answers. Early cognitive psychologists began to realize that if the human mind indeed functions like a communication system, the size of their ingrained set would be enormous:

> Let us consider some very simple quantitative aspect of speech. In the first place, each word is chosen from a vocabulary of fairly definite size. Basic English contains 850 words; other languages have considerably more, but Basic will do for our purpose, both because it is a definite number and also because it obviously has less than the maximum vocabulary a man can use. If a man can make a different response to any word of Basic which he hears, he must have at least 850 possible states of each part of the neural mechanism between stimulus and response. Now suppose that he hears a two-word sentence, he must have a certain number of possible states: but the number is now the square of 850, namely 732,500. A three-word sentence requires over 60,000,000 states of any mechanism which will produce an appropriate response to the sentence. … As the length of a speech message goes up there is bound to come a point at which it is drawn from a set of possibilities larger than the number of states the nervous system can take up.
>
> (Broadbent 1958, 38–39)

In the early days of cognitive psychology, psychologists such as Broadbent considered this inconceivable number of ingrained possible states that would be required for distinctive reactions to lengthy speech messages to be a reason for expecting certain limitations in the capacities of humans to process these messages (e.g. Broadbent 1958, 40). However, a few years later, cognitive psychologists such as Neisser considered the incredible number of ingrained possible states that would be required for distinct reactions – and the idea that these reactions are passive because they are preprogrammed and triggered by signals from the environment – to be a reason for arguing against the comparison of humans with communication systems. Already in his most influential book, *Cognitive Psychology*, which was a

landmark in the new psychological discipline, Neisser (1967, 175) described this view as "less congenial, because it interprets cognition as a passive rather than a constructive process." Later, inspired by James J. Gibson's theory of direct perception, Neisser (1976) would formulate his dissatisfaction with this passive view of the mind even more explicitly. Instead of a linear model like Shannon's, in which input is transformed into output, he suggested a cyclical model in which the organism is not completely passive but has a cognitive schema that directs its search for information. "The act of picking up information changes the schema, enabling it to pick up new information that in turn will change it further" (Neisser 2007, 288). Apparently, Neisser's judgments concerning the relevant similarities between Shannon's model and the human mind differed from those of early cognitive psychologists. This illustrates that these judgments, which are required for the application of models to phenomena, are non-trivial philosophically.

In sum, this inquiry into the application of Shannon's model to psychological phenomena such as the 'magical number seven' reveals that, from a philosophical perspective, it is not a straightforward matter to connect a model to a phenomenon. My case study thus illustrates that it is a skillful practice that involves making non-trivial assessments concerning relevant similarities. However, it leaves open several questions related to the non-trivial skills that are required for it. To start with, it does not reveal why, at first sight, the non-trivial application of Shannon's model appears to be elementary. Further, it does not reveal the reason why Neisser's assessments differed from those of early cognitive psychologists like Miller and Pollack. Surely all these men were competent scientists. They all possessed the skills required for making philosophically non-trivial assessments. It is not the case that, for instance, Neisser lacked the skills to apply Shannon's model or that the others lacked the skills to recognize relevant dissimilarities between the model and the phenomenon. Therefore, an explanation of the different assessments cannot be that they were not equally skilled. Answering these questions requires a more thorough analysis of the skillful application of Shannon's model in psychology. I will deal with these questions by invoking the picture that Diego Fernandez-Duque and Mark L. Johnson (1999) provide of the way in which Shannon's model of a communication system is used in psychology.

5.3.4. *The Skill of Conceptualizing and the Role of Metaphors*

Fernandez-Duque and Johnson (1999) offer an interesting analysis of the application of Shannon's model of a communication system in psychology. They argue that the early cognitive psychologists' comparison of the mind with an information-processing device is one of the central metaphors used to conceptualize psychological phenomena. According to them, the models used by the early cognitive psychologists are "conceptual metaphors." They define a conceptual metaphor, which is a central notion in their account, as "a conceptual mapping of entities, properties, relations, and structures from a domain of one kind (the source domain) onto a domain of a different kind (the target domain)" (Fernandez-Duque and Johnson 1999, 84). The source domain of the 'mind-as-information-processing-device' metaphor consists of Shannon's model of a communication system. The target domain is comprised of cognitive phenomena like that of the 'magical number seven,' visible in Pollack's experiment. In this conceptual mapping, properties of the receiver are mapped onto the ear, and properties of the signal are mapped on to a randomly chosen tone produced by an oscillator connected to a loudspeaker. Each of these sub-mappings "takes some entity or structure in the source domain and constructs a counterpart to it in the target domain" (Fernandez-Duque and Johnson 1999, 85). In this way, the metaphor conceptualizes the target domain: it adds conceptual structure to the description of the phenomenon.

Where I, in my analysis of the application of Shannon's model, talk about the act of specifying the model – for instance, by identifying signals with randomly chosen tones – Fernandez-Duque and Johnson speak of conceptualizing the target domain by means of a conceptual mapping. Like my observation that there are differences in the way scientists carry out the exercise of specifying a model, Fernandez-Duque and Johnson (1999, 86) argue that there is more than one way to map entities in the source domain on to entities in the target domain. This, however, does not imply that the mapping is ambiguous:

> [T]here is nothing at all vague about the mapping, once it has been interpreted in a particular way, depending on specific technical knowledge of information-processing devices and on a particular version of

> the mapping. It is precisely such specification of the metaphor that defines a particular theoretical and experimental viewpoint.
>
> (Fernandez-Duque and Johnson 1999, 87)

According to Fernandez-Duque and Johnson (1999, 105), the use of conceptual metaphors helps scientists develop a notion of the phenomena that has a sufficient conceptual structure to "provide adequate explanations of what [the phenomenon] is and of how it works." This aspect of the analysis of Fernandez-Duque and Johnson may help explain the non-trivial skills that are required for applying a model to a phenomenon. First, the use of conceptual metaphors facilitates the skillful activity of identifying characteristic features of the phenomenon. For instance, in the case of the 'mind-as-information-processing-device' metaphor, a characteristic feature of the source domain is the limited capacity of communication channels. Using information-processing devices as a metaphor for the mind therefore suggests that these limitations are also a characteristic feature of the target domain. This motivates the investigation of the limitations of human information processing that become apparent from the number of mistakes made by subjects in psychological experiments of absolute judgment. Accordingly, the metaphor guides the scientists in what they regard to be relevant, how they frame hypotheses, how they construct experiments, and how they interpret data (Fernandez-Duque and Johnson 1999, 111). In this view, the basis of the skills required for the judgments concerning relevant similarities is the conceptual metaphor. Metaphors are useful for singling out certain aspects of the phenomenon. However, this can also be a limitation of their use, because, as Neisser's critical remarks on the use of Shannon's model indicate, metaphors can be too restrictive. In his view, this use led to a disregard of relevant characteristics of human cognition.

Second, the use of conceptual metaphors facilitates the skillful process of reasoning about the phenomenon due to certain features of the source domain. According to Fernandez-Duque and Johnson (1999, 85), these features enable scientists to use their "knowledge of the source domain plus inference patterns drawn from the source domain to reason about the target domain." The features of the source domain that, in their account, facilitate reasoning about the phenomena, play the same role as the virtues of theories and models in the notion of intelligibility developed by De Regt (2004, 103; 2009, 31). According

to De Regt, to facilitate the skillful activity of reasoning about the phenomenon, the virtues of theories and models should match the skills of their users. In a similar vein, Fernandez-Duque and Johnson (1999, 85) argue that "the potential source domain for conceptual metaphors is highly constrained by the nature of our bodies, our brain capacities, and the environments we inhabit." They do not explain, however, how the virtues of the source domain facilitate reasoning about the phenomenon. Therefore, in section 5.4 I will look in detail at an example of the use of a conceptual metaphor and analyze the skillful process of reasoning with it.

The analysis of Fernandez-Duque and Johnson can be used to explain why, in the examples of experiments in cognitive psychology mentioned above, it appears to be straightforward how model and phenomenon should be connected. It is common nowadays to use information-theoretical concepts in descriptions of human behavior. A plausible reason for this is the dominant role of the computer in present society, which provides a fruitful source of metaphors. Familiarity with these metaphors makes it obvious to regard a listener as a receiver who decodes the information in the received signals. This would not have been obvious when information-processing devices were uncommon and other metaphors were used to account for human behavior – such as *L'homme machine* (La Mettrie 1748/1999) or the steam engine (Freud 1940, 80). In other words, at present it is common to conceptualize human behavior in terms of informational concepts, and therefore the mappings of psychologists such as Miller and Pollack seem obvious nowadays. This analysis reveals the context-dependent nature of the skills that are required for the application of models to phenomena. In order to apply a model, scientists have to be familiar with its concepts. Only then can the conceptual structure of the model be used to account for the phenomenon. The use of information-theoretical models could gain acceptance in psychology only after the psychologists developed the relevant skills to reason about information-processing devices.

This analysis does not, however, explain why Neisser, who surely had the skills to apply Shannon's model to psychological phenomena, did not regard it as contributing to his understanding of those phenomena. Neisser undoubtedly possessed the skills to work successfully with information theory. As a student at Harvard, specializing in

psychology, he already heard about it from his advisor Miller, and he followed Miller when the latter moved to a new psychology department being established at the Massachusetts Institute of Technology (Neisser 2007, 278). He was very interested in the new ideas about applying information theory in psychology and even wrote a very influential book about it:

> Sometime in the early 1960s, all of this began to jell. Perception, the span of attention, visual search, computer pattern recognition, human pattern recognition, problem solving, and remembering were all interrelated aspects of information processing. Perception and pattern recognition were input, remembering was output, and everything in between was one or another kind of processing. This was already a rather obvious idea (cf. Broadbent 1958), but no one had put it forward clearly and effectively. I could write a book!
>
> (Neisser 2007, 282)

Just like early cognitive psychologists such as Miller, Neisser was familiar with information-theoretical concepts. He was able to make judgments concerning the relevant similarities between Shannon's model and cognitive phenomena and possessed the skills to reason via the model. Yet, in contrast to the early cognitive psychologists, he did not consider the model as providing understanding of cognitive phenomena. In his view, the model had no relevant similarities with the phenomena and did not provide mechanisms that could be used to reason about the functioning of the brain:

> I am more skeptical than Broadbent about the value of information *measurement* … He argues that the cognitive mechanisms must have a finite informational capacity – in terms of bits per second – and that filtering mechanisms are needed if their capacity is not to be overloaded. This is surely true in some sense, but it does not help us understand the mechanisms in question. One might as well say that the heart, which pumps only about 100 cc. of blood per stroke, has limited capacity compared with, say a fire engine. This would also be true, but by itself would be of little help in understanding the physiology and "hemodynamics" of the heart. Perhaps it is for this reason that Broadbent's later papers (e.g., 1963) have emphasized flow charts rather than "bits." (Neisser 1967, 208)

Next to the context-dependent factor of familiarity with certain metaphors, there can be other contextual factors that influence whether scientists regard models intelligible. I submit that an important cause

for the difference in the assessment of intelligibility of Shannon's model between Neisser and the early cognitive psychologists such as Miller and Pollack was the difference in character and purpose of their investigations. The early cognitive psychologists often carried out investigations of applied psychology in the fashion of engineers because, for instance, it concerned research projects commissioned by the army, whereas Neisser, who worked as a theoretical scientist with an appointment at the university, was occupied only with theoretical psychology. Therefore, the early cognitive psychologists and Neisser were interested in different aspects of cognition. Information-theoretical accounts that use Shannon's model describe cognitive processes as passive. The early cognitive psychologists, who were mainly interested in human failures and shortcomings in the processing of information, did not consider this passivity relevant. Instead, they focused on the similarities between the capacity limitations of communication systems and humans. Neisser, however, regarded the difference between the passive nature of communication systems and the "constructive" nature of cognitive processes highly relevant. For instance, he argued that, instead of passively receiving and processing speech signals, a hearer actively constructs an internal model that matches these signals (Neisser 1967, 193–198). Neisser deemed it very important that this constructive aspect of human cognition was accounted for in cognitive psychology and he considered it problematic that it was lacking in accounts based on Shannon's model. Therefore, instead of using his familiarity with the source domain, such as his experiences with the limitations of communication equipment, to assess the relevance of aspects in the target domain, Neisser in a sense worked the other way around. He believed that cognition is constructive, which motivated him to regard the similarities with communication channels as irrelevant. In other words, he used his acquaintance with the target domain to assess the relevance of aspects in the source domain. In sum, Neisser's disapproval of the use of Shannon's model in psychology shows that in addition to the contextual factor of the availability or prevalence of metaphors, there may be other contextual factors that influence if and how scientists understand phenomena.

In my view, a shortcoming of Fernandez-Duque and Johnson's account is that the mapping between the source domain and the target

domain of a metaphor is considered to be unidirectional, going from source to target (Fernandez-Duque and Johnson 1999, 85). In their account, the source domain determines which aspects of the target domain are relevant, and which inference patterns drawn from the source domain are used to reason about the target domain. However, in my view the mapping between source and target is not strictly unidirectional. The use of the 'mind-as-information-processing-device' metaphor may influence not only the way in which the mind is understood but also the way in which information-processing devices are understood. Applying information-theoretical models in psychology, such as Shannon's model – which Fernandez-Duque and Johnson consider a typical example of the use of a conceptual metaphor – involves the act of specifying the model. This act can have effects upon the target domain, namely the addition of a conceptual structure, as well as upon the source domain, namely the stretching of the meaning of the concepts in that domain. For instance, the mapping between signals that carry messages and randomly chosen tones stretches the concept of message. This indicates that the view of Fernandez-Duque and Johnson concerning the directionality of the mapping requires adjustment.

Another shortcoming of the picture provided by Fernandez-Duque and Johnson is that it sketches the role of metaphors in science in broad outlines without offering a detailed analysis of the role of metaphors in concrete scientific practices. Fernandez-Duque and Johnson show that the 'mind-as-information-processing-device' metaphor inspired many psychologists, arguing that this metaphor guided them in their conceptualization of the phenomena. They even claim that conceptualizations are always based on metaphors. "The metaphors circumscribe the phenomena, define the basic concepts, guide the research program, and determine the inferences drawn about the phenomena" (Fernandez-Duque and Johnson 1999, 110). Although they have supplemented their argumentation with several quotes and examples to illustrate the role of metaphors, these examples concern mainly rather general theoretical reflections of scientists instead of concrete courses of action in scientific practice. For instance, they quote Broadbent, who compared the limited capacity of the brain with the limitation of sending messages in Morse code with a buzzer that cannot send a dot and a dash at the same time:

> Any hypothetical account of brain function must in the future consider on the one hand the size of the brain (how many buzzers there are) and on the other hand the rate at which that brain will make reactions to a given set of incoming stimuli (the number of dots and dashes per second). (Broadbent 1958, 5)

Although this quote illustrates Broadbent's theoretical ideas on hypothetical accounts of brain function, it does not reveal how these ideas are used in actual scientific practice to understand phenomena that are for instance observed in concrete psychological experiments. In their account, Fernandez-Duque and Johnson focus on the conceptualization of the phenomena and they point to the guiding role of metaphors. However, they have not looked in detail at how scientists conceptualize the phenomena in concrete scientific endeavors and how the guidance of metaphors functions in concrete cases. I am sympathetic to their ideas about the important role that conceptualizations play in accounts of phenomena. This topic appears to be closely related to the topic of surplus meaning that I discussed in the previous chapter on neo-behaviorism. For instance, Tolman argued that the conceptual substrate of models, which is required for their application, originates from the surplus meaning of their theoretical terms. In this view, the surplus meaning of the theoretical terms supplies the conceptual substrate that is necessary to account for these phenomena. Metaphors could very well be a source of this surplus meaning. The question if it is the only source, as the account of Fernandez-Duque and Johnson seems to suggest, is not answered by them.

In the next section, I will analyze in detail how the conceptualization of phenomena is realized in a concrete case of scientific practice, and how the conceptual substrate that is brought in by this conceptualization can be used to reason about the phenomena. My aim is to get a clearer idea of the skills required for this. I will examine the work of Broadbent, who was one of the leading early cognitive psychologists, and who made every effort to make his colleagues familiar with the application of information-theoretical models to psychological phenomena.

5.4. *The Information-Theoretical Approach at Work: Donald E. Broadbent's Account of Attention*

In this section I will focus on a concrete example of a psychological phenomenon and the way it was understood in cognitive psychology, namely the phenomenon of attention. A pioneer in this area was Donald Eric Broadbent (1926–1993), who in retrospect can be seen as one of the early cognitive psychologists although he saw himself more as a behaviorist and even disliked the strong cognitive interpretation given to his views (Baars 1986, 394). He is considered to be one of the initiators of the "paradigm shift" from behaviorism to cognitive psychology because he recognized the prospects of information theory for psychology. He was one of the first psychologists to apply concepts from communication engineering. His *Perception and Communication* (1958) is seen as a landmark in the development of information-processing psychology (Baars 1986, 394). In this book Broadbent applied information theory to the problem of attention. He explained the results of experiments on attention by hypothesizing underlying functional stages of information processing and their order of occurrence – and at a time when behaviorism was still the dominant paradigm in psychology, this was a radical move.

Broadbent's career choice was influenced by his experiences as a conscript in the British Army. Because of his fascination with flying, he was determined to complete his military service as a RAF pilot. He successfully passed the selection procedure of the RAF and later recalled that he was very impressed by the personnel selection tests. He went to the United States to learn to fly, and there he physically experienced the ergonomic problems of equipment use, which were to some extent related to the problems of applied psychology he would deal with in his professional career as psychologist. More importantly, in the United States he came in contact with psychology, which was studied much more widely there than in Britain. These experiences made him aware of the prospects of applied psychology. Back in England, in 1947, he decided to study experimental psychology at Cambridge. The head of the department was Sir Frederick C. Bartlett, who had been involved in wartime research on human performance under various demanding conditions. From Bartlett Broadbent took the idea that psychological principles should be derived from real-life situations (Weiskrantz

1994, 35). He became familiar with the novel field of cybernetic control systems explored by Kenneth J.W. Craik, who had died at a young age just before Broadbent entered the department. From the various manuscripts left by Craik Broadbent took the idea of an engineering approach to psychology, which involved applying control systems analysis to human performance in order to understand human nature (Baddeley and Weiskrantz 1993, xi). These experiences with applied psychology, which could yield practical knowledge that was useful for the army, for instance, had a major impact on Broadbent's choice of research topics in the rest of his scientific career. Although he became famous for his major theoretical contributions to cognitive psychology, his primary field of research was applied psychology.

After finishing his studies in 1949, Broadbent accepted a job with the Royal Navy that consisted of studying the effect of noise on performance. Originally, the intention was that he would perform his experiments in a naval laboratory. Broadbent, however, preferred to stay in Cambridge and asked if he could perform his experiments at the Applied Psychology Unit of the Medical Research Council (MRC) based in Bartlett's department. When the naval laboratory realized that the experiments would involve very loud and disturbing noises they approved this request (Baddeley and Weiskrantz 1993, xii). Broadbent would stay in Cambridge for 25 years. The experiments that involved the loud noises were quite boring, however, and fortunately Norman H. Mackworth, the director of the Applied Psychology Unit, allowed Broadbent to widen his scope and perform studies of vigilance tasks, which involved experiments on problems of gunnery and air traffic control systems. In these experiments, subjects only have to respond to very infrequent signals but may have to monitor their occurrence over long periods. Mackworth had performed experiments on these tasks during the war, in view of the problem of detecting submarines by radar from the air (Broadbent 1958, 108). Broadbent was especially interested in tasks in which listeners have to attend to several signals that arrive more or less simultaneously. This research topic was also interesting for the navy because it offered suggestions on how to deal with situations where a great deal of information is is being offered and some of it has to be discarded, as is the case in gunnery and air traffic control systems. Broadbent's work revealed that in these cases there should be reflection on the way in which this information is presented:

> [W]hen some information must be discarded, it is not discarded at random. Thus if some of the material is irrelevant it is better for it to come from a different place from the relevant material, or to be louder or softer, or to have different frequency characteristics, or to be on the eye instead of the ear. (Broadbent 1958, 34)

Broadbent realized that his experiments at the *Applied Psychology Unit* were all related to the phenomenon of attention, which was a topic that had been left largely untouched by traditional neo-behaviorist psychology. Inspired by the freshly developed information theory, Broadbent started to employ the vocabulary of communication engineering. Together with E. Colin Cherry, who was a professor of telecommunication at *Imperial College London*, he developed experiments in which two different auditory stimuli, usually speech, are simultaneously presented to a listener, one in each ear, normally using a set of headphones (Weiskrantz 1994, 36–37). As an adherent of information theory, Cherry used Shannon's model of communication systems as his point of departure in this psychological research, which he described as an investigation into the capacities of listeners to obtain information from noisy channels. He examined how listeners recognize what one person is saying when others are talking at the same time, which he called the "cocktail party problem" (Cherry 1953, 976). He suggested that humans have a filtering mechanism that filters out most of the irrelevant information. According to Cherry (1953, 976), if the principles of this filtering mechanism were known, one could design a machine for carrying out such a filtering operation. It was this challenge, finding the characteristics of the filtering mechanism, that Broadbent took up with his experiments on attention. His general idea was that the nervous system has a limited capacity and that the filtering mechanism is required as an economical way of managing the amount of information that passes through it (Broadbent 1958, 41). With this stress on the limits of information processing, the work of Cherry and Broadbent is similar to Miller's studies on the 'magical number seven' that I discussed earlier. An important difference between the research projects on both sides of the ocean, however, was that the British researchers not only investigated the structural limits in human information processing but also tried to determine which operations are performed by the nervous system on the information that is processed (Gardner 1985, 91). With his "non-positivistic approach,"

Broadbent (1958, 301–302) attempted to "find out what happens inside the organism." Because Broadbent's source of inspiration for the development of his account was the brand-new theory of information developed by communication engineers, it was natural to apply the methods of these engineers and represent the operational processes by means of a flow chart. He may well have been the first psychologist to apply this representational technique (Weiskrantz 1994, 37). In 1958, when he became the director of the *Applied Psychology Unit,* he presented his engineering approach and his findings in his most influential book, *Perception and Communication*. Like Miller's work in America, the studies of Cherry and Broadbent inspired the information-processing approach to psychology in England.

I will investigate concrete cases of the use of this information-theoretical approach in psychology to understand the psychological phenomenon of attention and I will describe the skills that are required for the application of information-theoretical models to this phenomenon. I will examine the surplus meaning of Broadbent's theoretical terms that (to put it in Tolman's terms) supply these models with a "conceptual substrate" (Tolman 1949, 48–49) and I will analyze how the information-theoretical conceptualization is used to reason about attention. My line of approach is to focus on the way in which Broadbent introduced his information-theoretical approach in psychology. I will examine how he contrasted the information-theoretical terms with the terms used in neo-behaviorism and how he used them to formulate an account of attention. Subsequently, I will investigate how he tried to ensure that his fellow psychologists possessed the required skills to use his approach to account for the phenomenon of attention as it was investigated in concrete psychological experiments.

5.4.1. *Broadbent's Conceptual Framework*

In addition to traditional behaviorist terminology, such as "stimulus" and "response," Broadbent used the information-theoretical terminology of communication engineering. He argued that the traditional behaviorist terminology could be used to describe observables, whereas terms from information theory could be used for theorizing about the unobservable processes in the nervous system. As I discussed in chapter 4, in the 1950s neo-behaviorists increasingly

came to appreciate the use of hypothetical constructs. Among neo-behaviorists it became fashionable to introduce physiologically inspired hypothetical constructs. Broadbent (1958, 306), however, disapproved of this, because he thought that psychologists ran the risk of having their psychological theory disproved by irrelevant physiological research. As an example, he mentioned the physiologically inspired psychological theory of Donald O. Hebb (1949):

> [T]he psychological essence of this theory is that the perception of patterns can be accounted for by the linking of unit elements in the nervous system into sequences. This theory is worded physiologically so that the elements of the theory are identified as cell assemblies and phase sequences. Should a physiological experiment cast doubt on the latter, the psychological side of the theory will be in danger of being neglected: although it may well be true even though the elements are not physically what Hebb supposed them to be.
>
> (Broadbent 1958, 306)

In his view, it is better to describe the unobservable processes in terms of information flow because the terminology of information theory is more general than that of physiology. It might very well be that such a description can be readily attached to physiological knowledge when the latter becomes available. For instance, the components of Broadbent's model of attention, such as a selective filter that blocks unattended information and a short-term store that can store information, are based on information theory. As such, they are not provided with a physiological interpretation. Nevertheless, Broadbent (1958, 303) argued that these components "could be recognized if it were possible to observe them directly: a filter or a short-term store might take different physiological forms, but it could be decided with reasonable ease whether any particular physiological structure was or was not describable by these terms." According to him, "[i]nformation theory is desirable as allowing future contact with physiology but never assuming physiological detail" (Broadbent 1958, 305). With the introduction of concepts from information theory, psychology can remain autonomous while leaving room for future connections with the adjacent sciences.

Furthermore, Broadbent was careful not to magnify the differences between his approach and dominant neo-behaviorist ideas. He argued that most of the results of his experiments could also be explained by

means of a stimulus-response theory as presented in Hull's *Principles of Behavior*. For instance, the finding that intense stimuli are more likely to produce a response is known in Hull's system as "stimulus intensity dynamism." However, a markedly anti-behaviorist aspect of using terms from information theory is that it invites theorizing about unobservable processes that take place within the nervous system. Broadbent stressed that this aspect should be seen as an advantageous addition to traditional neo-behaviorism that does not necessarily lead to conflicts with neo-behaviorist theories:

> By saying that intensity confers priority in a competition to pass through a filter at the entrance to the nervous system, we are merely adding to the Hullian statement about behaviour. Reinforcement theorists may regard this addition as speculative, since they sometimes accept uncritically the view that statements about internal processes are operationally untestable. But at least the addition is not in conflict with reinforcement. (Broadbent 1958, 247)

Broadbent (1958, 266) suggested keeping the words "stimulus" and "response" for observables and speaking of "information" when drawing inferences about unobservables. This could avoid a danger common to speculative physiology, namely "that if we use words normally given an objective meaning we will think our theory objective even though it [may] be wildly speculative" (Broadbent 1958, 304). As an example, Broadbent mentioned the frequent use of unobservable responses in stimulus-response theories.

Another advantage of applying terms from information theory is that it facilitates analysis of limitations in the human processing of information. As I already discussed above, one of the central concepts in information theory is the limited capacity of information channels. Broadbent, whose main research domain was applied psychology, stressed the importance of this concept in psychology. For instance, he argued that speech messages bring the brain of the listener into one of its possible brain-states. Because humans are capable of making different responses to many different sentences, Broadbent (1958, 38–39) considered this number of possible states to be quite large:

> The process we have been considering, in which one of a set of possible signals enters a system and one of another set emerges at the far end, is analogous to that of telephone or radio communication. The 'number of possible states' or the 'vocabulary' which we have considered so far

> is usually called the 'ensemble' of signals by communication engineers. They speak of the 'information' conveyed by a signal as increasing with the size of the ensemble from which it is drawn. If there are *n* possibilities, all equally probable, the information is usually defined as the logarithm of *n*. (Broadbent 1958, 39)

When the listener has to deal with a great deal of information in a short period, for instance, when more than one sentence is being presented at the same time, the amount of information that enters the nervous system in that period may surpass the amount of information it can handle given the number of possible brain-states. In other words, in some cases "the human being acts as a single communication channel and therefore cannot deal with two signals in rapid succession" (Broadbent 1958, 281). According to Broadbent (1958, 5), the limited capacity of communication systems is "a matter of central importance to communication engineers, and it is correspondingly forced on the attention of psychologists who use their terms."

According to Broadbent, one of the problems with neo-behaviorism is that it is unable to take into account that it generally depends on the context whether or not two stimuli should be regarded as similar. In his view, information theory does take the context into account. In this theory, the information content of a signal is equal to the choice that it enables the receiver to make within a set of *possible* messages. Therefore, whether two signals differ in meaning depends on their context (the set of possible messages). It is a major advantage of information-theoretical accounts that they enable one to "consider the whole ensemble of possible stimuli rather than simply the presence or absence of each one" (Broadbent 1958, 59). Broadbent criticized neo-behaviorism for its positivist attitude towards the metaphysical notion of possibility. He argued that because of the refusal to use this notion in the definitions of terms, stimuli-response terminology is vague:

> [T]o regard two stimuli as the same when one is chosen from two possibilities and another from a set of fifty, is … unjustifiable; even when the two are physically identical in every way. The word 'stimulus' is as vague as the word 'thing' in the phrase 'doing two things at once', and for the same reason: it contains enough truth for its unsatisfactory character to be overlooked. (Broadbent 1958, 35)

To some extent, Broadbent's own terminology is also vague. For instance, it is difficult to get clarity on his notion of channels. The most

important feature of channels in Broadbent's description is that they transmit signals that have some characteristics in common (Broadbent 1958, 291). However, in Broadbent's account it is not specified in detail what these common features are. An example of a possible common feature is that the signals originate from the same sense organ. Broadbent referred to such channels as sensory channels. Roughly speaking, the sense organs can be seen as the entrances of these channels. However, as is shown, for instance, by experiments on the role of auditory localization, this view of sensory channels is an oversimplification. Although the only sense organs affected in those experiments are the two ears, if the sound that reaches the listener arrives from different directions, the channels involved correspond to the different directions rather than to the two ears (Broadbent 1958, 14–15). Thus, in general, sensory channels transmit information from sensory perceptions that have certain features in common.

In sum, Broadbent introduced concepts in psychology that differed from traditional neo-behaviorist concepts, because their surplus meaning, or the conceptual content, originated from information theory. These concepts facilitated his theorizing. I will now investigate how he used these terms to account for the phenomenon of attention.

5.4.2. *The Filter Theory of Attention*

Broadbent conceptualized attention as a filter that filters out unattended information and passes through attended information. The main parts of his model of attention are a short-term store and a filter. Information that comes in through the senses is placed in a short-term store. After that, it is selectively filtered. Only information that passes through the filter enters a system of limited capacity. In Broadbent's account, this information becomes conscious and can enter long-term memory. The sense organs can take in more information than the nervous system can process. Therefore, the selective filter has to block some of the sensory channels. The blocked information is stored in a short-term buffer in which it decays in a few seconds. From this short-term store the information can be introduced again to the filter, which can let it through or send it back to the store. "The theory thus becomes one of recurrent circuit type, in which information is passed continuously around a loop until required. As has been said, this is in

fact a common device for short term storage in machines" (Broadbent 1958, 226–227).

In Broadbent's view, it is only logical that the capacity of the brain will limit the number of tasks that can be performed simultaneously and that part of the information presented must be discarded. Because not all the information that enters the nervous system can be processed, the hypothesis of a filter mechanism that operates on it is "almost purely a logical statement" (Broadbent 1958, 106). However, the principles that govern the discarding of information cannot be stated simply by logical analysis. Instead of using Hull's hypothetical-deductive methodology and starting with postulates from which these principles be deduced, Broadbent (1958, 7) stressed that what human behavior is like should be discovered first. For this reason, the structure of his *Perception and Communication* (1958), which describes the investigation of these principles, differs from most psychology books of that period:

> We are reversing, in the course of this book, the plan of many books on psychology. They often start with a general discussion of scientific method, then set up postulates, and then discuss predictions from those postulates. The type of such books is that of Hull (1943). We, on the other hand, have discussed results first, then theories, and finally broad principles. (Broadbent 1958, 307)

An inspiring starting point for the experimental inquiry into these principles was Cherry's (1953) "shadowing" experiment on selective listening in which subjects heard different messages through right and left ears of headphones. Cherry asked them to quickly repeat aloud ("shadow") the message presented to one specific ear. People become very good at this task with practice and are able to repeat the input with a delay of a few tenths of a second. Cherry discovered that, while performing this task, subjects were unable to report much of what was presented to the unattended ear. Although they could report some characteristics of the signal, such as whether it was music or speech, they generally were not able to report shifts of content or tongue:

> During the experiment [Cherry] changed the nature of the stimulus on the ear which was to be ignored. Although that ear was first stimulated with ordinary speech, it was then given reversed speech having the same spectrum as normal speech but no words or meaning. Other subjects heard a woman's voice follow that of the man who had started

> the passage, others heard German words instead of English ones, and still others heard a pure tone. These changes were all reversed before the end of the session, so that the subject was only exposed to the changed stimulus during the middle of the period while he was steadily repeating the speech on the other ear. After the experiment all subjects were interrogated about the speech they had to ignore. They could say nothing about its content or even what language it was in: a few noticed 'something queer' about reversed speech, but others did not. But the change from one voice to another was nearly always noticed, and the change to a pure tone was always noticed.
>
> (Broadbent 1958, 22–23)

Apparently when the selective filter is tuned to one sensory channel, it filters out most – but not all – of the information transmitted by other sensory channels. The subject receives some information even from the rejected ear, which means that a small amount of information passes through blocked channels. Broadbent suggested that this small amount of information is used by the filter mechanism to determine if these channels are to be kept blocked or not:

> Little mechanism is required to decide which sensory channel a particular word has arrived by, to reject words from one channel, and to pass those from another channel on to a further mechanism for the analysis of the remaining information they convey. Such a system is therefore an economical way of keeping down the amount of information passed through the main part of the mechanism. … We may call this general point of view the Filter Theory, since it supposes a filter at the entrance to the nervous system which will pass some classes of stimuli but not others. (Broadbent 1958, 41–42)

Broadbent used experimental investigations to determine the role that this small amount of information that passes blocked channels plays in the selection procedure and to reveal the factors that determine if information on a sensory channel passes through the selective filter. As I will show when I discuss some of the experiments, these factors included the physical intensity of the signals that conveyed the information, the absence of recent inputs on the channel, and the position of the sensory channel in the hierarchy of the different sense channels (Broadbent 1958, 265).

In addition to these findings, experimental investigation showed that a shift of the selective process from one sensory channel to another takes a determinate time (Broadbent 1958, 299). This small lapse of

time is noticeable, for instance, when persons carry out tasks of prolonged work. If a person is presented with a series of stimuli, a fresh stimulus immediately following every response, after some time there will occasionally be a very slow response. These slow responses are called "blocks" in psychology (Broadbent 1958, 128). Broadbent's interpretation of these blocks is that they represent interruptions in the intake of information from one source, owing to intake of information from another source. The filter, which is tuned to a specific sensory channel, occasionally passes stimuli from another channel for a second or so, and during this time the information on the original channel is blocked. The frequency of shifts by the filter depends on the relative intensity and novelty of the stimuli on the channels involved. An increase in either one of these qualities makes it more likely that a stimulus passes the filter (Broadbent 1958, 134). A similar effect is visible in the vigilance tasks mentioned earlier. After prolonged observation of one source of information, an observer will show brief intervals in which information is taken in from another source. Broadbent argued that because the filter possesses a bias in favor of channels that have not been active recently, it is likely that attention wanders after one source of information has been controlling responses for some time. In general these breaks in the intake of task information are brief, but increase in frequency as the task is continued (Broadbent 1958, 120).

In the last chapter of *Perception and Communication* (1958), Broadbent gave twelve principles that together constitute his abstract theory of attention and immediate memory. He supplemented these principles with an information flow diagram that reflects the processes involved in the phenomenon of attention (see figure 5.4.2). The experimental findings about attention could be explained by means of the principles and the flow chart (Broadbent 1958, 297–299).

The principles and the flow chart were presented as a comprehensive account of the phenomenon of attention. Broadbent wrote that his intention with his book was to show that nervous systems are networks of the type shown in figure 5.4.2, "and of no other type" (Broadbent 1958, 304). In what follows I will concentrate on the principles concerning the basics of the filtering of information by the nervous system (namely principles *A*, *B*, *C*, *D*, and *J*). I will not deal with the principles concerning the short-term store because discussing them

would not yield any new insights for my philosophical analysis of Broadbent's application of information theory:

(A) A nervous system acts to some extent as a single communication channel, so that it is meaningful to regard it as having a limited capacity.

(B) A selective operation is performed upon the input to this channel, the operation taking the form of selecting information from all sensory events having some feature in common. Physical features identified as able to act as a basis for this selection include the intensity, pitch, and spatial localization of sounds.

(C) The selection is not completely random, and the probability of a particular class of events being selected is increased by certain properties of the events and by certain states of the organism.

(D) Properties of events which increase the probability of the information, conveyed by them, passing the limited capacity channel include the following: physical intensity, time since the last information from that class of events entered the limited capacity channel, high frequency of sounds as opposed to low (in man), sounds as opposed to visual stimuli or touch as opposed to heat (in dogs).

. . .

(J) A shift of the selective processes from one class of events to another takes a time which is not negligible compared with the minimum time spent on any one class. (Broadbent 1958, 297–299)

Broadbent realized that this account of attention could fall on fertile ground only if the information-theoretical description was intelligible to his readers. Psychologists were not familiar with the information-theoretical approach and the use of flow charts. Therefore, the theory

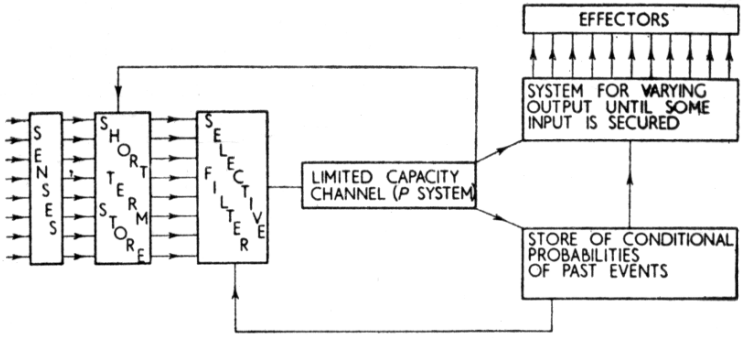

Figure 5.4.2. Information flow diagram of attention (Broadbent 1958, 299)

of information, and the use of information flow diagrams had to be carefully explained (Boden 2006, 291). In *Perception and Communication* (1958) Broadbent gave a thorough analysis of how to explain psychological phenomena by analyzing the flow of information through the nervous system and depicting this information flow using flow charts such as figure 5.4.2. I will briefly examine how he did this because it indicates which skills – in Broadbent's view – are relevant for the use of the flow chart as a model by means of which the theoretical principles can be applied to the phenomenon of attention. After that I will analyze another aid that Broadbent used to make his account of attention intelligible, namely a mechanical model that he presented to clarify the working of the filter mechanism.

5.4.3. *The Introduction of Flow Charts*

Broadbent illustrated what he meant by the flow of information by drawing an analogy with a railway system (see figure 5.4.3). Trains departing from London represent information entering the nervous

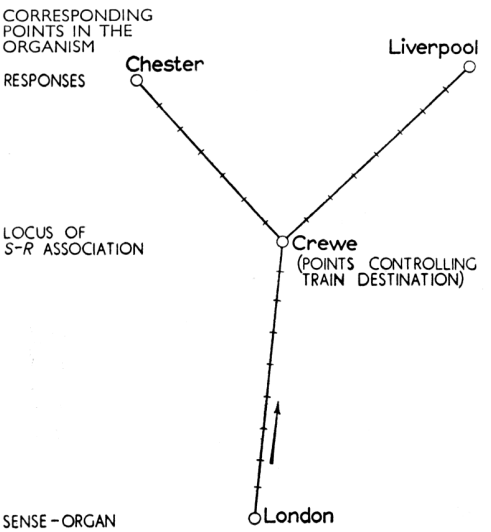

Figure 5.4.3. A railway system analogous to the flow of information through the nervous system in conditioning (Broadbent 1958, 188)

system. Like trains, which travel from one station to another, information flows from one stage to another. And similar to trains, which arrive at another destination when the points are switched, information can take different courses in the nervous system and result in different output. For example, if an unconditioned stimulus like a noise that initially does not lead to a conditioned response is always followed by a reward such as food, it transforms into a conditioned stimulus that does result in a conditioned response such as salvation:

> When an unconditioned stimulus is delivered, the information about this event travels through the organism and eventually appears as an output, such as turning the head to examine the source of stimulation. After conditioning the first stages of this journey must be the same, but at some point the information takes a different course and emerges in a different output, such as salivating. An analogy would be a railway system in which trains from London normally arrive at Chester, until the points are switched over at Crewe. After that operation the trains from London arrive at Liverpool. (Broadbent 1958, 187–188)

Although this analogy looks rather simple, Broadbent deemed it necessary to introduce it because it highlights significant aspects of information flow analyses that are virtually absent from neo-behaviorist analyses of psychological phenomena. Like the description of events happening at different stages of the train journey, an informational description aims at describing the operations that information undergoes from the moment it enters a system until it eventually produces an output. These operations are "informational events" (Palmer and Kimchi 1986, 39) which – like the event of the journey – are (spatio)temporally ordered. This ordering, which is generally neglected in neo-behaviorist analysis, is an important aspect of an information flow analysis. For example, conditioned responses can be inhibited. In a neo-behaviorist analysis of this phenomenon there is no room for mentioning the locus of inhibition, whereas in an information flow analysis there is: information that enters the nervous system may be discarded before or after the stimulus-response association. In terms of the analogy, if trains have disappeared, this might have happened before or after they reached Crewe. Broadbent regarded ignorance about the locus of inhibition an inadequacy of the neo-behaviorist analysis that, he argued, had led to several misinterpretations. Therefore, he considered it desirable to give an information flow analysis

instead (Broadbent 1958, 190; 1956a, 359). The ordering of the informational events can be represented with a flow chart that specifies how the information flows through the system. In these diagrams, such as figure 5.4.2, the informational events are depicted as a box and incoming and outgoing arrows depict their interrelationships. In Broadbent's view, the analogy of the railway system with its connections between different stations is a helpful aid to becoming familiar with the technique of anatomizing the psychological phenomena into these informational events with their ordering relations. In other words, it helps to master the skill of conceptualizing phenomena in information-theoretical terms. It is this skillful technique that scientists have to master before they can work successfully with the information flow chart and understand Broadbent's account of attention.

5.4.4. *The Mechanical Model for Attention*

Understanding the phenomenon of attention by means of Broadbent's flow chart and his twelve principles requires familiarity with informational descriptions. The psychological phenomenon is decomposed into informational events, such as the event of information filtering, that are described by theoretical principles. The ordering relations between these informational events are depicted by means of a flow chart. As I argued above, Broadbent tried to make his audience familiar with this kind of descriptions by means of the analogy with the railway system, for instance. However, mastering the skill of anatomizing phenomena in informational events with ordering relations between them is not sufficient for an information-theoretical understanding of them. In addition, each informational event also has to be understood. Broadbent's twelve principles of attention are meant to account for the informational events in his flow chart of attention. However, like Hull's principles of behavior discussed in chapter 4, which contain theoretical terms that are only partially defined, Broadbent's principles also contain concepts of which the meaning is not completely specified, such as the communication channel or the selective filter. And like Hull, who often supplemented the formal definitions of hypothetical constructs with informal formulations in which the constructs were depicted as mechanisms mediating between stimuli and responses, Broadbent tried to make his account intelligible by introducing

mechanical models that he used to articulate the meaning of his theoretical terms. I will analyze in detail how Broadbent used such a mechanical model to elucidate the meaning of his theoretical terms.

A year before presenting his theory of attention in *Perception and Communication* (1958), Broadbent published a very similar version of it in an article entitled "A Mechanical Model for Human Attention and Immediate Memory" (1957). He realized that his audience might be unacquainted with his information-theoretical account and therefore consider his theory "an abstract [one] using unfamiliar terms" like "communication channel" and "channel capacity" (Broadbent 1957, 206). Accordingly, he was afraid that the theory was "difficult to communicate to others without putting them to the trouble of learning the necessary vocabulary" of information theory (Broadbent 1957, 205). Moreover, he was afraid that it was open to misinterpretation (Broadbent 1957, 205). To avoid this, Broadbent presented an expository device for his abstract theory in the form of a mechanical model. It was meant as a device for communicating the outline of the abstract theory – "directed largely at those who find such a theory unintelligible in its original form" (Broadbent 1957, 213) – and for demonstrating how this theory connects a number of facts about attention "in a way which most people can understand" (Broadbent 1957, 209). Like the train analogy, this mechanical model was very simple. This might suggest that these expository devices were developed for an audience of lay people. However, this is clearly not the case, because the article about the mechanical model appeared in *Psychological Review*, and the

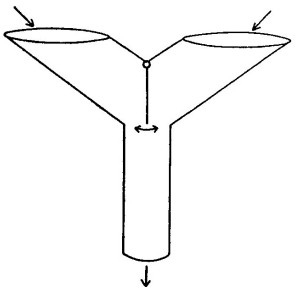

Figure 5.4.4.a. A simple mechanical model for human attention (Broadbent 1957, 206)

train analogy was presented in *Perception and Communication* (1958), which was Broadbent's most important scientific work.

Broadbent presented two versions of the mechanical model, namely a simple version comprising only the selective filter and an extended version that also comprises the short-term store. I will only look at the simple mechanical model, which consists of a Y-shaped tube mounted vertically (see figure 5.4.4.a), and a set of small balls. The stem of the Y tube is so narrow that it can take only one ball at a time. At the junction of stem and branches is a hinged flap that can be pivoted about its upper edge so as to close off either of the branches (Broadbent 1957, 206). The flap moves freely so that a ball dropped into one arm of the Y will knock the flap aside and fall into the stem. When the flap allows balls from one branch, it automatically closes the other (see figure 5.4.4.b).

According to Broadbent, his abstract theory of attention, formulated in the form of theoretical principles, is applicable to this mechanical model. This may sound peculiar because the principles were meant to describe human attention, and not the behavior of a mechanical model. However, Broadbent argued that the theoretical principles are formulated using terms from information theory, and "information concepts are applicable to any system, whatever its physical nature, and so may equally fit a model or a man" (Broadbent 1957, 206). In terms of the semantic view of theories discussed in chapter 3, the mechanical model is one of the models of Broadbent's theory: it is an interpretation that satisfies all the theoretical principles. Of course,

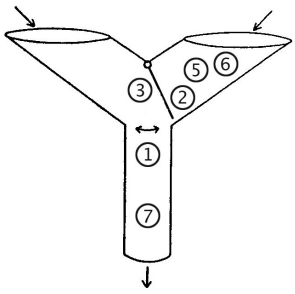

Figure 5.4.4.b. The mechanical model in action

this requires that there be a mapping between the theoretical terms of the principles and the components of the model such that the model satisfies the theoretical principles. Broadbent (1957, 206–207) provided such a mapping: the balls represent the information from various stimuli, the branches represent different sensory channels, the narrow stem represents the communication channel with limited capacity, the flap represents the selective operator, and the dropping of a ball into one of the arms represents the delivering of input to the channel.

It is clear that with this mapping the Y tube satisfies principles *A* and *B* (see section 5.4.2). Because the narrow stem (representing a communication channel) can take only one ball at a time (representing a certain amount of information) it is indeed meaningful to regard it as having a limited capacity. Furthermore, the hinged flap (representing a selective operator) can block or pass balls through that are dropped into the branch (representing information delivered into a sensory channel) and thus it indeed performs a selective operation. For the other three theoretical principles it is less obvious that the model satisfies them. To illustrate that they do, Broadbent discussed some thought experiments with the mechanical model. For instance, he argued that the flap may not block the balls completely at random. The probability that balls from one branch could pass through is increased by certain properties of the balls, such as the force by which they are thrown down, and by certain conditions of the tube, such as the steepness of the branches. He thus tried to demonstrate that the mechanical model fully satisfies the abstract theory.

As I will argue below, Broadbent compared the implications of his thought experiments with results of actual psychological experiments about attention. The idea was that the resemblance between the behavior of the mechanical model, which satisfies the theoretical principles of attention, and the behavior of the subjects in the psychological experiments illuminates the relation between the abstract theory and the experimental results. In other words, the mechanical model can be used to "mediate" (cf. Morrison and Morgan 1999b, 17–18) between this theory and the phenomenon of attention. Determining the resemblances between model and phenomenon requires the selection of relevant features of the model and the phenomenon and the evaluation of the possible similarities between these features. As discussed

in chapter 3, this is a skillful practice. Furthermore, the comparison between the behavior of the model and the behavior of the subjects involves the skillful practice of reasoning via the model. I will analyze the required skills by looking closely at the comparisons that Broadbent made between the behavior of the mechanical model and the results of the experiments on attention.

5.4.5. *Analogies between the Mechanical Model and the Phenomenon of Attention*

According to Broadbent (1957, 207), the behavior of the model resembles the behavior of humans in several respects. In his article, he described a number of relevant features of the behavior of the mechanical model, and he mentioned a number of psychological experiments on attention whose results, he argued, bore a resemblance to these features of the behavior of the mechanical model. He did not describe the experiments in detail. Instead, he only summarized the experimental results and gave references to publications about these experiments. I will examine Broadbent's description of the relevant behavior of the model, and review the experiments he associated with them.

The *first* situation discussed by Broadbent is that of two balls dropped simultaneously, one into each of the branches. Because the balls strike the flap on both sides, it does not move and the balls become jammed in the junction. However, if the flap is pivoted and closes off one of the branches before the balls are inserted, then the ball entering the other branch will emerge successfully (Broadbent 1957, 207).

The analogous situation for humans takes place in experiments carried out by Broadbent (1952) and E.C. Poulton (1953). In these experiments subjects heard two messages in different voices, and in Poulton's experiment also coming from different spatial directions. Subjects had to respond to only one of the two messages, namely the one that started with a specific call sign. For instance, in Poulton's experiment, where air traffic control tower communication was simulated, subjects had to respond only to aircraft calling one particular tower, such as "Lakenheath tower," by writing down the aircraft number. The subjects performed their task very well, except for the occasional situation in which the two voices spoke simultaneously. Then only about

50% of the subjects responded correctly to the message. By way of explanation, Broadbent supposed that "one of the two voices is selected for response without reference to its correctness, and that the other is ignored. It is understandable, then, that simultaneous auditory call signs should be ineffective: if one of the two voices is selected (attended to) in the resulting mixture there is no guarantee that it will be the correct one" (Broadbent 1952, 53). However, the performance of the subjects increased significantly when they were told in advance which of the voices they had to answer. In the case of simultaneous speech they were able to attain an efficiency of about 70% (Broadbent 1952, 54).

The *second* situation discussed by Broadbent is one in which the two balls are not strictly simultaneous. In that case, the first to arrive will gain an advantage by knocking the flap aside and shutting out the other (Broadbent 1957, 207).

The analogous situation for humans takes place in an experiment by J.C. Webster and P.O. Thompson (Broadbent actually referred to an experiment by W. Spieth, J. Curtis, and J.C. Webster that was published in 1954 but probably meant an experiment by Webster and Thompson that was published the same year; see also Broadbent 1958, 15). Like Poulton, Webster and Thompson simulated conversation in an air traffic control tower. Subjects heard airplane tower phraseology over six loudspeakers placed at different locations and had to respond to all messages. For each message they had to activate a switch corresponding to the loudspeaker from which it came and repeat back what they heard. Occasionally, messages coming from two loudspeakers overlapped in time. The degree of overlap was systematically varied (Webster and Thompson 1954, 396–398). This experiment demonstrated that in the case of two messages overlapping in time, the subjects responded much more accurately to the message that arrived first than they did the other.

The *third* situation discussed by Broadbent is one in which the Y-shaped tube is not mounted in a perfectly vertical position. In that case, the ball in the more vertical branch will have an advantage over a simultaneous ball in the other because the flap will hang to one side (Broadbent 1957, 207).

The analogous situation for humans takes place in experiments done by Broadbent (1954a). In these experiments subjects had to

perform a five-choice task while being exposed to high-pitched or low-pitched loud noise. The experimental setup consisted of five separate lights and five buttons each corresponding to a light. When a lamp lit up the subject had to touch the appropriate button as quickly as possible, which caused another lamp to light. Accordingly, there were no rest intervals between the signals. During this task, touches given to incorrect buttons as well as brief pauses by the research subjects of periods of two seconds or more were recorded (Broadbent 1958, 95). Compared to the performance of subjects that received low-pitched noise, subjects that received high-pitched noise made significantly more errors, which demonstrated that high-pitched noise is more likely to distract attention than low-pitched noise (Broadbent 1958, 99).

The *fourth* situation discussed by Broadbent is one in which one ball is thrown violently down its branch. In that case it may succeed in forcing the flap against the unassisted weight of a ball on the opposite branch (Broadbent 1957, 207).

The analogous situation for humans takes place in experiments by D.E. Berlyne (1950) and Broadbent (1954a). In Berlyne's choice-response experiment, subjects were asked to respond to any one of two simultaneously presented visual stimuli of different brightness by pressing a corresponding key. It appeared that they were more likely to choose the brighter one. In Broadbent's experiment, the five-choice task discussed above, subjects were not only exposed to high-pitched or low-pitched loud noise but also to noise of lower physical intensity. Compared to the performance of subjects under that circumstance, subjects that received loud noises made significantly more errors. As Broadbent formulated it concisely:

> Twenty-four subjects were used on this occasion, divided into three equal groups. Each group received its own intensity of noise: 80, 90 and 100 db. Each man worked for ½ hr periods, in successive days, one day receiving a noise containing only frequencies above 2000 c/s, and the other day receiving a noise containing only frequencies below that point. The high pitch noise produced more errors throughout, but the difference was insignificant at the two lower intensities. It was highly significant at 100 db: indeed, high-pitched 100 db noise produced twice as many errors as high or low pitched 80 or 90 db noise, which were all very similar to one another. (Broadbent 1958, 99)

The *fifth* situation discussed by Broadbent is that of the overswinging flap. "After a single ball has been passed through the system, the door will swing back from the position into which it was pushed. Naturally it will overswing, and temporally close the branch which has just been used" (Broadbent 1957, 207).

The analogous situation for humans takes place in an experiment by Poulton (1956) that demonstrated that a stimulus has an extra advantage for a response if it comes on a previously quiet channel as opposed to a previously busy one. This experiment was again one in which air traffic control tower communication was simulated. Subjects had to respond only to aircraft calling one particular tower. All the loudspeakers were quiet during the intervals between the calls except one, which was busy all the time. On the busy loudspeaker the interval between calls was occupied by other conversations. It was found that a call from a usually quiet loudspeaker tended to capture the attention of the subject even when this call was irrelevant and coincided with a relevant one from the busy loudspeaker (Poulton 1956, 338).

The *sixth* situation discussed by Broadbent is one in which balls are being inserted into one branch at a certain rate of delivery. In that case, "the effect of increasing the rate of delivery of the balls through the same branch is not the same as that of adding the same number of extra balls to the other branch. There is more risk of jamming in the latter case" (Broadbent 1957, 207).

The analogous situation for humans takes place in an experiment by R. Conrad (1951). In this experiment, the participants were engaged in a clock-watching task called the "speed and load test." The experimental setup consisted of two, three, or four clock dials and the same number of buttons corresponding to these dials. Subjects had to attend to the dials (see figure 5.4.5). When the rotating pointer of any of these dials approached the 12:00 or 6:00 position, they had to press the appropriate button. In different experimental sessions, the number of dials used and the rotation speed of the pointers of the dials were varied. It was recorded whether or not the subjects pressed the button at the right moments. Missed responses and responses at wrong moments were recorded as erroneous. The subjects were tested at five speeds. They had to respond 40, 60, 80, 100, or 120 times per minute. Not only did higher speeds appear to result in more errors but this effect also appeared to be drastically amplified when more

dials were used. Doubling the number of dials resulted in considerably more errors than doubling the speed of the dials, even though the number of additional signals to which the subjects had to respond was the same.

The *seventh* situation discussed by Broadbent is one in which two balls are dropped into the model and jamming is avoided by the flap being coincidentally to one side. One of the balls can pass through, while the other is blocked. "[T]he impeded ball will not therefore disappear. It will emerge later, when the door next swings back to the opposite branch. Surprisingly enough, this also happens with man. Simultaneous stimuli either jam or produce successive responses" (Broadbent 1957, 208).

The analogous situation for humans takes place in experiments by Broadbent (1954b; 1956b). In these experiments the subjects were given a pair of headphones, and two series of digits were presented simultaneously, one to each ear. The subjects were asked to repeat the digits. It appeared that both series could elicit responses but would do so one after the other. In the experiment a set of three digits, such as "7 1 3" was applied to one ear, and a different set such as "2 5 6" to the other ear. The subject's natural response was either "7 1 3 2 5 6" or "2 5 6 7 1 3" but never "7 2 1 5 3 6". Deliberate instructions to repeat the digits in order they were presented produced a very low proportion of correct responses (Broadbent 1956b, 145). However, when there was a time interval of about one and a half or two seconds between succeeding pairs of simultaneously presented

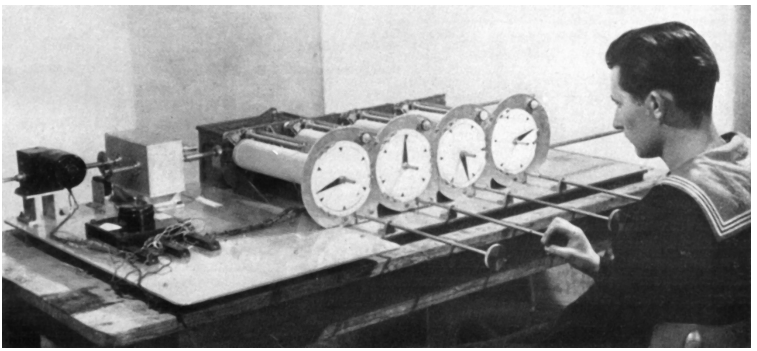

Figure 5.4.5. Speed and Load Stress (Conrad 1951, 3)

digits, the subjects showed considerable improvement. According to Broadbent (1954b, 195; 1958, 212) this implies that a time interval of between one and two seconds is required for attention to shift away from one sensory channel to another and back to the first.

By mentioning the relevant properties of the model and discussing the similarities that these properties have with characteristic properties of the phenomenon of attention, Broadbent showed how his mechanical model mediates between the theoretical principles and the phenomenon. According to Broadbent it is not necessary that model and phenomenon be similar in all respects. Dissimilarities do not have to be considered disadvantages if it is clear that they are irrelevant:

> [The mechanical model] is admittedly ludicrous as a description of what really happens in the brain, but this is a positive advantage. Psychologists are not likely to mistake this model for speculative neurology, and so they should concentrate their experiments on the essentials of the theory rather than the irrelevant properties of the model.
>
> (Broadbent 1957, 209–210)

In all seven situations of the mechanical model it is easy to have insight into its behavior. The model has several virtues, such as its visualizability and causal-mechanical nature. It is so simple that it is straightforward to mention some of the characteristic features of its behavior in specific circumstances. Everybody who possesses the skills of visualization and causal reasoning, and who has experience with causal agency, is able to reason with this model. For each of the seven situations, Broadbent mentioned a situation concerning the phenomenon of attention that can easily be regarded as analogous to it. Therefore, he argued that "the Y tube does seem at this time to have related a number of facts about perception and put them in a way which most people can understand" (Broadbent 1957, 209).

In sum, Broadbent's ability to apply the mechanical model successfully to the phenomenon of attention was a skillful activity, based on skills such as conceptualizing phenomena in information-theoretical terms, visualization, and causal reasoning. These skills were required for selecting the relevant features of model and phenomenon, evaluating the possible similarities between these features, and reasoning via the model.

5.5. *Relevant Skills for Successfully Applying Models to Phenomena*

A model is intelligible to its users if they possess certain skills and the model certain virtues, such that the combination of skills and virtues facilitates successful application of the model. The purpose of the case study of cognitive psychology is to analyze the skills that are involved in the application of a model to a phenomenon. In several examples in the case study it might seem that applying models is so straightforward that it is exaggerated to call it a skillful activity. For instance, at first sight, it seems trivial to connect Shannon's model of a communication system to Miller's phenomenon of the 'magical number seven.' However, differences in the selection of relevant features, and differences in the evaluation of similarities, as expressed, for instance, in Neisser's disapproval of the early cognitive psychologists' use of Shannon's model, give evidence for the contrary. Several aspects of the application of a model to a phenomenon are (at least philosophically) non-trivial.

In applying a model to a phenomenon, the conceptual substrate of the model is mapped onto the phenomenon. This conceptual substrate facilitates reasoning about the phenomenon. In theoretical models this conceptual substrate is contained in the surplus meaning of the theoretical terms. The source of this surplus meaning can be a metaphor such as the 'mind-as-information-processing-device' one. Whether or not scientists consider a model applicable to a phenomenon depends on contextual factors such as the availability of certain metaphors. It is common nowadays to employ information-theoretical concepts to describe human behavior. Therefore, the application of Shannon's information-theoretical model appears to be elementary. However, the use of these concepts in the domain of psychology presupposes the availability of the 'mind-as-information-processing-device' metaphor, which became prevalent only after major advances in the development of information-processing technology. Furthermore, there are other contextual factors that influence the assessment of scientists concerning the applicability of a model to a phenomenon. For instance, in the case of Neisser's disapproval of the application of Shannon's model in psychology, one of these factors was Neisser's interest in theoretical psychology rather than applied psychology.

Both the activity of conceptualizing the phenomenon and the activity of using this conceptualization to reason about the phenomenon are philosophically non-trivial activities. They require skills to assess the relevant similarities between the source domain and the target domain and skills to use the concepts of the source domain to reason about the target domain. For example, differences in the way in which scientists connected model and phenomenon illustrate that the activity of conceptualizing the phenomenon of the 'magical number seven' by means of Shannon's model is not merely a matter of course. Whereas Pollack specified the receiver as the listener, Miller specified the communication channel as such. That conceptualizing the phenomenon involves learned abilities is illustrated by means of Broadbent's analogy of the railway system that helps to master the technique of anatomizing the psychological phenomena into these informational events with ordering relations between them. Broadbent's mechanical model of attention illustrates that reasoning about the phenomenon by means of the conceptualization also involves skills. Broadbent conceptualized attention as a filter mechanism. In situations in which a lot of information enters the nervous system, such as a cocktail party in which one listens to one person while others are speaking at the same time, this mechanism filters out most of the irrelevant information. Broadbent explained his ideas concerning the principles that governed the discarding of information by means of a mechanical model. In this model, a flap that could close off branches of a Y-shaped tube represented the information filter and balls that enter the branches of the tube represent information. Reasoning via this model requires experience with causal agency and involves skills such as visualization and causal reasoning. These skills enable recognizing characteristic consequences of the model and making assessments about relevant similarities between the behavior of his mechanical model and the behavior of subjects in psychological experiments on attention. In sum, from a philosophical point of view, the application of a model to a phenomenon is a non-trivial activity that requires skills.

CHAPTER 6

Conclusion

In chapter 3 I developed a philosophical framework for scientific understanding of phenomena. A central idea in this framework is that scientists scientifically understand a phenomenon if they are able to apply a scientific model successfully to that phenomenon. A key notion in this framework is the intelligibility of a model, which is described as a value attributed to the model by its users. This value reflects their ability to apply the model successfully to a phenomenon. A prerequisite for the successful application of a model to a phenomenon, and thus for scientific understanding, is that the model is intelligible to its users. Other key notions in this framework are the skills of the scientists and the virtues of the models. The intelligibility of a model depends on the virtues of the model and on the skills of the scientists. At the end of chapter 3 I formulated five key questions concerning the key notions of this study. In this chapter I will use my analysis of the case studies in chapters 4 and 5 to answer these questions.

6.1. *Is the intelligibility of models an epistemic value and how does it function in scientific practice?*

The epistemic significance of the intelligibility of scientific models is one of the central issues in the case study of neo-behaviorism. Because of their positivist attitude towards science, neo-behaviorists rejected understanding as a genuine aim of science, viewing prediction and control to be its only aims. If they valued intelligibility of models at all, they regarded it as subordinate to positivist norms such as objectivity and verifiability. An analysis of the actual scientific practice of neo-behaviorism, however, shows that in spite of their positivist attitude even neo-behaviorists aimed at intelligible models. Their models were intelligible to them due to the surplus meaning of the theoretical terms in these models. The meaning of the theoretical terms generally exceeded their operational definitions because these terms were

named, for example, "demand" and "reinforcement," and because of the informal formulations that were supplemented to their formal definitions. Neo-behaviorists were able to apply their models to behavioral phenomena as a result of this surplus meaning. The application of theoretical terms in a specific domain required the establishment of operational definitions. This involved judgments to determine if these operational definitions can be considered adequate means to apply the theoretical concepts they represent. These judgments were facilitated by the surplus meaning of the theoretical terms. For instance, due to the surplus meaning of 'reinforcement,' which originated in everyday experiences, Hull recognized that in one experimental situation repeatedly receiving a shock after hearing a noise counts as 'reinforcement,' whereas in another experimental situation obtaining food after depressing a bar counts as 'reinforcement.' Consequently, he was able to apply his models of behavior that contained the theoretical term 'reinforcement' in different domains. Removing the surplus meaning of the terms (if at all possible) would make the models unintelligible, and that would drastically reduce their applicability. Consequently, even in the case of neo-behaviorism, the intelligibility of models has epistemic significance.

Like empirical accuracy, consistency, scope, simplicity, and fruitfulness, the intelligibility of models is one of the epistemic values that are constitutive of science. As I have argued, the various epistemic values are not completely independent. For instance, scientists do not consider a model to be intelligible if it is not empirically accurate. Intelligibility is also related to fruitfulness. According to Tolman, a model is fruitful if it supplies a conceptual structure that facilitates its application to phenomena. For instance, Tolman's model of the behavior of rats in mazes supplies a conceptual structure consisting of goals ('demands') and expectations ('hypotheses'). This conceptual structure enabled Tolman to apply his theoretical model to the concrete situation of rats in mazes. The feature of supplying a conceptual structure is a necessary condition for the intelligibility of a model. However, it is not a sufficient condition. In addition, the application of the model should be *successful*, which is the situation in which Tolman calls the model a "happy" one and which requires that the application meets all the epistemic values of science, and not only that of fruitfulness.

6.2. *What kinds of skills are required for the successful application of a scientific model to a phenomenon?*

The question of what skills are required for the successful application of models to phenomena is one of the central issues in the case study of cognitive science and to a lesser degree also in the case study of neobehaviorism. The case studies illustrate that the successful application of models to phenomena involves several skillful activities.

First, the successful application of models to phenomena involves making a connection between features of the model and features of the phenomenon. This requires making judgments concerning relevant similarities between model and phenomenon. For instance, it requires a judgment to recognize that obtaining food after depressing a bar and receiving an electric shock can both be conceptualized as instances of 'reinforcement.' By specifying the meaning of the theoretical terms in his learning model, such as 'reinforcement,' Hull was able to connect his model to concrete psychological phenomena. The different ways in which Pollack and Miller connected Shannon's model of a communication system to the phenomenon of the 'magical number seven' illustrates that the specification of the meaning of the theoretical terms in a model is not a matter of course. Whereas Pollack specified the receiver as the listener, Miller specified the communication channel as such. Broadbent, who realized that conceptualizing cognitive phenomena in information-theoretical terms is a technique that has to be mastered, introduced a railway system analogy as a helpful aid for becoming acquainted with the technique of anatomizing psychological phenomena into informational events.

Second, the successful application of models to phenomena involves reasoning via the model. That this is a skillful activity is illustrated in the example of Tolman who was able to recognize qualitatively characteristic consequences of his model of the behavior of rats in mazes by means of the technique of imagining how he would behave if he were a rat. This technique involved empathic imagining, which Tolman used to imagine the 'demands' and 'hypotheses' of the rats in the maze, and means-end reasoning, which he used to imagine how rats would behave as a result of these 'demands' and 'hypotheses.' Another example is Broadbent who was able to recognize qualitatively characteristic consequences of his mechanical model of attention, consisting

of a Y-shaped tube with branches in which balls could be placed, due to his experience with causal agency and his skills of visualization and causal reasoning.

6.3. *Which kind of virtues can render a model intelligible to its users?*

Whether the users of a model are able to apply that model successfully to phenomena depends on the match between their skills and the virtues of the model. The case studies, which focus on the use of theoretical models, show that the intelligibility of this kind of models depends on the surplus meaning of the theoretical terms in the model. The surplus meaning of theoretical terms can have various sources. Their origin can be informal, mechanistic interpretations that are supplemented by the formal definitions of the terms. Another source can be the naming of the theoretical terms. The use of the 'mind-as-information-processing-device' metaphor in cognitive psychology shows that technological metaphors also function as sources for surplus meaning.

The surplus meaning of the theoretical terms can render a theoretical model intelligible to its users. For instance, due to the surplus meaning of terms such as 'demand' and 'hypotheses,' Tolman's model of the behavior of rats in mazes possesses the virtue of being anthropomorphically interpretable and facilitating means-end reasoning. Hull also used theoretical terms that were anthropomorphically interpretable. Moreover, he used terms such as 'habit strength' that had a causal-mechanical surplus meaning that made them accessible to causal reasoning. Similarly, Broadbent's mechanical model of attention was intelligible because of its causal-mechanical surplus meaning.

6.4. *On what kind of pragmatic and contextual factors does intelligibility depend?*

The attribution of the value of intelligibility to a model depends on several pragmatic and contextual factors. For instance, it can depend on the availability of certain metaphors, such as the 'mind-as-information-processing-device' metaphor on which the information-theoretical models of early cognitive psychologists were based and which facilitated the application of these models to cognitive phenomena.

These models did not provide understanding for psychologists who were not familiar with this metaphor.

Another factor can be the objectives of the researcher. For instance, because early cognitive psychologists like Pollack and Broadbent were mainly engaged in applied psychology, their objectives differed from Neisser's, who was primarily occupied with theoretical psychology. Because of that, they were interested in other aspects of cognition than Neisser was. As a result, they did not share Neisser's objection to the application of Shannon's model of a communication system to cognitive phenomena. Whereas they focused on the similarities between the model and the phenomena and considered the model to be intelligible, Neisser focused on the differences, such as the fact that cognitive processes are active, whereas in Shannon's model information processing is passive, and consequently argued that the model did not provide understanding.

That intelligibility is a pragmatic and context-dependent concept might give the impression that it depends on the idiosyncrasies and changing tastes of scientists. However, within scientific communities scientists should be able to apply the same models. Therefore, it is important that scientists working in the same research tradition do not vary much in their views about intelligibility. For instance, when cognitive psychology was developed as a new scientific discipline, it was important that psychologists possessed the skills to work with information-theoretical concepts. At that time, however, the 'mind-as-information-processing-device' metaphor was not common knowledge. Therefore, early cognitive psychologists made an effort to promote this metaphor. They tried to familiarize fellow psychologists with the use of information theory and attempted to give them the required skills. This was necessary because only if fellow psychologists were familiar with the metaphor were they able to apply the information-theoretical models of the early cognitive psychologists.

For instance, Miller showed how experiments in absolute judgment could be reinterpreted as experiments on the capacity of people to transmit information, and Broadbent constructed simple models that were meant to cultivate the techniques that were necessary to work with information-theoretical concepts. One of these techniques was the atomization of phenomena into informational events, which is necessary to account for these phenomena by means of flow charts.

One way in which Broadbent tried to give a feeling for the use of flow charts, in which information flows from one stage to another, was by drawing an analogy with railway systems. Conceptualizing the phenomena as informational events involves the recognition of relevant similarities between the information-theoretical models and the psychological phenomena. Broadbent tried to elucidate this to his colleagues by means of a simple mechanical model in which tubes and flaps represented information-theoretical concepts such as communication channels and information filters. The use of this model involves skills that practically everybody masters.

6.5. *Is the characterization of science advocated in this study useful for the explanatory and normative tasks of philosophy of science?*

Major aims of philosophy of science are the description, explanation, and normative appraisal of (the results of) scientific practices. A common feature of these aims is to provide a more or less general characterization of science. The characterization of science advocated in this study asserts that intelligibility of models is an epistemic value of science. In the case study of neo-behaviorism, this characterization is used mainly to provide a more comprehensive explanation of developments in psychology that are, in themselves, well known from textbooks on the history of psychology. In the 1950s neo-behaviorists gradually came to appreciate the surplus meaning of theoretical terms. They started to advocate the use of hypothetical constructs, which possess surplus meaning, instead of intervening variables, which were considered not to possess surplus meaning. This development in psychology, which eventually resulted in the rise of cognitive psychology, can be understood as being driven by the normative question of how the value of intelligibility of models should be incorporated into psychology. For instance, Tolman converted publicly to the use of hypothetical constructs because he realized that the surplus meaning of these hypothetical constructs provided the models in which they were used with a conceptual substrate that might facilitate the successful application of the models.

Although the analysis of the case studies has been mainly descriptive and explanatory, the account of understanding developed in this study also has normative implications. It entails that, because models

in science should be intelligible, the use of hypothetical constructs is to be preferred over the use of intervening variables. Thus, contrary to the opinion of some strict neo-behaviorists such as Marx, Tolman's conversion should be appraised positively. Scientists should aim at intelligible models.

In sum, scientific understanding is an important aim of science that has epistemic significance. Adding the value of intelligibility of models to the list of epistemic values that are constitutive of science results in a characterization of science that can be used in the explanation and normative appraisal of scientific practices and results.

References

Baars, B.J. 1986. *The Cognitive Revolution in Psychology*. New York: Guilford Press.

Baddeley, A., and L. Weiskrantz, eds. 1993. *Attention: Selection, Awareness, and Control: A Tribute to Donald Broadbent*. Oxford: Clarendon Press.

Bailer-Jones, D.M. 1999. Tracing the Development of Models in the Philosophy of Science. In *Model-Based Reasoning in Scientific Discovery*, edited by L. Magnani, N.J. Nersessian, and P. Thagard, 23–40. New York: Kluwer Academic/Plenum.

———. 2003. When Scientific Models Represent. *International Studies in the Philosophy of Science* 17:59–74.

Barnes, E. 1992. Explanatory Unification and Scientific Understanding. In *PSA 1992*, vol. 1, edited by D. Hull, M. Forbes, and K. Okruhlik, 3–12. East Lansing: Philosophy of Science Association.

Beach, F.A. 1959. Clark Leonard Hull. In *Biographical Memoirs*, vol. 33, edited by the National Academy of Sciences of the United States of America, 125–141. Washington: National Academy Press.

Bergmann, G. 1940a. On Some Methodological Problems of Psychology. *Philosophy of Science* 7:205–219.

———. 1940b. The Subject Matter of Psychology. *Philosophy of Science* 7:415–433.

———. 1953. Theoretical Psychology. *Annual Review of Psychology* 4:435–458.

———, and K.W. Spence. 1941. Operationism and Theory in Psychology. *Psychological Review* 48:1–14.

Berlyne, D.E. 1950. Stimulus Intensity and Attention in Relation to Learning Theory. *Quarterly Journal of Experimental Psychology* 2:71–75.

Boden, M.A. 2006. *Mind as Machine: A History of Cognitive Science*. Oxford: Clarendon Press/Oxford University Press.

Bogen, J., and J.F. Woodward. 1988. Saving the Phenomena. *Philosophical Review* 97:303–352.

Boltzmann, L.E. 1896/1964. *Lectures on Gas Theory*. Translated by S.G. Brush. Berkeley: University of California Press.

Boon, M., and T. Knuuttila. 2009. Models as Epistemic Tools in Engineering Sciences: A Pragmatic Approach. In *Philosophy of Technology and Engineering Sciences*, edited by A.W.M. Meijers, 687–719. Amsterdam: Elsevier.

Bridgman, P.W. 1927. *The Logic of Modern Physics*. New York: Macmillan.

Broadbent, D.E. 1952. Listening to One of Two Synchronous Messages. *Journal of Experimental Psychology* 44:51–55.

———. 1954a. Effects on Behaviour from Noises of High and Low Pitch. *Cambridge University, Applied Psychology Unit, Report* 222.

———. 1954b. The Role of Auditory Localization in Attention and Memory Span. *Journal of Experimental Psychology* 47:191–196.

———. 1956a. The Concept of Capacity and the Theory of Behaviour. In *Proceedings of the Third London Symposium on Information Theory*, edited by E.C. Cherry, 354–359. London: Butterworths.

———. 1956b. Successive Responses to Simultaneous Stimuli. *Quarterly Journal of Experimental Psychology* 8:145–152.

———. 1957. A Mechanical Model for Human Attention and Immediate Memory. *Psychological Review* 64:205–215.

———. 1958. *Perception and Communication*. London: Pergamon Press.

Brown, H.I. 1988. *Rationality*. London: Routledge.

Burian, R.M. 2001. The Dilemma of Case Studies Resolved: The Virtues of Using Case Studies in the History and Philosophy of Science. *Perspectives on Science* 9:383–404.

Carnap, R. 1939. Foundations of Logic and Mathematics. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.

———. 1956. The Methodological Character of Theoretical Concepts. In *Minnesota Studies in the Philosophy of Science*, vol. 1, *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, edited by H. Feigl and M. Scriven, 38–76. Minneapolis: University of Minnesota Press.

———. 1963. Intellectual Autobiography. In *The Philosophy of Rudolf Carnap*, edited by P.A. Schilpp, 1–84. LaSalle: Open Court.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.

———, T. Shomar, and M. Suárez. 1995. The Tool Box of Science: Tools for the Building of Models with a Superconductivity Example. *Poznań Studies in the Philosophy of the Sciences and the Humanities* 44:137–149.

Cherry, E.C. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustic Society of America* 25:975–979.

Conrad, R. 1951. Speed and Load Stress in a Sensory-Motor Skill. *British Journal of Industrial Medicine* 8:1–7.

da Costa, N.C.A., and S. French. 2003. *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

De Regt, H.W. 2004. Discussion Note: Making Sense of Understanding. *Philosophy of Science* 71:98–109.

———. 2005. Scientific Realism in Action: Molecular Models and Boltzmann's *Bildtheorie*. *Erkenntnis* 63:205–230.

———. 2009. Understanding and Scientific Explanation. In *Scientific Understanding: Philosophical Perspectives*, edited by H.W. de Regt, S. Leonelli, and K. Eigner, 21–42. Pittsburgh: University of Pittsburgh Press.

———. Forthcoming. Explanation. In *The Continuum Companion to the Philosophy of Science*, edited by S. French and J. Saatsi. London: Continuum Press.

———, and D. Dieks. 2005. A Contextual Approach to Scientific Understanding. *Synthese* 144:137–170.

———, S. Leonelli, and K. Eigner, eds. 2009. *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

Dowe, P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Eigner, K. 2009. Understanding in Psychology: Is Understanding a Surplus? In *Scientific Understanding: Philosophical Perspectives*, edited by H.W. de Regt, S. Leonelli, and K. Eigner, 271–297. Pittsburgh: University of Pittsburgh Press.

Feest, U. 2005. Operationism in Psychology: What the Debate Is About, What the Debate Should Be About. *Journal of the History of the Behavioral Sciences* 41:131–149.

Fernandez-Duque, D., and M.L. Johnson. 1999. Attention Metaphors: How Metaphors Guide the Cognitive Psychology of Attention. *Cognitive Science* 23:83–116.

Flexer, A. 1995. Connectionists and Statisticians, Friends or Foes? In *From Natural to Artificial Neural Computation: Proceedings of the International Workshop on Artificial Neural Networks, Malaga-Torremolinos, Spain, June 7–9, 1995*, edited by J. Mira and F. Sandoval, 454–461. Berlin: Springer.

Freud, S. 1940. *Gesammelte Werke*, vol. 15. London: Imago.

Friedman, M. 1974. Explanation and Scientific Understanding. *Journal of Philosophy* 71:5–19.

Frigg, R. 2006. Scientific Representation and the Semantic View of Theories. *Theoria* 55:49–65.

———, and S. Hartmann. 2006. Models in Science. In *Stanford Encyclopedia of Philosophy*, edited by E.N. Zalta. Stanford: The Metaphysics Research Lab.

Furedy, J.J. 1988. On the Relevance of Philosophy for Psychological Research: A Preliminary Analysis of Some Influences of Andersonian Realism. *Australian Journal of Psychology* 40:71–77.

Gardner, H. 1985. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Garner, W.R., and H.W. Hake. 1951. The Amount of Information in Absolute Judgments. *Psychological Review* 58:446–459.

Giere, R.N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

———. 1999a. *Science Without Laws*. Chicago: University of Chicago Press.

———. 1999b. Using Models to Represent Reality. In *Model-based Reasoning In Scientific Discovery*, edited by L. Magnani, N.J. Nersessian, and P. Thagard, 41–57. New York: Kluwer Academic/Plenum.

———. 2004. How Models Are Used to Represent Reality. *Philosophy of Science* 71:742–752.

Greenwood, J.D. 1999. Understanding the "Cognitive Revolution" in Psychology. *Journal of the History of the Behavioral Sciences* 35:1–22.

Gutting, G. 1980. Science as Discovery. *Revue Internationale de Philosophie* 34:26–48.

Hanson, N.R. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hartmann, S. 1999. Models and Stories in Hadron Physics. In *Models as Mediators: Perspectives on Natural and Social Science*, edited by M.S. Morgan and M. Morrison, 241–281. Cambridge: Cambridge University Press.

Hebb, D.O. 1949. *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

Hempel, C.G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

———. 1969. Logical Positivism and the Social Sciences. In *The Legacy of Logical Positivism*, edited by P. Achinstein and S.F. Barker, 163–194. Baltimore: Johns Hopkins Press.

———, and P. Oppenheim. 1948. Studies in the Logic of Explanation. *Philosophy of Science* 15:135–175.

Hergenhahn, B.R. 1997. *An Introduction to the History of Psychology*. 3rd ed. Pacific Grove: Brooks/Cole.

Hoyningen-Huene, P. 1987. Context of Discovery and Context of Justification. *Studies in History and Philosophy of Science* 18:501–515.

Hull, C.L. 1935. The Conflicting Psychologies of Learning: A Way Out. *Psychological Review* 42:491–516.

———. 1937. Mind, Mechanism, and Adaptive Behavior. *Psychological Review* 44:1–32.

———. 1938. Logical Positivism as a Constructive Methodology in the Social Sciences. *Einheitswissenschaft* 6:35–38.

———. 1943a. The Problem of Intervening Variables in Molar Behavior Theory. *Psychological Review* 50:273–291.

———. 1943b. *Principles of Behavior*. New York: Appleton-Century-Crofts.

———, and H.D. Baernstein. 1929. A Mechanical Parallel to the Conditioned Reflex. *Science* 70:14–15.

Humphreys, P. 1989. *The Chances of Explanation*. Princeton: Princeton University Press.

Innis, N.K. 1999. Edward C. Tolman's Purposive Behaviorism. In *Handbook of Behaviorism*, edited by W.T. O'Donohue and R.F. Kitchener, 97–117. San Diego: Academic Press.

Kirschenmann, P.P. 2001. Local and Normative Rationality of Science. In *Science, Nature and Ethics: Critical Philosophical Studies*, 3–14. Delft: Eburon.

Kitchener, R.F. 2004. Logical Positivism, Naturalistic Epistemology, and the Foundations of Psychology. *Behavior and Philosophy* 32:37–54.

Kitcher, P. 1981. Explanatory Unification. *Philosophy of Science* 48:507–531.

———. 1989. Explanatory Unification and the Causal Structure of the World. In *Scientific Explanation*, edited by P. Kitcher and W.C. Salmon, 410–505. Minneapolis: University of Minnesota Press.

Knuuttila, T., and M. Merz. 2009. Understanding by Modeling: An Objectual Approach. In *Scientific Understanding: Philosophical Perspectives*, edited by H.W. de Regt, S. Leonelli, and K. Eigner, 146–168. Pittsburgh: University of Pittsburgh Press.

Koch, S. 1941. The Logical Character of the Motivation Concept. *Psychological Review* 48:15–38 and 127–154.

———. 1954. Clark L. Hull. In *Modern Learning Theory*, edited by W.K. Estes, S. Koch, K. MacCorquodale, P.E. Meehl, C.G. Mueller, W.N. Schoenfeld, and W.S. Verplanck, 1–176. New York: Appleton-Century-Crofts.

———. 1964. Psychology and Emerging Conceptions of Knowledge as Unitary. In *Behaviorism and Phenomenology*, edited by T.W. Wann, 1–41. Chicago: University of Chicago Press.

Krech, D. 1950. Dynamic Systems, Psychological Fields and Hypothetical Constructs. *Psychological Review* 57:283–290.

Kuhn, T.S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

———. 1973/1977. Objectivity, Values, and Theory Choice. In *The Essential Tension*, 320–339. Chicago: University of Chicago Press.

Lacey, H. 2005. *Is Science Value Free? Values and Scientific Understanding*. 2nd ed. London: Routledge.

La Mettrie, J.O. de. 1748/1999. *L'Homme-Machine*. Paris: Denoël/Gonthier.

Leahey, T.H. 1980a. *A History of Psychology: Main Currents in Psychological Thought*. Englewood Cliffs: Prentice Hall.

———. 1980b. The Myth of Operationism. *Journal of Mind and Behavior* 1:131–140.

Leonelli, S. 2007. *Weed for Thought: Using* Arabidopsis Thaliana *to Understand Plant Biology*. Ph.D. diss., VU University Amsterdam. http://hdl.handle.net/1871/10703.

———. 2009. Understanding in Biology: The Impure Nature of Biological Knowledge. In *Scientific Understanding: Philosophical Perspectives*, edited by H.W. de Regt, S. Leonelli, and K. Eigner, 189–209. Pittsburgh: University of Pittsburgh Press.

Lindzey, G. 1953. Hypothetical Constructs, Conventional Constructs, and the Use of Physiological Data in Psychological Theory. *Psychiatry* 16:27–33.

———, ed. 1954. Handbook of Social Psychology. Cambridge: Addison-Wesley.

Longino, H.E. 1990. *Science as Social Knowledge*. Princeton: Princeton University Press.

Lumsdaine, A.A. 1953. Audio-Visual Research in the U.S. Air Force. *Audio-Visual Communication Review* 1:76–90.

McAllister, J.W. 1986. Theory-Assessment in the Historiography of Science. *British Journal of the Philosophy of Science* 37:315–333.

MacCorquodale, K., and P.E. Meehl. 1948. On a Distinction between Hypothetical Constructs and Intervening Variables. *Psychological Review* 55:95–107.

MacKay, D.M. 1952. The Nomenclature of Information Theory. In *Cybernetics: Circular, Causal and Feedback Mechanisms in Biological and Social Systems, Transactions of the Eighth Conference,* edited by H. von Foerster, 222–235. New York: Josiah Macy, Jr. Foundation.

———. 1969. *Information, Mechanism and Meaning*. Cambridge: MIT Press.

Mackenzie, B.D. 1977. *Behaviourism and the Limits of Scientific Method*. London: Routledge/Kegan Paul.

McMullin, E. 1983. Values in Science. In *PSA 1982*, vol. 2, edited by P.D Asquith and T. Nickles, 3–28. East Lansing: Philosophy of Science Association.

Marx, M.H. 1951. Intervening Variable or Hypothetical Construct? *Psychological Review* 58:235–247.

Maze, J.R. 1954. Do Intervening Variables Intervene? *Psychological Review* 61:226–234.

Miller, G.A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63:81–97.

Morgan, M.S., and M. Morrison, eds. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Morrison, M., and M.S. Morgan. 1999a. Introduction. In *Models as Mediators: Perspectives on Natural and Social Science,* edited by M.S. Morgan and M. Morrison, 1–9. Cambridge: Cambridge University Press.

———, and M.S. Morgan. 1999b. Models as Mediating Instruments. In *Models as Mediators: Perspectives on Natural and Social Science,* edited by M.S. Morgan and M. Morrison, 10–37. Cambridge: Cambridge University Press.

Neisser, U. 1967. *Cognitive Psychology*. New York: Appleton-Century-Crofts.

———. 1976. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco: W.H. Freeman.

———. 2007. Ulric Neisser. In *A History of Psychology in Autobiography*, vol. 9, edited by G. Lindzey and W.M. Runyan, 269–301. Washington: American Psychological Association.

Newton-Smith, W.H. 2000. Explanation. In *A Companion to the Philosophy of Science*, edited by W.H. Newton-Smith, 127–133. Oxford: Blackwell.

Nickles, T. 1986. Remarks on the Use of History as Evidence. *Synthese* 69:253–266.

———. 2001. Discovery. In *A Companion to the Philosophy of Science*, edited by W.H. Newton-Smith, 85–96. Oxford: Blackwell.

Palmer, S.E., and R. Kimchi. 1986. The Information Processing Approach to Cognition. In *Approaches to Cognition: Contrasts and Controversies*, edited by T.J. Knapp and L. Robertson, 37–77. Hillsdale: Erlbaum.

Partridge, D. 1987. What's Wrong With Neural Architectures. *University of Exeter, Computer Science Department, Research Report* 142.

———. 1991. *A New Guide to Artificial Intelligence*. Norwood: Ablex.

Polanyi, M. 1967. *The Tacit Dimension*. New York: Anchor Books.

Pollack, I. 1952. The Information of Elementary Auditory Displays. *Journal of the Acoustical Society of America* 24:745–749.

Popper, K.R. 1934/1959. *The Logic of Scientific Discovery*. London: Hutchinson.

———. 1935. "Induktionslogik" und "Hypothesenwahrscheinlichkeit." *Erkenntnis* 5:170–172.

———. 1962/1989. *Conjectures and Refutations: The Growth of Scientific Knowledge*. 5th rev. ed. London: Routledge.

Poulton, E.C. 1953. Two-Channel Listening. *Journal of Experimental Psychology* 46:91–96.

———. 1956. Listening to Overlapping Calls. *Journal of Experimental Psychology* 52:334–339.

Radder, H. 1996. *In and About the World: Philosophical Studies of Science and Technology*. Albany: State University of New York Press.

———. 1997. Philosophy and History of Science: Beyond the Kuhnian Paradigm. *Studies in History and Philosophy of Science* 28:633–655.

———. 2006. *The World Observed/The World Conceived*. Pittsburgh: University of Pittsburgh Press.

Reichenbach, H. 1938. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.

Salmon, W.C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

———. 1990. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

———. 1998. *Causality and Explanation*. Oxford: Oxford University Press.

Sanders, C. 1972. *De behavioristische revolutie in de psychologie (The Behavioristic Revolution in Psychology)*. Deventer: Van Loghum Slaterus.

Schickore, J., and F. Steinle. 2006. Introduction: Revisiting the Context Distinction. In *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*, edited by J. Schickore and F. Steinle, vii–xix. Dordrecht: Springer.

Schultz, D.P., and S.E. Schultz. 2000. *A History of Modern Psychology*. Orlando: Harcourt Brace.

Schurz, G. 1999. Explanation as Unification. *Synthese* 120:95–114.

Scriven, M. 1962/1988. Explanations, Predictions and Laws. In *Theories of Explanation*, edited by J.C. Pitt, 51–74. Oxford: Oxford University Press.

Shannon, C.E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27:379–423 and 623–656.

———, and W. Weaver. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Smith, L.D. 1986. *Behaviorism and Logical Positivism: A Reassessment of the Alliance*. Stanford: Stanford University Press.

Spence, K.W. 1944. The Nature of Theory Construction in Contemporary Psychology. *Psychological Review* 51:47–68.

———. 1948. The Postulates and Methods of "Behaviorism." *Psychological Review* 55:67–78.

———. 1950. Cognitive Versus Stimulus-Response Theories of Learning. *Psychological Review* 57:159–172.

Sperry, R.W. 1995. The Future of Psychology. *American Psychologist* 50:505–506.

Spieth, W., J. Curtis, and J.C. Webster. 1954. Responding to One of Two Simultaneous Messages. *Journal of the Acoustical Society of America* 26:391–396.

Still, A. 1997a. Tolman, Edward Chace. In *Biographical Dictionary of Psychology*, edited by N. Sheehy, A.J. Chapman, and W.A. Conroy, 576–578. London: Routledge.

———. 1997b. Hull, Clark L. In *Biographical Dictionary of Psychology*, edited by N. Sheehy, A.J. Chapman, and W.A. Conroy, 284–286. London: Routledge.

Suárez, M. 1999. The Role of Models in the Application of Scientific Theories: Epistemological Implications. In *Models as Mediators: Perspectives on Natural and Social Science*, edited by M.S. Morgan and M. Morrison, 168–196. Cambridge: Cambridge University Press.

———. 2003. Scientific Representation: Against Similarity and Isomorphism. *International Studies in the Philosophy of Science* 17:225–244.

———. 2004. An Inferential Conception of Scientific Representation. *Philosophy of Science* 71:S767–S779.

Suppe, F. 1977. *The Structure of Scientific Theories*. Urbana: University of Illinois Press.

Suppes, P.C. 1960. A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. *Synthese* 24:287–301.

———. 1962. Models of Data. In *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, edited by E. Nagel, P.C. Suppes, and A. Tarski, 252–261. Stanford: Stanford University Press.

———. 1967. What is a Scientific Theory? In *Philosophy of Science Today*, edited by S. Morgenbesser, 55–67. New York: Basic Books.

———. 2002. *Representation and Invariance of Scientific Structures*. Stanford: Center for the Study of Language and Information Publications.

Thomson, W. (Lord Kelvin). 1884/1987. Lecture XI. In *Kelvin's Baltimore Lectures and Modern Theoretical Physics: Historical and Philosophical Perspectives*, edited by R. Kargon and P. Achinstein, 106–114. Cambridge: MIT Press.

Tolman, E.C. 1932. *Purposive Behavior in Animals and Men*. New York: Century.

———. 1935/1966. Psychology Versus Immediate Experience. In *Behavior and Psychological Man*, 94–114. Berkeley: University of California Press.

———. 1936/1966. Operational Behaviorism and Current Trends in Psychology. In *Behavior and Psychological Man*, 115–129. Berkeley: University of California Press.

———. 1938/1966. The Determiners of Behavior at a Choice Point. In *Behavior and Psychological Man*, 144–178. Berkeley: University of California Press.

———. 1949. Discussion from "Interrelationships between Perception and Personality: A Symposium." *Journal of Personality* 18:48–50.

———. 1952. Autobiography. In *A History of Psychology in Autobiography*, vol. 4, edited by E.G. Boring, H.S. Langfeld, H. Werner, and R.M. Yerkes, 323–339. Worcester: Clark University Press.

Trout, J.D. 2002. Scientific Explanation and the Sense of Understanding. *Philosophy of Science* 69:212–233.

Van Fraassen, B.C. 1980. *The Scientific Image*. Oxford: Clarendon Press.

———. 2000. The Semantic Approach to Scientific Theories. In *The Philosophy of Science: A Collection of Essays*, vol. 2, *The Nature of Scientific Theory*, edited by L. Sklar, 175–194. New York: Garland.

———. 2004. Science as Representation: Flouting the Criteria. *Philosophy of Science* 71:S794–S804.

Van Lunteren, F.H. 1991. *Framing Hypotheses*. Ph.D. diss., Utrecht University.

Vicedo, M. 1993. Is the History of Science Relevant to the Philosophy of Science? In *PSA 1992*, vol. 2, edited by D. Hull, M. Forbes, and K. Okruhlik, 280–290. East Lansing: Philosophy of Science Association.

Watson, J.B. 1913. Psychology as the Behaviorist Views It. *Psychological Review* 20:158–177.

Webster, J.C., and P.O. Thompson. 1954. Responding to Both of Two Overlapping Messages. *Journal of the Acoustical Society of America* 26:396–402.

Weiskrantz, L. 1994. Donald Eric Broadbent: 6 May 1926–10 April 1993. In *Biographical Memoirs of Fellows of the Royal Society*, vol. 40, 33–42. London: Royal Society.

Whitehead, A.N., and B.A.W. Russell. 1927. *Principia Mathematica*. 2nd ed. Cambridge: Cambridge University Press.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Woodward, J.F. 1989. Data and Phenomena. *Synthese* 79:393–472.

———. 2000. Data, Phenomena, and Reliability. *Philosophy of Science* 67:S163–S179.

———. 2009. Data and Phenomena: A Restatement and Defense. *Synthese*. DOI 10.1007/s11229-009-9618-5.

Zuriff, G.E. 1985. *Behaviorism, A Conceptual Reconstruction*. New York: Columbia University Press.

Index of Names

List of Figures

Samenvatting (Summary in Dutch)

Het begrijpen van het begrip van psychologen: De toepassing van begrijpelijke modellen op verschijnselen

Het begrijpen van verschijnselen is een belangrijk epistemisch doel van de wetenschap. Onder wetenschappers en wetenschapsfilosofen bestaat echter de neiging het belang van wetenschappelijk begrip te bagatelliseren door het te omschrijven als een psychologisch bijproduct van wetenschappelijke verklaringen. De neobehaviorist Edward C. Tolman had het bijvoorbeeld over de psychologische behoefte om de "innerlijke spanningen" te verlichten die ontstaan bij de confrontatie met onverklaarde verschijnselen.

Volgens traditionele opvattingen in de wetenschapsfilosofie verwijst begrijpen naar psychologische en pragmatische aspecten van wetenschappelijke verklaringen, en daarmee niet naar de epistemologische aspecten ervan. In filosofische beschouwingen over wetenschap zou daarom geen plaats zijn voor wetenschappelijk begrijpen. Het scherpe onderscheid tussen pragmatische aspecten, die afhankelijk zijn van de persoon die een verklaring geeft of ontvangt, en epistemische aspecten, die dat niet zijn, is echter problematisch. In dit proefschrift betoog ik dat wetenschappelijk begrijpen een pragmatische notie is die van epistemisch belang is.

De stelling die wordt verdedigd in dit proefschrift is dat wetenschappers een verschijnsel begrijpen als ze in staat zijn er met succes een wetenschappelijk model op toe te passen. Een centrale notie hierbij is de begrijpelijkheid (of intelligibiliteit) van het model. Dit is een waarde, toegekend aan een model door de gebruikers ervan, die weergeeft in hoeverre ze in staat zijn het model succesvol toe te passen. Deze toepasbaarheid hangt niet alleen af van de eigenschappen van het model maar ook van de vaardigheden van de wetenschappers, bijvoorbeeld de vaardigheid om de relevante overeenkomsten tussen model en verschijnsel te beoordelen en de vaardigheid om karakteristieke consequenties van het model in te zien. De notie van de

begrijpelijkheid van een model, en dan met name het epistemisch belang en het pragmatische karakter ervan, is in dit proefschrift onderzocht met behulp van twee historische gevalstudies uit de psychologie.

De casus over het neobehaviorisme

Het hoofddoel van de gevalstudie over het invloedrijke neobehaviorisme is het aantonen van het epistemisch belang van wetenschappelijk begrijpen door het ontkrachten van een schijnbaar plausibel tegenvoorbeeld. Tolmans uitlating over de psychologische aspecten van wetenschappelijke verklaringen illustreert de positivistische wetenschapsvisie van de neobehavioristen, waarin geen rol is weggelegd voor wetenschappelijk begrip. Een analyse van de wetenschappelijke praktijk van deze psychologen laat echter zien dat de neobehavioristen ondanks hun positivistische beginselen impliciet wel degelijk streefden naar het opstellen van begrijpelijke modellen.

Een voorbeeld van zo'n model is Tolmans model voor het gedrag van ratten in een doolhof. Dit model, dat hij opstelde ter verduidelijking van zijn door het logisch positivisme geïnspireerde methode van het operationele behaviorisme, was begrijpelijk voor Tolman doordat het de mogelijkheid bood de situatie van de ratten in het doolhof op een antropomorfe wijze te interpreteren. Doordat de termen in het model zoals 'behoefte' en 'verwachtingen,' die door hem waren geïntroduceerd als theoretische termen, konden worden verbonden met ervaringen uit het dagelijks leven, overstegen ze qua betekenis hun operationele definities. Deze zogenoemde "surplusbetekenis" stelde Tolman in staat om zich met behulp van het model in te leven in de ratten en zich een voorstelling te maken van hun gedrag. Dat dit zijn model begrijpelijk maakte, was voor Tolman echter geen reden om zich positief uit te laten over het gebruik van theoretische termen met surplusbetekenis. Volgens hem moest zijn model gezien worden als een tussenstap in de ontwikkeling naar objectieve wetenschappelijke uitspraken. Om tot zulke uitspraken te komen was het volgens hem nodig de theoretische termen in zijn model om te zetten in objectieve termen vrij van surplusbetekenis. Ik betoog echter dat dit vanuit epistemologisch oogpunt ongewenst is.

Dat de surplusbetekenis van theoretische termen epistemische relevantie heeft, blijkt uit een analyse van neobehavioristische modellen

die zijn opgesteld door Clark L. Hull. De theoretische termen in deze modellen zoals de 'sterkte van een gewoonte' hadden een surplusbetekenis waarvan de oorsprong niet alleen lag in het gebruik van betekenisvolle termen uit het dagelijks leven, maar ook in de informele mechanistische interpretaties die Hull meegaf aan deze termen. Zo bouwde hij bijvoorbeeld een mechanisch apparaat ter illustratie van zijn theoretische principes over het vormen van gewoonten. De surplusbetekenis die dit gaf aan zijn theoretische termen zoals 'gewoonte' maakte het hem mogelijk om de relevante overeenkomsten tussen zijn modellen en de verschijnselen te beoordelen en de karakteristieke consequenties van de modellen in te zien. Zonder deze surplusbetekenis zou Hull niet in staat zijn geweest zijn theoretische modellen toe te passen op de relevante verschijnselen, waardoor deze modellen dan geen epistemische waarde zouden hebben.

Rond 1950 gingen de neobehavioristen, die aanvankelijk vanwege hun positivistische uitgangspunten gekant waren tegen het gebruik van theoretische termen met surplusbetekenis, dit gebruik positiever waarderen. Ook Tolman schaarde zich onder de voorstanders. Deze ontwikkeling, die wel wordt beschouwd als een belangrijke factor in de overgang van neobehaviorisme naar cognitieve psychologie, kan worden gezien als het gevolg van een impliciet ongenoegen over de logisch-positivistische wetenschapsvisie, waarin de begrijpelijkheid van wetenschappelijke modellen onvoldoende wordt gewaardeerd. Het besef ontstond dat de surplusbetekenis van theoretische termen in wetenschappelijke modellen een belangrijke epistemische functie heeft, namelijk het begrijpelijk maken van deze modellen.

De casus over cognitieve psychologie

Het hoofddoel van de gevalstudie over cognitieve psychologie is het analyseren van de pragmatische en contextuele aspecten van wetenschappelijk begrip. Het gaat hierbij met name om de vaardigheden van wetenschappers die nodig zijn om een model toe te passen op een verschijnsel, zoals het herkennen van verbanden tussen eigenschappen van het model en eigenschappen van het verschijnsel, en het afleiden van consequenties van het model. De modellen die in de beginjaren van de cognitieve psychologie werden ontwikkeld, waren gebaseerd op de metafoor van de mens als informatieverwerker. Het basisidee

was dat, analoog aan een telegraaf of telefoon, de informatieverwerking van een mens kan worden beschreven in termen van de nieuw ontwikkelde informatietheorie, zoals informatiebron, zender, ontvanger en communicatiekanaal. Op het eerste gezicht lijkt het toepassen van informatietheoretische modellen wellicht dusdanig triviaal dat het overdreven is om hier te spreken van "benodigde vaardigheden". Dat het toepassen van deze modellen echter niet triviaal is, blijkt uit de verschillende manieren waarop wetenschappers dit deden.

Cognitieve psychologen verschilden bijvoorbeeld in de wijze waarop ze met behulp van een informatietheoretisch model psychologische verschijnselen conceptualiseerden. Waar Irwin Pollack de proefpersoon in een bepaald experiment zag als de ontvanger van informatie, beschouwde zijn collega George A. Miller deze persoon juist als communicatiekanaal, een duidelijk ander aspect van hetzelfde informatietheoretische model. Dit laat zien dat het conceptualiseren van cognitieve verschijnselen in informatietheoretische termen niet vanzelfsprekend is. De psycholoog Donald E. Broadbent, die het gebruik van de terminologie van de informatietheorie in zijn vakgebied aanmoedigde, zag het als een techniek die psychologen in de vingers moesten krijgen. Dat het beheersen van deze techniek tegenwoordig mogelijk overkomt als triviaal komt waarschijnlijk door de vertrouwdheid die er nu is met concepten uit de informatietheorie.

Broadbents ontwikkeling van een mechanisch model voor het psychologische verschijnsel 'aandacht' laat zien dat hij besefte dat niet alleen het conceptualiseren van psychologische verschijnselen in informatietechnologische termen een vaardigheid is, maar dat ook het redeneren met modellen die zijn gebaseerd op deze concepten vaardigheden vergt. Aandacht kon volgens hem worden geconceptualiseerd als een mechanisme waarmee informatie wordt gefilterd. In een situatie waarin veel informatie het zenuwstelsel binnenkomt, zoals tijdens een cocktailparty waarin iemand luistert naar een persoon terwijl tegelijkertijd ook anderen spreken, zorgt dit mechanisme ervoor dat irrelevante informatie wordt uitgefilterd. Broadbents uiteenzetting van de informatietheoretische principes waarop dit mechanisme berust, werd door hem verduidelijkt met behulp van een mechanisch model. In dit model wordt de informatiefilter gerepresenteerd door een klep die een arm van een Y-vormige buis kan afsluiten en wordt informatie gerepresenteerd door ballen die de armen van deze buis

binnenkomen. Het kunnen werken met dit model vraagt bekendheid met de werking van causale mechanismen en bekwaamheid in visualiseren en causaal redeneren. Omdat vrijwel iedereen deze vaardigheden bezit, was Broadbent in staat om met dit model zijn theoretische visie op het verschijnsel aandacht duidelijk te maken aan collega's die niet vertrouwd waren met informatietheoretische concepten. Broadbents manier om zijn collega's inzicht te verschaffen in dit psychologische verschijnsel illustreert dat het voor de begrijpelijkheid van een model van belang is dat er een goede aansluiting is tussen de eigenschappen van dit model, zoals de visualiseerbaarheid ervan, en de vaardigheden van de gebruiker.

Of wetenschappers de waarde van begrijpelijkheid toeschrijven aan een concreet model hangt niet alleen af van hun vaardigheden, maar ook van andere pragmatische en contextuele factoren. Zo verschilde de cognitieve psycholoog Ulric Neisser bijvoorbeeld van mening met collega's als Pollack en Broadbent over de begrijpelijkheid van informatietheoretische modellen zoals dat van Broadbent. Doordat Neisser zich primair bezig hield met theoretische psychologie was hij geïnteresseerd in andere aspecten van cognitie dan Pollack en Broadbent, die zich met name richtten op toegepaste psychologie. Terwijl Neisser zich concentreerde op het wezen van cognitie, dat hij beschouwde als een actief proces, hadden zijn collega's vooral belangstelling voor de beperkingen van menselijke informatieverwerking in praktische situaties. Neisser vond het relevant dat informatieverwerking in de informatietheoretische modellen wordt beschreven als een passief proces, en voor hem was dit een reden om te beargumenteren dat deze modellen geen begrip geven van cognitieve processen. Psychologen als Pollack en Broadbent vonden echter met name die aspecten van de modellen relevant die ze konden koppelen aan de beperkingen van menselijke informatieverwerking, en zij waren van mening dat informatietheoretische modellen wel degelijk inzicht geven in deze beperkingen.

Doordat begrijpelijkheid een pragmatisch en contextueel concept is, zou de indruk kunnen ontstaan dat begrijpen volledig afhankelijk is van de karaktertrekken en de wisselende smaak individuele van wetenschappers. Echter, het is wenselijk dat binnen een wetenschappelijke discipline verschillende wetenschappers in staat zijn dezelfde modellen toe te passen. Aangezien de begrijpelijkheid van modellen een

grote rol speelt bij de toepassing van modellen op verschijnselen is het daarom belangrijk dat wetenschappers niet al te zeer verschillen in hun oordeel over de begrijpelijkheid van de modellen die binnen hun discipline worden gehanteerd. Naar mijn idee is dit de reden waarom Broadbent het van belang vond dat zijn mechanische model van aandacht aansloot bij vaardigheden die vrijwel iedereen bezit, en waarom hij zich inspande om zijn collega's de vaardigheid van het conceptualiseren van cognitieve verschijnselen in informatietheoretische termen bij te brengen. Hij besefte dat zijn vakgenoten slechts dan in staat zouden zijn om met behulp van informatietheoretische modellen cognitieve verschijnselen te begrijpen, wanneer ze vertrouwd waren met de metafoor van de mens als informatieverwerker en wanneer ze beschikten over de relevante vaardigheden.

Kortom, wetenschappers moeten streven naar begrijpelijke modellen, niet alleen omdat dit hen zou kunnen helpen bij het verlichten van "innerlijke spanningen" die ze ervaren bij de confrontatie met onverklaarde feiten, maar vooral omdat enkel begrijpelijke modellen succesvol kunnen worden toegepast op verschijnselen. Wetenschappelijk begrip is niet slechts een psychologisch bijproduct van wetenschappelijke verklaringen. Begrijpen is een centraal doel van de wetenschap.