RELIABILITY AND VALIDITY OF EXPERIMENT IN THE NEUROBIOLOGY OF LEARNING AND MEMORY

by

Jacqueline Anne Sullivan

M.S., Neuroscience, University of Pittsburgh, 2003

B.A., Philosophy, Clark University, 1995

Submitted to the Graduate Faculty of

the Department of History and Philosophy of Science, University of

Pittsburgh in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF HISTORY AND PHILOSOPHY OF SCIENCE, UNIVERSITY OF PITTSBURGH

This dissertation was presented

by

Jacqueline Anne Sullivan

It was defended on

June 29, 2007

and approved by

Peter K. Machamer, History and Philosophy of Science, University of Pittsburgh

Sandra D. Mitchell, History and Philosophy of Science, University of Pittsburgh

Edouard Machery, History and Philosophy of Science, University of Pittsburgh

Kenneth F. Schaffner, History and Philosophy of Science, University of Pittsburgh

Edda Thiels, Neurobiology, University of Pittsburgh

German Barrionuevo, Neuroscience, University of Pittsburgh

Richard Grush, Department of Philosophy, University of California San Diego

Dissertation Director: Peter K. Machamer, History and Philosophy of Science, University of Pittsburgh

RELIABILITY AND VALIDITY OF EXPERIMENT IN THE NEUROBIOLOGY OF LEARNING AND MEMORY

Jacqueline Anne Sullivan, PhD

University of Pittsburgh, 2007

The concept of *reliability* has been defined traditionally by philosophers of science as a feature that an experiment has when it can be used to arrive at true descriptive or explanatory claims about phenomena. In contrast, philosophers of science typically take the concept of *validity* to correspond roughly to that of generalizability, which is defined as a feature that a descriptive or explanatory claim has when it is based on laboratory data but is applicable to phenomena beyond those effects under study in the laboratory. Philosophical accounts of experiment typically treat of the reliability of scientific experiment and the validity of descriptive or explanatory claims independently. On my account of experiment, however, these two issues are intimately linked. I show by appeal to case studies from the contemporary neurobiology of learning and memory that measures taken to guarantee the reliability of experiment often result in a decrease in the *validity* of those scientific claims that are made on the basis of such experiments and, furthermore, that strategies employed to increase validity often decrease reliability. Yet, since reliability and validity are both desirable goals of scientific experiments, and, on my account, competing aims, a tension ensues. I focus on two types of neurobiological experiments as case studies to illustrate this tension: (1) organism-

level learning experiments and (2) synaptic-level plasticity experiments. I argue that the express commitment to the reliability of experimental processes in neurobiology has resulted in the invalidity of mechanistic claims about learning and plasticity made on the basis of data obtained from such experiments. The positive component of the dissertation consists in specific proposals that I offer as guidelines for resolving this tension in the context of experimental design.

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

I would like to sincerely thank the members of my dissertation committee: Germán Barrionuevo, Rick Grush, Edouard Machery, Sandy Mitchell, Ken Schaffner and Floh Thiels for their guidance and support during the writing of this dissertation. I especially want to thank my Director and friend, Peter Machamer, for his unfailing enthusiasm for the project, his steadfast patience with me through the trials and tribulations of writing it up, and his consistent and strong support throughout the duration of my graduate career in HPS. I also want to give special thanks to my mentor in neuroscience and friend, Floh Thiels, who taught me how to be both an experimentalist and a theorist.

Integrative projects in the philosophy of science cannot be undertaken without someone believing in them and without financial support. I want to thank Edward Stricker, Alan Sved and all of the faculty in the Department of Neuroscience for providing me with wonderful academic support and training during my M.S. venture, without which this dissertation would not be possible. Special thanks go out to Germán Barrionuevo and Don DeFranco, for the support and guidance that they gave me as members of my M.S. committee in Neuroscience. I especially want to thank Carol Colby, Peter Strick, Jay McClelland and Dave Touretzsky for believing in the project, providing me with the opportunity to receive an even more well-rounded education in neuroscience and for the generous financial support that they provided to me through the Center for the Neural Basis of

Cognition (CNBC). A large portion of this financial support was part of a generous grant from the National Science Foundation (NSF) for the Integrative Graduate Education and Research Training Program (IGERT). So, thanks also to NSF.

At the end of the day you realize that there are so many people you need to thank for keeping you on the straight path through graduate school, browbeating you when necessary, and extending a kind word when things got stressful. So, in addition to the aforementioned people who also served these functions, I want to thank: Merrilee Salmon, Jim Lennox, Ted McGuire, John Norton, John Earman, Paolo Palmieri, Paul Griffiths, Allan Gotthelf and Jim Bogen. A special thank you also goes out to my friends Rita Levine and Joann McIntyre for all their encouragement and support along the way through grad school. And Joann, thank you always for making sure I was properly nourished.

My time in both HPS and Neuroscience brought me into contact with absolutely unique and just remarkable people–all of the aforementioned people, and my fellow HPS and Neuroscience cohorts who became close friends. In HPS: Uljana Feest, Julie Zahle, Erik Angner, Gualtiero Piccinini, Andrea Scarantino, Brian Hepburn and Balazs Gyenis. In Neuroscience: Beata Jarosiewicz, Nadia Chaudhri, Beatriz Kanterewicz, Guillermo Burgos, Joaquin Burgos, Jon Johnson, Isabel Johnson, and Luc Johnson. Thank you for fabulous dinner parties, for your support through the laughter and the tears of grad school, and for your continuing friendships.

Last, but not least, I want to thank my family: my mom, Kay, my dad, Jack, and my little brother, Mike for their love and support during these years that I spent in graduate school in my hometown of Pittsburgh.

1.0 INTRODUCTION

The dissertation is about scientific experimentation in general and the neurobiological study of learning in particular. The focus is on the relationship between the laboratory and the world. It is a truism that the phenomenal world is not orderly, but messy. The aim of the scientific laboratory is to impose order, to control nature in a fashion conducive to understanding it. But in this process, the phenomena that we desire to understand are simplified and artificially altered. The result is a lack of similarity between what is produced in the laboratory and what is encountered in the world and this puts constraints on the sorts of conclusions that can be extended from the one context to the other. And yet, if a phenomenon of interest is not studied in a controlled environment, but in circumstances that closely resemble the phenomenal world in which it was identified, the control requisite to draw meaningful conclusions about it is lacking. Experimentation in science involves just this kind of Catch-22: exertion of control in the laboratory can jeopardize the legitimacy of the extension of findings from the laboratory to the world; the absence of control may compromise the detection of accurate claims about the world and its contents. My first aim in this dissertation is to understand the specific nature of this Catch-22 and its unique manifestation in the contemporary neurobiology of learning and memory. The second aim is to develop strategies *within* the context of neurobiological experiments in order to cope with it.

1.1 FOUNDATIONS

The philosophy of experiment is a relative latecomer to the philosophy of science. Despite the importance of experimentation to science, compared to other aspects of science that have been a target of philosophical analysis (e.g., theory, explanation), scarce philosophical attention has been directed at understanding it. Yet, the increase in philosophical publications on experiment from Pierre Duhem's *The Aim and Structure of Physical Theory* (1954) to Ian Hacking's *Representing and Intervening* (1983) to Peter Galison's *How Experiments End* (1987) in combination with the work of Peter Achinstein (1983, 2000, 2001), James Bogen and James Woodward (1988, 1992), Nancy Cartwright (1983, 1999), Allan Franklin (1986, 1999, 2002, 2005), Rom Harré (1998), Jerrold Aroson, Rom Harré, Eileen Way (1995), Bruno Latour (1988), Deborah Mayo (1991, 1996, 2000) and James Woodward (1989, 2000) may be regarded as forming a foundation for the field. More recent philosophical work on experimentation has concentrated on specific areas of science (e.g., experimental biology, Bechtel 2005; Weber 2005) or specific experimental methods and techniques (e.g., functional imaging (Bogen 2001, 2002; Delehanty 2005). I take this dissertation to be making a contribution to this growing body of literature, with the target of my analysis being experimentation on learning in cellular and molecular neurobiology and the epistemological problems to which such experiments give rise.

In this dissertation, I conceive of experimentation as a process, which I refer to simply as the "experimental process". My account of this process is informed by Bogen and Woodward's (1988) and Woodward's (1989, 2000) distinction between *data* and *phenomena* and the related distinction that Woodward (1989, 2000) draws between two phases of experimentation: *data production* and *data interpretation*. Briefly, Bogen and Woodward (1988) claim that data are the output of scien-

tific experimentation. On their view, data are observable within the context of the laboratory and once gathered, data can be used to substantiate the existence of worldly phenomena (e.g., bubble-chamber photographs are data; the phenomena that such data were used to detect are weak neutral currents). They claim, on the other hand, that "phenomena are detected through the use of data, but in most cases are not observable in any interesting sense of that term" (Bogen and Woodward 1988, 306) (e.g., weak neutral currents are not directly observable). Since I am interested in the question of how an effect under study in a scientific laboratory relates to a phenomenon or set of phenomena in the world, I take there to be something out in the world over and above data that is the phenomenon of interest–even if the only way that we can get a handle on that phenomenon is by carving the world up in a particular way with our concepts and theories. On my account, there is something "in the world" or at least "outside the laboratory" that sets the experimental process in motion.[1]

Woodward identifies "two components in the overall process by which data are used to reach conclusions about phenomena" (Woodward 2000, S165): (1) *data production* and (2) *data interpretation*. On his account, *data production* "involves setting up systems of causal interaction or locating preexisting systems which allow for the production of effects which can be perceived and interpreted by investigators and which permit investigators to discriminate among competing phenomena claims" (Woodward 2000, S165). *Data interpretation*, on the other hand, "involves the use of arguments, analytic techniques, and patterns of reasoning which operate on the data so produced to reach conclusions about phenomena" (Woodward, 2000, S165). I take Woodward's

[1] In a sense, the data-phenomena distinction does capture the case of learning, if we can take data points to be observational "in the world". After all, when we identify learning, we do not perceive the learning directly, but use "data points". For example, in the case of learning to read, we point to Mikey at age 5 before he learned to read and Mikey at a later point in time, reading *Where the Wild Things Are* (1963), with no problem–to get at the phenomenon. Of course, Mikey might have memorized the story and not be reading, so we may require additional data points to determine if Mikey can read.

framework for understanding the "overall process by which data are used to reach conclusions about phenomena" (Woodward 2000, S165) as the foundation for my account of the experimental process.

In this dissertation I am interested in two types of experiments in the neurobiology of learning and memory: (1) *synapse-level plasticity* experiments and (2) *organism-level learning* experiments. The experimental process as I understand it is depicted in *Figure 1*. On this framework, the process is initiated when an empirical question about a phenomenon of interest is posed. The phenomenon may be a worldly phenomenon, like learning. On the other hand, it may be a phenomenon that is exclusively (as far as we know) produced in a laboratory, such as an electrophysiologically induced increase in the strength of a neural synapse (e.g., long-term potentiation or LTP described in Chapter 2). The posing of this question initiates what I refer to as the process of *data production* (designated as "1" in *Figure 1*). Data production involves two stages, which I refer to as *design* and *implementation*.[2] In the design stage, an experiment is designed and a protocol for running individual trials is developed. In the implementation stage, the protocol is implemented in the laboratory in an individual trial of the experiment. The aim of data production is to produce data that can be used to arrive at true claims about an effect produced in a laboratory. Using Woodward's terminology, I take the second stage of the experimental process to be *data interpretation*. On my account, data interpretation involves two steps (designated "2" and "3" in *Figure 1*): (2) the application of a claim arrived at as a result of a data production process to the effect produced in the laboratory (i.e., the claim is taken to be *true* of the effect) and (3) the application of a claim to the original phenomenon of interest in the world that prompted the experimental process in the first place (i.e., the claim is taken to be *true* of the original phenomenon of interest).

[2]For the sake of simplicity, these two stages are not designated in *Figure 1*, but they are both contained in the part of the process captured by the arrow designated "1".

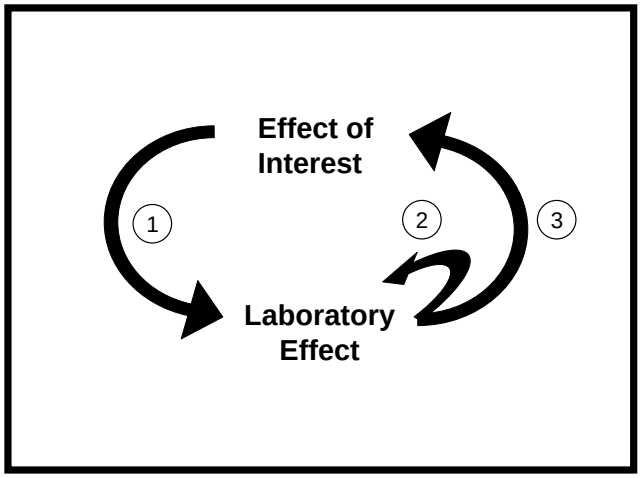


Figure 1: The Experimental Process.

1.2 RELIABILITY AND VALIDITY

In conjunction with this framework for understanding the experimental process, I provide non-traditional accounts of two aspects of this process, *reliability* and *validity*. *Reliability* has been characterized traditionally by philosophers of science as a feature that an experiment has when it can be used to produce evidence for knowledge claims (e.g., Mayo 1996, 2000, Woodward 2000). Philosophers of science have, however, provided different accounts of how experiments come to be *reliable*. For example, Mayo (1991, 1996, 2000) has claimed that "in order to determine what inferences are licensed by data it is necessary to take into account the error probabilities of the experimental procedure" (Mayo 1996, 442) used to produce that data. For Mayo (1991, 1996, 2000), *error statistical methods* used to reduce the error in an experimental procedure or arrangement are an important way in which to ensure the reliability of that experimental procedure or arrangement. In contrast, on Woodward's account of reliability, in order for "data to provide evidence for some phenomena-claim", "a systematic pattern of counterfactual dependence or sensitivity of both the data and the conclusions investigators reach on the phenomena themselves" must be present (Woodward 2000, S166). While Woodward admits that such patterns are rarely achieved in practice, scientific investigators "focus on a large number of highly specific local empirical facts about the causal characteristics of the detection or measurement process and investigate these by means of strategies [. . .] that need not involve explicit derivation" (Woodward 2000, S170). They use what they learn "about the kinds of mistakes that do or do not occur with the repeated use of a measurement procedure in various applications [. . .] to ground judgments of reliability" (Woodward, 2000, S168). In other words, while such repetitions of a procedure are not necessary or sufficient for reliability, the information gained from such repetitions can serve as a foundation for

grounding judgments about the reliability of a procedure.

While I agree with these general understandings of reliability, there are two primary differences between these accounts and the account that I provide in this dissertation. First, I ascribe reliability to a data production process as a whole, which I take to include the design of an experiment and the implementation of that experiment across individual trials in a laboratory. On my account, a complete data production process is reliable if it results in statistically analyzed data that can be used to discriminate one hypothesis from a set of competing hypotheses about an effect produced in the laboratory. Second, while I agree with philosophers of science like Woodward (2000) who maintain that the repetition of an experiment in the laboratory is important for grounding judgments about the reliability of that experiment, I emphasize the importance of such repetition in this dissertation, in a way not explicitly taken by other philosophers of science. I take such repetitions of an experiment to be important not only because they enable an understanding of the types of errors that may occur with respect to a given experiment and how to correct for such errors in the future, but also because they enable an investigator to achieve some degree of certainty that an effect produced in the laboratory is not an artifact but rather the direct result of a relevant experimental manipulation. Getting clear on what reliability is, how it functions in neurobiological experiments and why repetition is important for reliability in many if not all scientific experiments is the focus of Chapter 3.

A second desirable feature of the experimental process, which differs from *reliability*, is *validity*. On traditional scientific accounts, validity is taken to be a feature that is ascribed to experiments or tests (e.g., Campbell and Stanley 1963; Messick 1989). An experiment (or test), on such accounts, is taken to be *valid* if it supports the intended conclusion that is drawn from its results (e.g. Campbell and Stanley 1963; Messick 1989). Recently, an interest in the validity of

experiment has emerged in the philosophy of experiment. Specifically, Francesco Guala (2003, 2005) has sought to understand the tension that arises when an investigator desires to have the conclusion of an experimental result apply not only to the experimental situation in the laboratory but also to "circumstances of interest" (Guala 2003, 1198) outside the laboratory. Guala uses the distinction between *internal* and *external* validity (e.g., Cook and Campbell 1979) to comprehend the nature of this tension. On Guala's account, then, the *internal validity* of an experimental result is established when that result captures a causal relationship that *is* operative *in* the context of the laboratory. That experimental result is *externally valid*, then, when it captures a causal relationship that is operative in "a set of circumstances of interest", external to the laboratory setting.

My account of validity and data interpretation more generally differ from traditional accounts of validity in science of the kind Guala is interested in, in so far as I take validity to be a feature of interpretive claims rather than of experimental results. Experimental results, on my account, are statistically analyzed sets of data and interpretive claims are what arise when a hypothesis that has been discriminated by the data is applied to (a) the circumstances of the laboratory and (b) the circumstances of interest outside the laboratory. The explicit idea that application of the term "validity" should be restricted to interpretive claims is unique to the account of validity that I provide in this dissertation. However, I take it to be implicit in Guala's account in so far as he claims that validity assessments concern the "attribution" of a causal claim (that A causes B in an experiment E) to a circumstance inside or outside a laboratory (Guala 2003, 1198). What is novel in my dissertation is the application of this understanding of validity to experiments in neurobiology and the attempt to determine why interpretive claims arising from neurobiological experiments may lack validity as a result of specific features of those experiments. In addition, I use distinctions drawn between different types of validity in the psychological literature (e.g.,

Brunswick 1943; Cook and Campbell 1963; Cronbach and Meehl 1955; Messick 1989; Shadish, Cook and Campbell 2002) to provide a more comprehensive understanding of what validity is and what some of its dimensions are.

1.3 CONTEMPORARY NEUROBIOLOGY OF LEARNING AND MEMORY

With the aforementioned general framework of the experimental process serving as the foundation of this dissertation, the target of my analysis is the contemporary cellular and molecular neurobiology of learning and memory. The ultimate aim of neurobiological experiments on learning is, broadly, "to explain in molecular detail how specific memories are encoded and maintained within the human brain" (Barco, Bailey, Kandel 2006, 1525).[3] The process by which memories are encoded and maintained in the mammalian brain is *learning*. The terms "learning" and "memory" are often used interchangeably by neurobiologists who work on learning because their referents are identical–both point to changes in the behavior of either an organism (organism-level learning) or a synapse (synapse-level learning) as indicative of learning. Both are typically defined by

[3] Although I am abstracting this quote from a more recent text, the claim that the goal of neurobiology is to identify the cellular mechanisms of learning and memory can be found in the majority of published papers and textbooks in this field of research. For example, see Dudai ([1989] 1994), Sweatt (2003) Squire and Kandel ([1999] 2003) I selected this quote from Eric Kandel and colleagues (2006) for several reasons. First, Kandel has functioned as the primary spokesperson for the neurobiological study of learning and memory, identifying the goals of the field and the methodological approaches necessary to realize them. Furthermore, of all contemporary researchers in the field of learning and memory, he has had the most profound impact on shaping the field by producing students and postdocs who have gone on to work on the same issues in their own laboratories. As such, Kandel has many descendants who have contributed to shaping the field in the light of his influence. Additionally, it is these researchers that are called upon to review grants and referee papers in cellular and molecular learning and memory research, ultimately dictating which papers are actually published. This situation has to a large extent dictated how the goals of the field have been formulated across research labs in the Western world. For the aforementioned reasons, the field of cellular and molecular neurobiology is an interesting sociological phenomenon in its own right.

neurobiologists as "changes in behavior as a result of behavioral experience" (e.g., Dudai ([1989] 1994, 2003; Sweatt 2003).

In this dissertation, I focus on two types of neurobiological experiments: (1) *organism-level learning* experiments, in which a normal organism or a genetically modified organism is trained in a learning paradigm that is meant to bring about a specific type of learning, and after some time frame after learning, its brain is removed and the activity of specific molecules are investigated [4] and (2) *synaptic-level plasticity* experiments, in which an electrical stimulus is inserted into the brain of an organism and is used to artificially induce a phenomenon at the level of synapses in the brain that is taken to mimic and potentially be the mechanism of organism-level learning. The brain of the animal is then removed some time after application of the stimulus and the activity of specific molecules is investigated. While I will primarily be concerned with these *in vivo* experiments in this dissertation, some of the conclusions at which I arrive are applicable to *in vitro* experiments in neurobiology as well.[5] In Chapter 2, I introduce the basic assumptions that serve as a foundation for these two types of experiments and at later points in the dissertation (Chapters 3-5), I describe in detail how these experiments are designed and implemented.

1.4 OUTLINE OF DISSERTATION

In this dissertation my primary interest is to determine whether or not the aforementioned aims of learning and synaptic plasticity experiments that neurobiologists have identified for themselves are

[4] I define what I mean by learning paradigm in detail in Chapter 3. For now, please take it simply to mean, "a method for inducing learning and identifying it when it takes place."

[5] To differentiate, "*in vivo*" electrophysiological experiments are run in the intact anesthetized organism, while "*in vitro*" experiments are run in slices of brain tissue essentially in a glass dish.

actually being met and whether, given the current state of experimentation on learning they ever could be met. I show that the overarching commitment in neurobiology to ensuring what I refer to as "the reliability of the data production process" has resulted in a lack of attention being paid to establishing the validity of the interpretive claims about organism-level learning and synapse-level plasticity that neurobiologists make.

In Chapters 2.0, 3.0, and 4.0, I essentially track the experimental process as we encounter it in the neurobiology of learning and memory, from the positing of an empirical question about a phenomenon of interest to interpretive claims made on the basis of data. In Chapter 2.0, I begin with a general discussion of what learning is, what the phenomena of learning are, and how we come to detect learning in the world. I then consider a set of historical influences on neurobiology that have shaped the fundamental conceptual and theoretical presuppositions of the field. It is these assumptions that inform both the design and implementation of experiments.

Chapter 3.0 has two specific aims. First, since I take data production to be a process and maintain that reliability (as a goal) is a constraint that is operative over this process, I argue for a process reliabilist account of data production that I see as an improvement on current understandings of reliability in the philosophy of experiment because it adequately captures the role of the experimenter in both the design and implementations phases of data production. Appealing to Goldman's (1986, 2004) account of process reliabilism, I argue that reliability is best ascribed to processes; data production is one such process. On my account, a data production process is reliable just so long as it results in statistically analyzed data that discriminates one hypotheses as true from a set of competing hypotheses about an effect produced in a laboratory. If an investigator wants a data production process to be reliable, certain features taken to guarantee reliability will be built into the experimental design and its adjoining protocol, which essentially specifies a process type.

Then, during the implementation stage of data production, measures will be taken to ensure that each individual instantiation of the process type exemplifies those fundamental features. In this way, concerns about reliability will begin and end with the data production process and so will be primarily operative within the context of the laboratory.

In Chapter 4.0, I direct my attention at the issue of the validity of data interpretation. I investigate current work on validity in the philosophy of experiment and historical work on validity in psychology. This analysis yields the idea that validity has parameters and that the legitimacy of interpretative claims ought to be investigated along these parameters. I focus on those three parameters that I take to be most important for my purposes in the dissertation (1) *construct validity*, (2) *ecological validity* and (3) *external validity*. I argue that although evaluations of the validity of interpretive claims are made in the context of data interpretation, establishing the validity of interpretive claims ought to shape the design and implementation of experiments. I show that this is not the case with respect to learning and memory experiments in neurobiology.

In Chapter 5.0 of the dissertation, I consider the relationship between reliability and validity of the experimental process in the abstract and then with respect to the two neurobiological case studies under consideration, namely (1) synapse-level plasticity and (2) organism-level learning experiments. I show that reliability and validity coincide with conflicting prescriptions. Reliability does not concern the accuracy of the match between an effect of interest and that effect produced in the laboratory. Instead it prescribes simplifying measures in order to narrow the set of competing hypotheses with respect to a laboratory produced effect. As it operates to constrain the process of data production, it restricts the extension of interpretive claims to the laboratory. Validity, on the other hand, pulls in the opposite direction, by prescribing that an investigator build into an experimental design those dimensions of complexity that accompany an effect of interest in the world.

Yet, the consequence of such moves is a decrease in the simplicity of the effect produced in the laboratory and an expansion of the set of competing hypotheses that pertain to that effect–in essence, a decrease in reliability. I show that with respect to the case studies under consideration, neurobiologists are far more concerned with securing the reliability of the process of data production than with establishing the validity of their interpretive claims beyond the context of the laboratory. However, I argue that this commitment to reliability is in direct contrast to the explanatory goals of neurobiology as expressed by those investigators who are conducting such experiments. On the basis of how I understand the relationship between reliability and validity, while I understand this commitment as rational, I view it as untenable given the fact that neurobiologists are ultimately aiming to produce mechanistic explanations of effects beyond the laboratory.

In Chapter 6.0 of my dissertation, I provide a set of guidelines geared to function to maintain the reliability and increase the validity of those neurobiological experiments under consideration. The proposals that I make rely on methodological tools that have been proposed historically by experimental psychologists and more recently, by cognitive psychologists. I claim that given the fact that validity has different parameters, any solutions must proceed in a piecemeal fashion. I arrive at the conclusion that at best what we can achieve is some middle-ground in terms of guaranteeing simultaneously the reliability of data production and any given parameter of the validity of interpretation. This coincides with what a number of philosophers of science have claimed with respect to other laboratory and also non-laboratory sciences (e.g., Cartwright 1999; Hacking 1992). I conclude in Chapter 7.0 with a summary of the dissertation and a discussion of future directions for research.

2.0 LEARNING: INITIAL DEPARTURE FROM THE PHENOMENA

2.1 INTRODUCTION

All experimentation has some starting point. On the model that I outlined in Chapter 1 (*Figure 1*), I imply that it begins with the natural world. Yet, this is only part of the story. As several philosophers of experiment have indicated (Duhem 1954; Galison 1987; Harré 1998; Aronson, Harré, Way 1995), experimentation begins with the theoretical presuppositions of a given area of science. In the words of Peter Galison: "Theoretical presuppositions serve to carve out a piece of the phenomenal world as interesting; [. . .] without experimental presuppositions of some sort even the gross properties of matter would be in doubt" (Galison 1987, 4). Galison goes on to suggest that theoretical presuppositions are where experimentation not only begins, but where it ends. While I am sympathetic to Galison's point, some areas of science seek to study phenomena that have already been carved up and grouped by means of ordinary language definitions. And often in such cases the conclusions drawn in the laboratory are extended back not merely to the theoretically carved phenomena, but to pre-theoretically or folk theoretically carved phenomena. In particular, I think it is common in areas of science that study phenomena of our common-sense daily experience that come pre-grouped under an ordinary language definition, like *learning phenomena*, to take their results back to the phenomenal world that has been pre-theoretically

carved.[1] This is what I aim to capture in the framework for the experimental process that I have provided, and what I hope to show is the case in the contemporary neurobiology of learning and memory.

The primary aim of this chapter, then, is to consider the phenomena of learning broadly by itemizing a subset of the various kinds of phenomena that fall under the common-sense heading of "learning" and to see precisely how the pre-experimental definitions of learning in neurobiology compare. I then go on to consider some historical influences on the experimental study of learning that have shaped pre-experimental conceptual and theoretical presuppositions of experimentalists in neurobiology. I use this to demonstrate the immediate contrast between the conditions in the phenomenal world in which learning phenomena are encountered and what transpires conceptually even before an experimentalist in neurobiology sets foot in the laboratory. I suggest that well before experimentation begins, it is only a very restricted class of phenomena at which the question, "What are the cellular and molecular mechanisms of learning and memory?" is ever directed.

2.2 LEARNING: CARVING OUT THE PHENOMENA

The term "learning" is difficult to define, despite its widespread use. On the OED definition, *learning* is defined as "the acquisition of knowledge; a process which leads to the modification of behavior or the acquisition of new abilities or responses, and which is additional to natural development by growth or maturation." On this definition, learning is a process that an organism undergoes that by definition consists in changes in that organism's behavior.

[1] An additional issue, which I have yet to think about in detail, concerns whether we should take seriously a common-sense notion of learning, or differentiate folk psychological concepts of learning from scientific concepts and preference one to the other.

The OED definition can be taken as the ordinary language definition of organism-level learning or in the special case, human learning, in so far it is generally aligned with our everyday ascriptions of learning. We ascribe learning to a sentient organism as a whole–to ourselves, other persons and other types of organisms (e.g., dogs, cats, birds). We can identify when a very young child learns to speak, to walk, to imitate and to pretend–things that have prior to that point not been exhibited in that child's behavior. He may learn the fact that a truck has different properties than a train and to identify different kinds of trucks (street trucks, toy trucks, cartoon trucks) as part of the same class of things. We also take note when a child learns how to read and to ride a bike. We also encounter failures to learn, such as our Japanese friend's inability to differentiate phonetically "r" from "l" despite our repeated attempts to teach him the difference, or, our grandmother's inability to remember the date no matter how many times on a given day we tell her. As adults, we learn new things each day in reading the *New York Times* or listening to *NPR*. We may read about a new dish and learn how to prepare it.

In some way or other, each and every one of us relies every day on what we have learned, whether it be sending an e-mail, navigating from our home to our place of work, saying "hello" to a friend, taking an exam, or flossing before bedtime. Our abilities to undergo modifications that we exhibit in our behavior are essential to the development of a personal identity, successful interactions with other social beings and our overall survival as individuals and members of a species. Learning, at the level of the behaving organism, on the broad OED definition, is everywhere we look and seems to be happening all the time. Furthermore, it is not unique to humans. We may say that our dog Kip knows how to present his paw when someone says "paw" or that he can identify his owners from strangers.[2] If a flock of starlings finds food outside our apartment building at 8

[2]It may even be indicated by the wag of his tail (Quaranta et al., 2007)

a.m. on Wednesday, they most likely will come back at 8 a.m. on Thursday, and will continue to come back long after there is no more food left to give them, until eventually, one day they stop. That's learning (albeit slow). As the neuroscientist and psychiatrist Larry Squire (1987) claims:

> [...] it has always been recognized that a complete account of learning and memory must accommodate many different kinds of behavioral change. Memory includes not only the conditioned reflex, but also the ability to remember one face out of a thousand that have been seen, the ability to memorize a poem, the ability to demonstrate an improved throwing arm, and the ability to find one's way around an old neighborhood. (Squire 1987, 3)

With such everyday phenomena in mind, I want to consider how some neurobiologists have defined learning. Three neurobiologists in particular have been concerned to provide explicit definitions of learning. I want to see how these definitions compare to the ordinary language definition and whether or not they capture the same kinds of phenomena as seem to be captured by the ordinary language definition. Then, I want to go on to consider whether all of these phenomena that we group under the broad heading of learning share any fundamental similarities in common or not.

It is fairly common in cellular and molecular neurobiology to define learning and memory interchangeably. This is in part because the fact that a subject has learned and the fact that a subject has a memory are both detected by reference to some kind of change in behavior. For example, when I have learned that a student's name is Christan, that fact is exhibited in my behavior by my calling her by name. Yet it could also be said that I remember her name, i.e., that I have a memory of it. Learning and memory should be distinguished in theory, even if they cannot be distinguished in practice. Learning involves the "encoding" or "storage" of information. Remembering, however, involves not only encoding but also *recall*. For all intents and purposes, I will mean by learning, *encoding*.

In his ([1989] 1994) textbook, *The Neurobiology of Learning and Memory: Concepts, Findings and Trends*, the neurobiologist Yadin Dudai defines learning as:

> an experience-dependent generation of enduring internal representations, and/or an experience-dependent lasting modification in such representations. 'Enduring' and 'lasting' mean at least a few seconds, but in most cases much longer, and in some cases up to a lifetime. The effects of rigid developmental programmes, injury, disease, and drugs are excluded. (Dudai [1989] 1994, 6)

In comparison to the ordinary language definition, Dudai leaves open the possibility that learning can occur without an observed modification in behavior. In fact, he defines an "internal representation", as a "neuronally encoded structured version of the world which could potentially guide behavior" (Dudai [1989] 1994, 5). He differentiates his understanding of "representation" from the "mental representations" of cognitive psychology, because he thinks the class of internal representations includes everything spanning from simple representations, e.g., a color, texture, or a pain in a simple neuronal circuit in a limb, to complex representations, like grandmother, your concept of self, and Jennifer Aniston.

Dudai's notion of "internal representation" is clearly vague. To criticize his notion would put the subsequent discussion in a context that, for the purposes of this dissertation, I do not wish to go.[3] I merely want to point out the similarity between the class of phenomena captured by the ordinary language definition and the class of phenomena captured by Dudai's definition. Both definitions capture the whole list of kinds of phenomena that I cited above, despite the differences in the detail of the definitions. The primary differences between Dudai's pre-experimental definition of learning and the ordinary language definition (OED) are (1) he is committing himself to a

[3]The notions of "representation" and "information" are used freely by scientists but are left for the most part unanalyzed despite definitions being offered for them. Addressing this issue would put me in the context of a different project, one that has become of recent interest to philosophers of science. (e.g., Machamer and Bogen, Center for Philosophy of Science Talk, 2006)

neuronal mechanism account of learning–that nervous systems are required for learning, that the units of the system, the neurons, do the encoding of the internal representations; (2) some internal representations are not exhibited in behavior, and that (3) modifications in internal representations are the result of experience, and not of drugs, injury, or insult to the brain. At least the point about experience being necessary for learning is captured in the ordinary language definition in the clause "in addition to natural development by growth or maturation". However, it is enough to point out that these two definitions pertain to similar or overlapping classes of phenomena and that ultimately when claims are made by neurobiologists like Dudai about the mechanisms of learning, they pertain to this broad class of phenomena. Again, we can call this a *theoretical* definition, but it will differ from those operational definitions that we encounter in the laboratory.

In their book *Memory: From Mind to Molecules*[4] the neuropsychiatrist Larry Squire and the neurobiologist Eric Kandel ([1999] 2003) offer a definition of learning that is in the same spirit of Dudai's but far more general:

> We can acquire new knowledge about the world because the experiences we have modify our brain. And, once learned, we can hold the new knowledge in our memories, often for a very long time, because aspects of the modifications are retained in our brains. At a later time, we can act on the knowledge that is stored in memory, behaving and thinking in new ways. Memory is the process by which what is learned persists across time. In this sense, learning and memory are inextricably connected. (Squire and Kandel ([1999] 2003, 2)

Squire and Kandel add that "what we learn about the world is not built into our brains at birth but is acquired through experience and maintained through memory" ([1999] 2003, 2). So, in general, their definition contains that idea that learning is the result of experience modifying the brain. They do not define experience, but whatever they mean by it is more broad than what

[4]Although this textbook is not strictly a neurobiology textbook, it is a book that is written from the perspective of neurobiology and a reductionist biological view of learning, and it is representative of the account of learning and memory and their mechanisms that Kandel has been promoting since the 1960s.

Dudai means, and so, closer to the ordinary language definition. They leave open the possibility that experience can be construed broadly to include changes in the brain that result from drugs or injury, although they make no explicit commitment to this idea. What I am interested in primarily is pointing out that this definition and the ordinary language definition essentially pick out the same class of phenomena.

I want to consider a final definition of learning that we encounter in the neurobiology of learning and memory literature offered by the molecular neurobiologist David Sweatt. In his textbook, *Mechanisms of Memory* (2003) Sweatt defines learning as:

> the acquisition of an altered behavioral response, due to an environmental stimulus; in other words, learning is when an animal changes its behavior pattern in response to an experience. [. . .] Note that what is defined is a change in a behavior from a pre-existing baseline. Don't get confused that I am defining learning as a response to an environmental stimulus, but rather as an alteration in that response due to an environmental stimulus. An animal has a baseline response, experiences an environmental signal, and then has an altered response different from its previous response (Sweatt 2003, 3).

Sweatt's definition of learning is, as he admits explicitly, from the standpoint of an experimentalist who seeks to understand the mechanisms of learning. First, similar to the ordinary language definition, Sweatt defines learning as a change in behavior. So, the detection of learning necessarily requires some comparison of a previous behavior with a subsequent behavior or the comparison of the frequency of a previous behavior with the frequency of that same behavior at a later point in time. In so far as Sweatt's definition captures this idea, it does not deviate from the ordinary language definition. Furthermore, having an experience on both definitions is taken to require something other than natural developmental changes or maturation. However, Sweatt is trying to carve up the world as if it were a laboratory. The result is that there are obvious disconnects between his definition of learning and how ordinary language users identify learning in

the world. Sweatt must assume several things. First, he must assume that what he calls "baseline responses" are detectable in the natural world. That is, on his definition, it must be possible that we can isolate some point in time at which an animal has had no prior experience with an environmental stimulus and use that as the point of reference for detecting all changes in behavior.[5] Yet, this seems an impossibility unless we have control over an organism's experiential and behavioral history, which we rarely do. We typically encounter organisms after they have had at least a small subset of experiences. Furthermore, Sweatt seems to be assuming that environmental stimuli are discrete–that we can readily circumscribe and detect environmental stimuli and pinpoint in time and space the environmental stimulus to which an organism is responding. However, sometimes it is even difficult in the laboratory to itemize the exact stimulus to which an animal is modifying its behavior. (Cf. Gibson 1960) In addition, Sweatt's definition does seem to rule out more complex forms of learning that seem to require more than just the presentation of a stimulus, but that build on the prior experience of the organism, like learning how to read sentences after learning how to read words, or developing a concept of self that has some history. So unlike the other definitions, Sweatt's definition of learning is dissimilar from the ordinary language definition to the extent that it seems to rule out cases of things that we may want to call learning. But what's worse–it does not seem to capture anything that we actually encounter in the natural world. However, I do not think he would be happy with this conclusion, given the number of different learning types that he includes in his taxonomy (Figure 11), of which he thinks the mechanisms can be identified.

What I have sought to establish here are the pre-laboratory definitions of learning that are common in contemporary neurobiology. Clearly Sweatt's definition is closer to an "operationalized" definition than the others. Still, he is interested in at least a subset of the phenomena that are picked

[5]Or, it must be possible to identify the frequency of an organism's behavior in response to an environmental stimulus at various points in time in order to detect changes in the frequency of a behavior.

out by the ordinary language definition, and given his taxonomy it is a rather large subset. I want to move now past these general definitions of learning to identify historical influences on the field of neurobiology that have to a great extent shaped the concepts, methods and trends that drive experimentation in the field.

2.3 EXPERIMENTAL STUDY OF LEARNING: HISTORICAL INFLUENCES ON NEUROBIOLOGY

I am not in a position here to offer a comprehensive history of the cellular and molecular neurobiology of learning and memory. What I offer here is a subset of what I take to be the most important historical and more recent influences on experimental work in this field. I have abstracted the subset that I present here from various sources including: theoretical and review papers in behavioral physiology, contemporary neuroscience and neurobiology (sp., Kandel and Spencer 1968; Bliss, Collingridge and Morris (eds.) 2003), historical accounts from psychology and history of psychology textbooks and very partial historical introductions from neuroscience and neurobiology textbooks (Squire and Kandel [1999] 2003; Dudai [1989] 1994; Sweatt 2003). The four interconnected historical influences that I take to have shaped the predominant assumptions of the field that I will focus on are: (1) traditional animal learning theory (2) the Hebbian postulate (Hebb 1949) (3) Kandel and Spencer's (1968) early conclusions about the relationship between learning and activity at the level of cells and synapses relative to their work in the sea mollusk *Aplysia Californicum* and (4) Bliss and Lomo's (1973) discovery of a long-lasting increase in synaptic strength (long-term potentiation (LTP) documented in Bliss et al. 2003; Lomo 2003).

2.3.1 Animal Learning Theory

Late 19th and early 20th century advances in physiology and experimental psychology had a profound influence on shaping the fundamental conception of organism-level learning, the appropriate methods for studying it and hypotheses about the neural concomitants and neural mechanisms of learning. The three scientists who had the most predominant influence in one or more of these areas are: Ivan Pavlov (1849-1936), Edward Lee Thorndike (1874-1949) and Burrhus Frederick Skinner (1904-1990).

The Russian physiologist Ivan Pavlov (1906, 1927) provided psychology with a rigorous method for studying learning in a controlled environment, namely, the learning paradigm of classical conditioning.[6] In classical Pavlovian conditioning, the correlation between two stimuli, an unconditioned stimulus (UCS; e.g., food) and a conditioned stimulus (CS; e.g., tone) is related to a single behavior, or unconditioned response (UR; e.g., salivation). Pavlov began his experiments by identifying the baseline responses of animal subjects to both the UCS and the CS. A baseline is a measurement of an animal's performance prior to training in a classical conditioning experiment. For Pavlov, baseline is defined for an unconditioned stimulus (e.g., meat) as the animal's response (e.g., salivation) to that stimulus prior to training. Baseline for the neutral stimulus (e.g., a bell) is defined as the animal's response to that stimulus prior to training. On Pavlov's operational definition of learning[7], learning occurs when, after repeated pairings of neutral and unconditioned stimuli, the presentation of the neutral stimulus alone elicits a measurable response (e.g., drops of salivation) that approximates the measurable response initially elicited by presentation of the

[6] I define what I mean by "learning paradigm" in Chapter 3. For now I take it simply to be a method for inducing learning and detecting it when it occurs.

[7] I define "operational definition" in Chapter 2. Here, take it to mean a definition that specifies the behavioral response that must be elicited in order for learning to be ascribed to an organism.

unconditioned stimulus alone. Pavlov took this elicitation of the conditioned response (CR) to the CS to be indicative of the formation of a stimulus-stimulus (S-S) relationship between the CS and the UCS related to a single behavior, which he called *associative learning*. On the basis of his experiments, Pavlov concluded that ALL organism-level learning was associative.

Classical conditioning is perhaps the most widely used paradigm in the contemporary neurobiology of learning and memory. I want to consider a few examples. First, experimentation on the nictitating membrane (NM) reflex has been primarily undertaken in the laboratories of John Disterhoft and Richard Thompson since the late 1970's (Disterhoft et al. 1977, 1985; Thompson et al. 1976, 1984) . The nictitating membrane is a third, internal eyelid in rabbits that is drawn laterally across the cornea when a noxious stimulus is applied to the eye. Several different Pavlovian conditioning procedures have been developed to study the brain systems and neural pathways that mediate this reflex. In these experiments, the baseline responses of the NM muscles to presentation of the unconditioned (UCS; sp., airpuff) and neutral stimuli (CS, sp., a tone) are each measured. Then, the UCS and the CS are presented repeatedly and contiguously over numerous trials. Learning is operationally defined as the elicitation of an NM reflex (conditioned response) to the CS alone that is comparable in measurement to the reflex elicited by the UCS alone prior to training. Experimenters have investigated the effects of lesioning specific brain areas on the acquisition of this type of learning (Mauk and Thompson, 1987). In another set of experiments (Berger, et al. 1978, 1980), chronic electrodes were positioned in specific brain structures in order to record the activity of individual neurons before, during and after training. The goal of these experiments is to locate the synaptic pathways that "contain" the memory trace or "engram" of this type of associative organism-level learning. In general, such experiments have been considered successful in identifying the synaptic pathways and molecular mechanisms of learning.

Classical conditioning is also used to study contextual, cue, and cue and contextual fear conditioning in the rat. Such work was initially undertaken by Joseph LeDoux and colleagues (e.g., see LeDoux et al. 1984, 1986, 1988).[8] In these experiments, a rat is placed into an operant chamber (described below and in Chapter 3.0) with a speaker and an electric floor grid. A tone (CS) is played and then a shock (UCS) is delivered through the floor grid. The animals come to associate the tone with the shock, and the brains of animals are removed at varying time points after training in order to detect the activity of molecules in those brain areas that have been previously shown to be implicated in this form of learning.[9]

In addition to the aforementioned paradigms, investigations of learning in model organisms such as the fruit fly *Drosophila Melanogaster*, the sea mollusk *Aplysia*, and the worm, *Caenorhabditis elegans*, typically involve classical conditioning of simple reflexes or behaviors. A host of neurobiological laboratories are dedicated to studying learning in these organisms.[10]

The American psychologist Edward Lee Thorndike recognized the importance of contiguity and repetition for learning, but thought that behaviors could also be changed in light of their consequences or effects. He introduced the "law of effect" to capture this point, which states that responses elicited just prior to a satisfying state of affairs are more likely to be repeated (Thorndike 1898, 551-552). Thorndike's exact experimental methods have had less of a profound influence on the contemporary study of learning in neurobiology than has the *law of effect*, which is the cornerstone of contemporary neurobiological and psychological investigations into reward-based learning and addiction. Yet, the law of effect is as celebrated as it is because of the American

[8]These are citations of early work, but there is a host of later work with scientists such as Kareem Nader, Glen Schafe and Robert Malinow to name only a few.

[9]Specifically, contextual fear conditioning is thought to involve two limbic system structures: the amygdala and the hippocampus, whereas cued fear conditioning is thought to only require the amygdala.

[10] I have in mind here cellular and molecular neurobiologists such as Tim Tully and Aike Guo (*Drosophila*), Eric Kandel, Tom Carew, Daniel Glanzman, Jack Byrne (*Aplysia*) and Catherine Rankin (*C. Elegans*).

psychologist B.F. Skinner (1938, 1969).

Like Thorndike, Skinner believed that learning could involve either the formation of relationships between two stimuli and a response or the formation of reward-based relationships (S-R) between a stimulus (literally "an operant") and a response. In designing his Skinner box: an enclosed chamber with a lever, food hopper, floor grid and an apparatus to record lever presses, Skinner developed a learning paradigm that could be readily used for a variety of different kinds of conditioning experiments. Rats are by nature curious and when placed in the box will eventually elicit the behavior of pressing the lever. Rewards for pressing the lever can be varied, but typically a rat will receive a food pellet for each lever press.[11] In *operant conditioning* experiments, the baseline response of the animal is the rate at which it initially presses the lever. On Skinner's operational definition of learning, a measurable increase over time in the rate (frequency) of lever pressing above baseline is taken as indicative of learning.

Combined, the definitions of learning and paradigms for its study suggest that all learning is *associative* in nature. This idea of associativity has historically shaped how experimental psychologists and later behavioral physiologists conceived of learning and its mechanisms, as I will explain in the next section.

2.3.2 The Hebbian Synapse and Long-Lasting Potentiation

The conceptions of learning and experimental methods developed by Pavlov, Thorndike and Skinner are still used today, but when they were introduced into the history of psychology, they had a profound impact on conceptions of learning in behavioral physiology and shaped hypotheses concerning the nature of the neural mechanisms of learning and memory. First, Pavlov's experimental

[11]The rewards for pressing the lever can be varied (in contemporary neuroscientific research, rats may receive nicotine, alcohol and cocaine as a reward for lever pressing).

work on conditioned reflexes in combination with the anatomical findings of Santiago Ramón y Cajal (see Hebb 1949, 230) and physiological findings of Charles Sherrington (1941) and Lorente de Nó (1938), led Donald Hebb (1932, 1949) to the idea of "the Hebbian synapse" (see *Introduction* to the *Organization of Behavior*, Brown and Milner [1949] 2002).

Hebb maintained that because repetition and contiguity were necessary for classical conditioning of a reflex arc, such repetition and contiguity must have some impact at the level of the neural units that comprise the brain and must be sufficient for learning to take place. This inference led Hebb (1949) to his famous postulate:

> Let us assume then that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. The assumption can be precisely stated as follows: *When an axon of cell A is near enough to excite a cell B and repeatedly and persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.* (Hebb 1949, 62)

These steps itemized in this postulate with respect to a given neural synapse are represented in *Figure 2*. Hebb made an additional assumption about what type of anatomical change might mediate such electrical changes. He cited the work of Cajal who claimed that it must be the case that an anatomical change occurs at a synapse in the form of, for example, a synaptic knob on a dendrite that enhances the connectivity between the two cells. The importance of Hebb's postulate for the neurobiology of learning and memory was profound because it described a phenomenon and a potential way to induce that phenomenon artificially. While attempts were made after the introduction of Hebb's postulate to produce the phenomenon at a nerve synapse, it was not until 13 years later, in a brain structure thought to be implicated in learning, the mammalian hippocampus[12], that the neurophysiologists Bliss and Lomo (1973) artificially induced a phenomenon in an

[12]The mammalian hippocampus is a structure that has been broadly implicated in learning and memory, especially memories for places and events. It is a structure located in the medial temporal lobe in the mammalian brain. The phys-

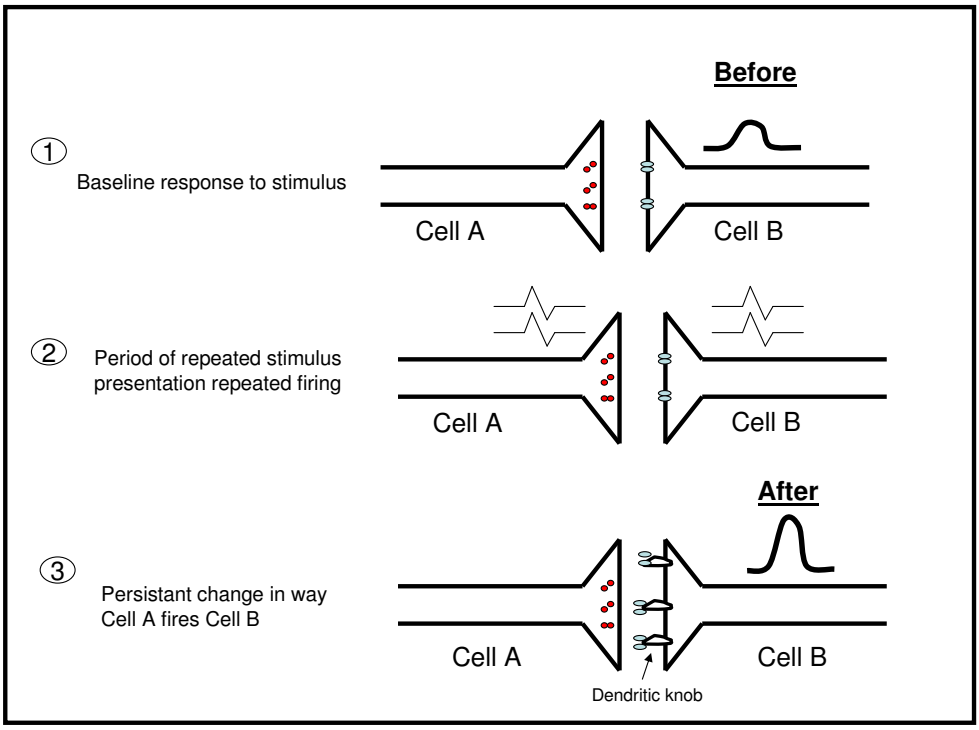


Figure 2: The Hebbian Postulate

anesthetized rabbit that instantiated the basic features of the Hebbian postulate in terms of how it was produced (contiguity and repetition) and how it was expressed (as a long-lasting change in synaptic strength–in the way that one set of cells fired those cells upon which they synapsed). Bliss and Lomo's finding, which they dubbed a "long-lasting potentiation of synaptic transmission" opened the door for a new field called "synaptic plasticity" that attracted physiologists, cellular and molecular biologists and biochemists (See Bliss, Collingridge and Morris 2003). Synaptic plasticity, in effect, became an important field of research in the cellular and molecular neurobiology of learning and memory that is viewed as complementing organism-level learning experiments. The aim of such work is to identify the cellular and molecular mechanisms involved in this, what is taken to be, "neural correlate" of learning and memory, with a basic assumption being that these mechanisms will turn out to be the same at both levels of organization (organism-level learning, and synapse-level plasticity).

2.3.3 Cellular and Neurophysiological Approaches to Learning

A second area in which the work of Pavlov, Thorndike and Skinner had a profound influence with respect to the development of the contemporary neurobiology of learning and memory is in the study of learning in model organisms with simple identifiable reflexes. These organisms can be used to study very simple forms of learning, and their simple nervous systems can be readily investigated using invasive techniques. The Nobel Prize winner Eric Kandel may be attributed as

---

iological pathways in this structure are well-demarcated, easy to find and their properties were to a significant extent detailed by Per Anderson in the late 1960's, paving the way for the structure to become a primary target of neurophysiological analysis. Furthermore, the famous patient H.M., who underwent complete removal (bilateral resection) of his hippocampi, exhibited memory deficits that Squire (2004) takes as having codified the distinction between declarative and non-declarative memory. I have left out a detailed discussion of the importance of the hippocampus here. But I refer the reader to my MS Thesis in Neuroscience, which is available on-line through the ETD website, http://etd.library.pitt.edu

the first behavioral physiologist to bring the techniques of classical and operant conditioning to bear on the issue of identifying the cellular and molecular mechanisms of learning and memory in model organisms. In a theoretical paper published in 1968, Kandel and Spencer explained how the learning paradigms of classical and operant conditioning could be translated to understand what was going on at the level of cells or synapses when an organism learns. This is represented in *Figure 3*. According to Kandel and Spencer, a natural stimulus in the environment of an organism could be interpreted as an electrical stimulus at the level of the cell; a muscle response or reflex could be interpreted as an effector response in a cell, which innervates the musculature. Learning experiments could be run in either intact or partially dissected animals; or in intact animals in which a piece of tissue was manually isolated. What is noteworthy about Kandel and Spencer's framework is that the learning paradigms taken to be operative at the level of the organism are the same as those taken to be operative at the level of the cell. In addition, events that happen at the level of cells and molecules can accompany a learning event (what Kandel and Spencer (1968) refer to as "cellular concomitants" of learning) and events that happen at the level of the cell and molecules may *causally* determine if learning occurs at all. What this picture in essence suggests, is that the traditional paradigms of classical and operant conditioning shaped not only the study of organism-level learning but also conceptions of learning at the level of the synapse. This claim preceded Bliss and Lomo's (1973) discovery, but it clearly coincided with a general consensus that had emerged in behavioral physiology and neurophysiology.

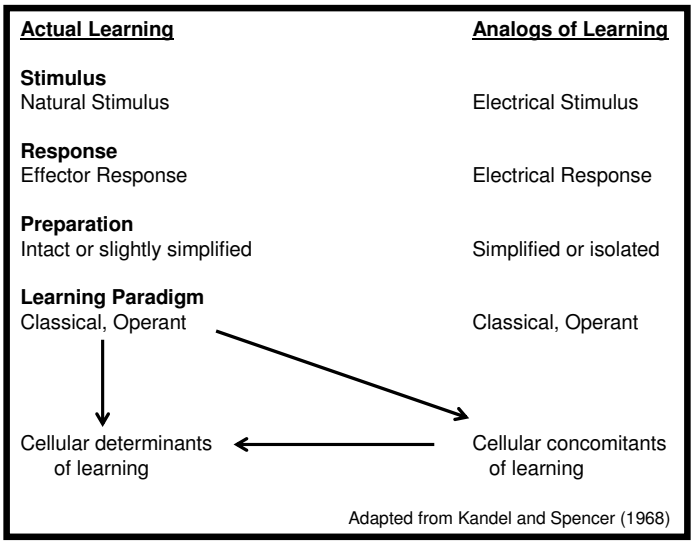


Figure 3: Translating Classical and Operant Conditioning at the Level of the Cell

2.4 CELLULAR AND MOLECULAR NEUROBIOLOGY TODAY: THE FUNDAMENTAL ASSUMPTIONS

Thus far I have described the historical influences on the contemporary neurobiology of learning and memory, but ultimately what I am after are the set of background assumptions, conceptual distinctions and theoretical presuppositions that an experimentalist accepts as true given his chosen field (e.g., see Kuhn 1962, Galison 1987, Hacking 1992) before he sets foot in a laboratory. I have already, in section 2.2 considered the conceptions of learning that initially inform experiments on learning in neurobiology, and in considering a set of historical influences I have identified another set of assumptions pertaining to the neural mechanisms of learning. I am interested here in several additional issues that neurobiologists have to take a stance on before they begin an experiment. In particular I am interested in how they answer the following set of questions: (1) What are the assumptions about the organism of interest? (2) What are the additional assumptions about the phenomena of interest? (3) What is the dominant model of the synapse?

2.4.1 Model of the Organism

The main organism of interest in contemporary neuroscience is the human organism. Knowledge of the structure of the primate nervous system and the stages of information processing from sensory inputs to behavioral outputs has been obtained from research on monkey subjects as well as via post-mortem anatomical studies on human subjects. Also relevant are functional imaging studies of human and monkey brains in the context of learning tasks. In addition, general knowledge of the mammalian brain has been obtained from invasive research on non-human animals, including such organisms as rodents, rabbits and cats. In this dissertation I am primarily interested

with a working input-output model of the mammalian organism which is represented in *Figure 4*. Yet, clearly the size and organization of the systems that this model contains and the flow of information from one component in the system to another will vary in degree of complexity from lower (rats) to higher-order (primates) mammals. The basic input-output model of the mammalian organism and the parts of the CNS taken to be relevant for experimental investigations of learning is represented in *Figure 4*.[13] With respect to any given organism-level learning event there will be 3 basic components with which an experimentalist in neuroscience may be concerned: (1) the properties of the stimulus or stimuli presented to an organism, (2) the variables internal to the organism that intervene between the presentation of a stimulus or stimuli and the organism's response; and (3) the organism's behavior in response to the stimulus or stimuli. Clearly each of these 3 basic components will be subject to different levels or layers of description, and a scientist will select that level/those levels of description for framing an experiment that he/she takes to be pertinent to his/her explanatory interests. Yet, in the context of learning experiments, one always begin and ends with the whole organism (where "ends" may be construed as one uses one's data to draw conclusions about an event pertaining to the whole organism).[14] So, I take it as necessary to begin this analysis with the "big picture" where experimentation on learning begins, and where the conclusions derived from such experimentation are ultimately ascribed.

With respect to each component of the model, specific kinds of simplifying assumptions are made. For example, the model is itself treated as a closed system. Each separately identified

[13]On this model, the initials are used to designate the following components of the model (from left to right of the figure): the primary sensory receptors: sensory (S), visual (V), auditory (A), taste (T), and olfactory (O) receptors; the thalamus (TH); the primary sensory, visual, auditory and taste and olfactory cortices (PSC, PVC, PAC, PTC, POC, respectively); the unimodal association area for a each given modality: sensory (SS), visual (VS), auditory (AS), taste (TS), and olfaction (OS); the basal ganglia and cerebellum (BSLGANG CEREBLLM); and motor neurons (MN).

[14]I discuss synaptic plasticity experiments in Section 2.4.3. In the context of synaptic plasticity experiments one begins and ends with the synapse, although such studies are interesting in so far as changes in plasticity at the level of brain synapses are models for what is assumed to occur during organism-level learning.

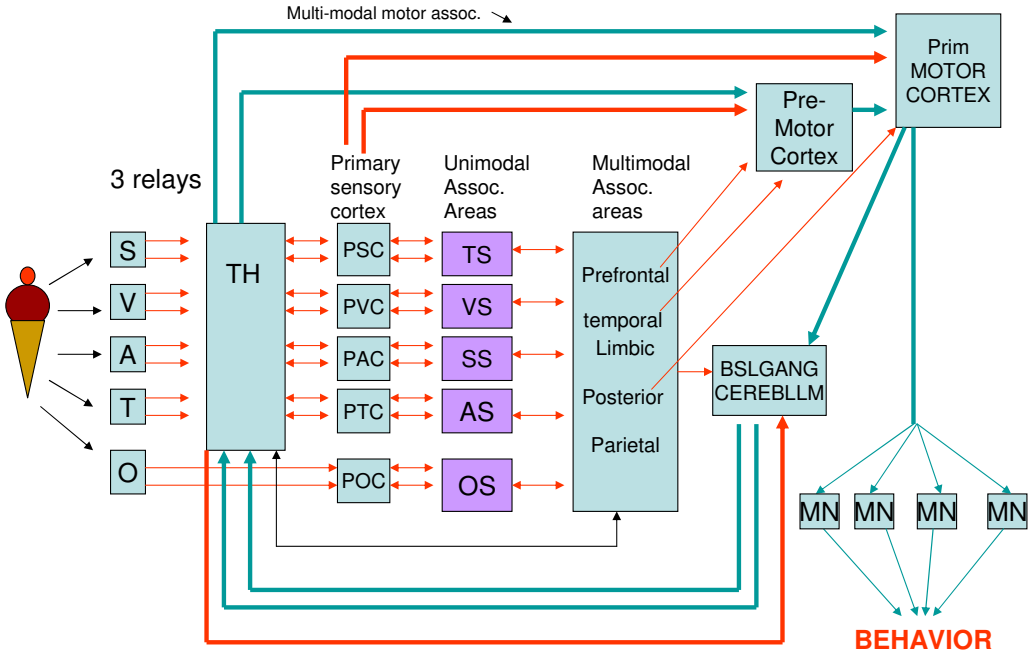


Figure 4: Input-Output Model of the Organism.

component or stage in the model is taken to be functionally separable from the others. Processes such as learning are assumed to have a definite beginning and a definite end in the context of experimentation. Each trial has a beginning and an end. Each block of trials has a beginning and an end. Each experiment has a beginning and an end. These beginnings and ends are fixed at the time at which the animal is placed in the experimental apparatus until the time that the experiment is concluded (sp., one trial of an experiment). This time frame is selected by the experimenter, but it neglects all processes that go on (1) prior to learning and (2) in between trials when the animal is in its home cage, when it is awake and behaving or resting.

In what follows, I will describe briefly the set of hypothetical events that occur from the time point at which a stimulus is presented to an organism, to the neural systems-level (sp., the level nearest to or just underneath the whole intact organism from the vantage point of the external world) events internal to the organism that intervene between the stimulus presentation and the organism's behavior with respect to that stimulus. I will parse the model in traditional neural systems-level terms, in which the axons of the neurons/cells that arise in and comprise one brain system are represented as synapsing directly onto the dendrites of those neurons that comprise another:

(1) Sensory information comes into the organism through sensory receptors. Stimuli and sensory experience have 4 basic attributes: (1) modality, (2) location, (3) intensity and (4) timing. Biological organisms have receptors that have the capacity to represent each type of information. The modality of a stimulus refers to "the type of energy transmitted by a stimulus and the type of receptors designed to detect that type of energy" (Kandel, Jessell and Schwartz (KJS) 2000, 413). Five classic sensory modalities have been identified: (a) touch, (b) taste, (c) smell, (d) vision and (e) hearing. In addition to these classic senses, pain, temperature, itch and proprioception (postur-

ing and body movement) as well as the vestibular sense of balance (position of the body within the gravitational field) are also detected by specific sensory receptors. For each class of sensory modalities there are specialized receptors that respond to different sets of stimulus energies within that modality (*Table 21-1*, KJS 2000, 414). These are typically referred to as submodalities. For example, a stimulus can be blue or red (color), stationary or moving (motion), hot or cold (temperature), bitter or sweet (taste). (KJS 2000, 414) The location of a stimulus refers to the position of that stimulus relative to space. Stimulus location is thought to be represented by the distribution of activity across a set of sensory receptors that detect when a stimulus is present. For example, sensory receptors in the somatosensory and visual systems have a topography that allows for the detection of stimulus size and position. Stimulus intensity is represented by a change in receptor amplitude, which reflects the amount of stimulus energy that has impacted on that receptor. Stimulus timing is defined as the time period between the initial firing of a receptor and the termination of its firing. Both the timing and intensity of a stimulus are represented by the firing patterns of a given receptor.

(2) Sensory receptors translate stimulus energy into electrical energy. Each sensory receptor is sensitive to 1 type of physical energy. Sensory receptors may be of 4 different types: (1) *mechanoreceptors* in the somatosensory and auditory systems (2) *chemoreceptors*, which mediate pain, itch, taste and smell, (3) *thermoreceptors* which mediate body temperature, temperature of air surrounding the organism and the temperature of things that we touch and (4) *photoreceptors* in the retina that are sensitive to and mediate electromagnetic energy. According to Kandel, Jessell and Schwartz "The sensory receptor is the first cell in each sensory pathway and transforms stimulus energy into electrical energy" (KJS 2000, 414). This electrical signal is called the receptor potential and the process by which the energy is converted into an electrical signal is referred to as

stimulus transduction.

(3) Electrical energy is conveyed from receptors to neurons that send projections to different nuclei of the thalamus. This is true for all sensory modalities except olfactory receptors that send information directly to cortex.

(4) Neurons that comprise thalamic nuclei process and filter sensory input before sending it to primary sensory cortex.

(5) Sensory input is conveyed from thalamic nuclei to neurons in primary sensory cortex. Sensory input that comes into thalamic nuclei is filtered and projected to primary sensory cortex. Input from each modality is sent to an area proximal to other input coming in from receptors responsible for that modality. Additional sensory input is sent directly to the cerebellum.

(6) Information from primary sensory cortex is conveyed to neurons in the unimodal sensory association cortex. The unimodal sensory association cortex is the area in which inputs from each modality are combined and represented with all inputs arising from that modality (e.g., with respect to vision: depth, color, motion, orientation). Information from primary association cortex is also sent to the premotor and primary motor cortices.

(7) Sensory information is conveyed from neurons in unimodal sensory association cortex to neurons in multimodal sensory association cortex. Neurons in unimodal sensory association cortex convey information to neurons in multimodal sensory association cortex where information from all modalities is combined. The multimodal sensory association areas include: (1) prefrontal cortex, (2) parietal cortex, (3) temporal cortex (limbic system (including amygdala and hippocampus)) and (4) posterior cortex. In addition, in these areas, current sensory information is thought to be integrated with stored information. More specifically, these association areas are said to have the following types of functions: (1) interpretation of sensory information; (2) association of percep-

tions with previous experience; (3) focusing of attention, and (4) exploration of the environment. [or "the production of a unified precept and the representation of the percept in memory" (KJS 2000, 345).]

(8) Sensory information is conveyed from neurons in multimodal sensory association cortex to neurons in multimodal motor association cortex. Neurons in multimodal sensory association cortex project to areas in the pre-motor and supplementary motor cortices, and the primary motor cortex. They also send projections to the basal ganglia and cerebellum.

(9) Neurons in multimodal motor association cortex project to neurons in unimodal motor association cortex. Information from pre-motor and supplementary motor cortices is conveyed directly to the primary motor cortex. Information from the basal ganglia and cerebellum are conveyed via some thalamic nuclei to the premotor cortex and via other thalamic nuclei to the primary motor cortex.

(10) Neurons in primary motor cortex synapse directly onto and activate motor neurons, which initiates movement and results in behavior. Information is carried from neurons in primary motor cortex via axons that comprise different corticospinal tracts descending from the cortex and cerebellum.

This input-output model of the organism as I have presented it is highly-simplified and idealized when we consider how it compares to an organism behaving in the world. First, the model assumes that from inputs to outputs this is a closed system – that somehow it is possible to take a behaving organism in the world, carve out the stimuli to which the organism is responding, identify the intermediate neural processes and the response, and investigate the causal relationships between them. Assuming that this is possible, it is clearly a more complex venture than this model suggests. And it leads to several problems. First, making such temporal slices with respect to an

organism in the world obscures that organism's history and all of the causal events leading up to the current situation of interest. Second, attentional processes do not feature prominently in this model. Furthermore, the flow of "information" from stimulus to response is generally considered to be a passive flow operating below the level of conscious awareness.[15] Additionally, experiments on learning in neurobiology typically begin with a specific type of learning that is assumed to involve a specific brain structure. On such a reduced model, only one structure of interest out of the whole set of structures itemized in *Figure 4* is taken to be relevant to the form of learning under study. In terminal learning experiments, the brain is removed and molecular activity in that area is studied. Yet this abstracts away from everything else that is going on in that organism's brain, so whatever the molecular story arrived at, it must be incomplete.

2.4.2 Taxonomy of Learning Types

One taxonomy of learning and memory types that is widely accepted throughout contemporary neuroscience was put forward by Larry Squire (2004) and is represented in *Figure 5*. This taxonomy is the culmination of experiments on learning from the 1950's to date. On this taxonomy, all learning phenomena that exist in the world are thought to be of one of two primary forms: (1) *declarative* and (2) *non-declarative*. *Declarative memory* is defined as

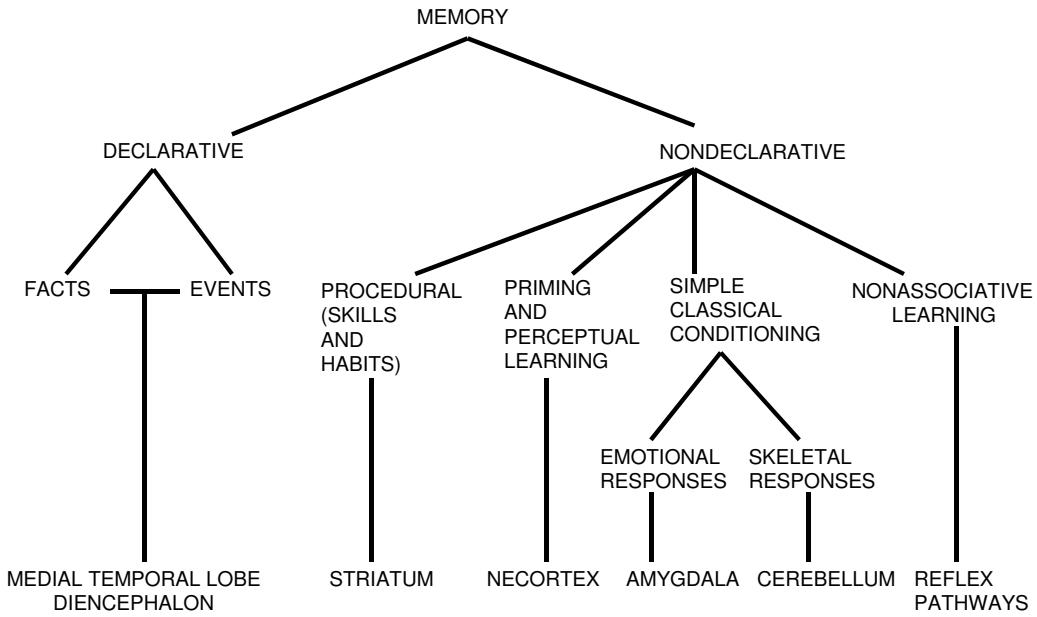
> the memory for events, facts, words, faces, music–all various pieces of knowledge that we have acquired during a lifetime of experience and learning; knowledge that can potentially be declared that is brought to mind as a verbal proposition or as a mental image. (Squire and Kandel [1999] 2003, 70-71)

[15] One might argue that this is only taken to be true in simple forms of learning or in simple organisms. However, if that is true, then it should cast some doubt on the idea that the cellular and molecular mechanisms of declarative and non-declarative memory will be the same, a claim made Kandel and Squire (2003).

While declarative memory is taken to involve items and events that can be consciously recalled, no emphasis is placed on whether so-called declarative learning, the encoding of memory itself, compared to the recall of encoded information, is conscious or not. It seems likely that it could be either, which is problematic for the strict division between declarative and procedural learning on the basis of awareness.[16]

Another problem with Squire's taxonomy is the fact that simple forms of classical conditioning are taken to involve two brain structures in the limbic system–the amygdala (contained in the limbic system) and the cerebellum, but not the medial temporal lobe, which includes the hippocampus. However, contextual-fear conditioning paradigms, which are a form of simple conditioning, have been shown to involve the hippocampus (e.g., Atkins et al. 1998). So the question is whether this form of learning is "declarative" or not. The declarative memory category remains divided between facts and events without other subdivisions, in part because little work can be done on declarative memory in non-human mammals, since they lack a language, and few paradigms that test declarative learning have been developed for monkeys and rats. In contrast to declarative memory, *non-declarative* memory is taken to be "revealed through unconscious performance", "inaccessible to conscious awareness", and "inflexible". Based on the types of learning itemized under "non-declarative" memory, it is a form of learning that directly involves an organism's sensory-motor system, but not "higher-order" structures such as the diencephalon. The basic idea seems to be that simple stimulus-response processes that are part of a reflex circuit are where the learning takes place–and these events take place without conscious awareness. However, if we could itemize all of the procedures in our lives that we learn, in particular, the interesting and complex ones such as

[16]This is a feature of Squire's taxonomy with which Sweatt (2003) takes issue. His solution is to provide a new taxonomy that differentiates between the categories of conscious and unconscious learning and working memory. This taxonomy is represented in *Figure 10*. I criticize Sweatt's taxonomy in Chapter 6.



Mechanisms conserved across declarative and non-declarative memory types.

Figure 5: Squire's Taxonomy adapted from Squire (2004).

riding a bike or learning how to read or to play the piano, conscious attentional processes must be present.[17] Also, that such procedures involve changes in the cortex is evidenced by the fact that various musicians have learning related changes in their diencephalon, in response to studying an instrument for years (cf. Pantev et al. 2003).

Another reason why the division between declarative and non-declarative memory is an over-simplification: it is well-documented that some forms of learning involve what would be classified as a mixture of both declarative and procedural learning. For example, in reward-based learning, which would be classified in this scheme as a form of "procedural" learning, it is well known that animals come to associate cues (e.g., a light coming on–"an event") with an action (e.g., lever-pressing) and a "reward". Yet, the drugs being taken are in the process having an impact on the performance of the procedure of lever pressing in a way not conscious to the animal. In fact, one would imagine that there are many cases in which the distinction between the two forms of memory collapses. [For a list of the experiments that have led to the conceptual distinctions drawn in the taxonomy, please see *Appendix A*.]

I want to turn away now from the level of the organism and the forms of learning of which it is capable, to address the unit of the organism that is taken to be fundamental for learning: *the synapse*.

2.4.3 Model of the synapse and learning-related synaptic change

A widespread assumption in neurobiology is that organism-level learning involves an activity dependent change in brain synapses. Such changes are thought to be mediated by the activation of molecular signaling cascades when an organism learns that cause structural changes in brain

[17] And this is why our piano instructors reprimand us when we do not seem to be paying attention while playing!

synapses. These structural changes are thought to underlie learning.

The dominant model of the synapse in learning is presented in Figure 6. The general assumption is that during a learning event a change takes place in the way in which a set of cells (A1. . .An) excites that set of cells to which it projects (B1. . .Bn). Hebb talked about individual cells with respect to his postulate, because he was primarily thinking of a simple reflex circuit, having a sensory neuron and a motor neuron. He took other "higher-order" cognitive forms of learning to require the activity of "assemblies" of cells–circuits of neurons connected by synapses.

In the learning and plasticity experiments that are the focus of this dissertation, the unit of analysis is one group of pre- and post-synaptic neurons that comprise sets of synapses (e.g., the Schaffer commissural fiber pathway in the hippocampus). Their analyses of synapses tend to stop at investigations of learning-induced or artificially-induced molecular changes in post-synaptic neurons. There are practical reasons for this stopping point, of course, but as a result, certain problems emerge. Specifically neurobiologists treat synapses as closed systems in much the same way that they treat of organisms as closed systems. While in learning experiments they may investigate the behavior of an organism that is taken to be downstream of molecular changes, they rarely look beyond the post-synaptic neurons in the synapses of interest to investigate the immediate downstream consequences for those neural circuits intermediary between the synapses of interest and the behavioral response. In plasticity experiments, investigations of neurons downstream of the synapses in which a plastic change has been effected are also not typically investigated. In addition, neurobiologists tend to treat individual synapses in the brains of organisms as "naive". In other words, they tend to think of a synapse as not having undergone a plastic change prior to the plastic change induced by the investigator–they treat the synapse as if it has no history of being

modified.[18]

The dominant contemporary view as to how synapses are changed does not differ significantly from Hebb's postulate, and has the same features as those described for Bliss and Lomo's long-lasting potentiation. One exception is that changes in synaptic strength can include either increases or decreases, long-term potentiation (LTP) and long-term depression (LTD), respectively. These two kinds of synaptic changes are represented in *Figure 6*. Both types of changes are taken to involve the repetitive and contiguous presentation of stimuli at the level of an organism, and direct electrical stimulation at the level of the cell.

2.5 CONCLUSION

In this Chapter I have sought to itemize the foundational assumptions in the neurobiology of learning and memory that are accepted by experimentalists who work in this field before they begin their experimental research. As I will show in the chapters to follow, such assumptions shape a variety of decisions that are made in the context of the laboratory. Such assumptions may also be viewed as ultimately impeding the ease of the move from the laboratory to the world. Changing such assumptions would require a radical ("paradigm" in the Kuhnian sense of a set of background assumptions, techniques, research standards) shift in the foundations of contemporary neuroscience. In light of the unlikelihood of such a shift taking place, in the chapters to follow, I attack the problem by means of a less direct route. Yet, given the theoretical presuppositions I identified in this chapter, we can predict that the order imposed in the laboratory will be of a particular kind–it will

[18] Among other ideas presented in this dissertation, I am indebted to Floh Thiels for candid discussion on this issue with respect to plasticity research.

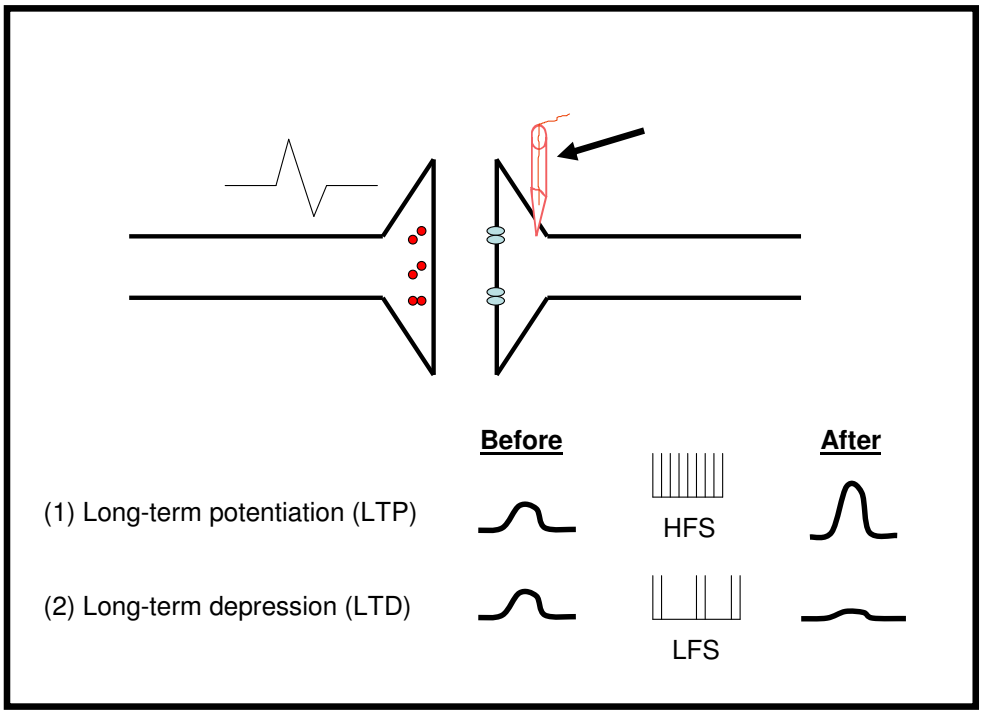


Figure 6: Two Forms of Synaptic Plasticity.

be meant, in part, to uphold the assumptions in which the science is rooted.

3.0 RELIABILITY OF EXPERIMENT IN THE NEUROBIOLOGY OF LEARNING AND MEMORY

3.1 INTRODUCTION

Reliability has traditionally been characterized by philosophers of science as a feature that an experiment has when it can be used to yield data that can support knowledge claims. To date, understandings of the reliability of experiment in the philosophy of science have been based primarily on analyses of experiments undertaken in the physical and chemical sciences and neuropsychology (e.g., Bogen and Woodward 1988; Hacking 1983, 1992; Mayo 1996, 2000; Woodward 2000). I am interested in whether such accounts can capture the nature of reliability in those laboratory sciences that have as their goal the discovery of the mechanisms of complex biological and psychological phenomena. In these areas of science, experimentalists play a crucial role in both establishing and maintaining the reliability of their experiments.

Experimentation in any branch of science begins with an empirical question or set of empirical questions about a phenomenon of interest or observed effect. In many domains of science, such as neurobiology, these empirical questions are questions about mechanisms. Such mechanistic questions concern what entities and activities are involved in the production of a phenomenon of interest (Bechtel and Richardson 1993; Machamer, Darden and Craver 2000). To answer mecha-

nistic questions, neurobiologists must attempt to isolate a system in a laboratory, intervene in that system in order to produce effects and investigate the consequences of such interventions on these effects (cf. Woodward 2003).

Data are the recorded measurements of scientific experimentation that are produced when a system is isolated in a laboratory and used to produce effects. When the effects produced in the laboratory are measurable and repeatable, they comprise data. In the neurobiology of learning and memory, data production is an intricate process often involving multiple experimental techniques and methods that have multiple discrete steps.

Some philosophers of science (Achinstein 1983, 2000, 2001; Bogen and Woodward 1988, 1992; Mayo 1996, 2000; Roush 2005; Woodward 2000; Kelly 1996; Wimsatt 1981) have suggested that the most important role for data is to function as evidence to support knowledge claims. These same philosophers have indicated that in order for data to be sufficient for this purpose it must be *reliable*, where "reliability" is attributed variously to the procedures, methods, techniques, devices and equipment used to produce, detect and analyze data. On such accounts, reliability is considered to be an *ideal* that may or may not be achieved in practice. A number of philosophical accounts have been offered that seek to identify what makes these various aspects of scientific experiments reliable. However, many of these accounts have not treated in-depth of a primary way in which reliability is achieved in sciences other than physics and chemistry; namely, by means of multiple repetitions of the "same" experiment. In this chapter, I address the issue of *repeatability* and its relationship to reliability in a way not previously addressed in the philosophical literature.

My primary aim in this chapter is to develop a basic account of reliability that can be used to capture its role in experimentation in those areas of neurobiology directed at the study of the cellular and molecular mechanisms of learning and memory. I take the development of such an

account to require a basic set of conceptual tools to be found in epistemology and the philosophy of science. I motivate my account in Section 3.2.1, with a critical analysis of Alvin Goldman's *process reliabilist* account of cognitive processes (Goldman 1979, 1986, 2004). I *tentatively* argue that Goldman's process reliabilism can be used to provide a detailed framework for understanding the precise role of reliability in neurobiological experiments. In Section 3.3., I provide a detailed analysis of a case study of experimentation from the neurobiology of learning and memory to highlight how concerns about reliability constrain the study of learning and its mechanisms. I go on to synthesize those conclusions that I reach in section 3.2. with the findings that I extrapolate from my analysis of the case study to develop my account of the reliability of experiment which illuminates the relationship between reliability and repeatability, capturing an important way in which reliability is achieved in neurobiological experiments.

3.2 RELIABILITY

On the OED definition, the term "reliable" is equated with words such as *trustworthy*, *safe* or *sure*. In statistics, "reliable" means "repeatable". When we speak about something being *reliable* we usually mean, intuitively, that it is reliable for some end; that it performs some function or can be counted on to repeatedly serve a specific purpose. Typically, we have to have had more than one encounter with a process to call it reliable. For example, a car is reliable in so far as it gets us from point A to point B; not just once, but *regularly*. However, though repetition in some cases may appear to be a requirement for reliability ascriptions, there is no fixed rule about how many times an event has to occur in order for it to count as *reliable*.

I consider an object or process *reliable* just so long as it performs some function or behaves

the same way repeatedly or at least can be expected to function or behave that way in the future (e.g., in cases in which the function or behavior has only occurred or been observed one time). While this is a vague criterion for reliability ascriptions, it does appear to be an acceptable foundation for accounts of reliability on offer in epistemology (e.g., Goldman 1986), philosophy of science (e.g., Woodward 2000) and science itself (e.g., Krauth 2000). So, I will also use it as my starting point here. I am interested in particular in the *reliability* of those data or evidence-producing processes that we encounter in experimental science. Experimentation is the most celebrated evidence-producing process. It has a definite beginning in experimental design, it proceeds in a set of discrete stages as this design is implemented to collect data in the laboratory and it has a projected end or result, namely, to produce evidence for knowledge claims.[1] According to the basic definition that underlies my account, *an evidence-producing process is **reliable** just so long as it results in data that can be used to discriminate one hypothesis from a set of competing hypotheses about an effect produced in a laboratory*. In order to develop a systematic account of what it means for an evidence-producing process in science to be reliable in this way and how such reliability may be attained, in what follows I investigate accounts of reliability on offer in those areas of philosophy in which concerns about the reliability of such processes have emerged: *epistemology* and the *philosophy of science*.

3.2.1 Reliability and Epistemology

In epistemology, reliabilism is a primary form of externalism. Externalist theories of knowledge are based on the assumption that the justification of a belief state may be determined by facts to

[1]For example, consider the OED definition of "process": (1) A continuous and regular action or succession of actions, taking place or carried on in a definite manner, and leading to the accomplishment of some result; a continuous operation or series of operations. (2) An artificial or voluntary operation; a course or method of operation; a systematic series of actions, physical or mental, directed to some end.

which the subject who holds the belief may or may not have access. Internalists hold the opposite view, namely that belief justification depends on the state of mind of a belief holder or on information to which he alone has reflective access. Externalist theories are more attractive for capturing the epistemic practices of scientists than internalist accounts. The nature of belief justification in science is *public* rather than private, making science a *large-scale community-based* effort to understand the world (e.g., see Galison 1987). An experimental process in science is deemed reliable for producing evidence for knowledge claims only when it has passed a public test of reliability. While individual scientists try to ensure the reliability of those processes they use to produce data, such processes are nonetheless scrutinized by members of the scientific community. For example, manuscripts accepted for publication in peer-reviewed journals are only those that contain descriptions of data production processes that supposedly have passed a test of reliability. In this way, justification for knowledge claims in science will never depend *exclusively* on the internal states of any one individual agent reflecting on the reliability of the processes by which such claims were produced. Rather, it will also depend to a significant extent upon favorable peer review.[2]

In light of the suitability of externalist accounts for capturing the public nature of the epistemic practices of scientists, I take an examination of such accounts to be relevant to my aims here. Of the available externalist accounts of knowledge in epistemology, including *causal*, *reliabilist* and *truth-tracking* theories, Alvin Goldman's (1979, 1986, 2004) *process reliabilism* is the only account that emphasizes how epistemology should concern itself with both the *local* and *global* reliability of the processes by which beliefs are produced. If a belief producing process is *globally*

[2]It should be added, however, that contrary to what some externalists claim, some processes that lead ultimately to the production of evidence upon which beliefs or knowledge claims are based, such as producing data, *are* refined and controlled by individual scientists in order to ensure the reliability of those processes. This suggests that perhaps there are circumstances in which internalist and externalist positions can find a common grounding when it comes to understanding the belief-producing practices of scientists. Sherri Roush (2005) makes a similar suggestion.

reliable, it is reliable across many if not all uses of that process and thus, in many situations. If, on the other hand, a belief producing process is only *locally reliable*, then it is reliable only in that local context in which it has produced a belief. In science, evidence producing processes must be both locally (i.e., in one lab) and globally (i.e., across labs) reliable. Many externalist accounts emphasize either the local or global reliability of cognitive processes. As Goldman is one of the only externalists concerned with both types of reliability, I want to consider his account here.

Although Goldman has refined his account since its inception (Goldman 1979) in response to his critics, the basic upshot has always been the same, namely, that a belief producing process (in any intellectual domain) is reliable if and only if it tends to produce a higher ratio of true compared to false beliefs and moves toward the elimination of error (Goldman 1986). Ascriptions of reliability to a belief-producing process require multiple trials of that process in order to establish a *probability* or *likelihood* that the process will produce more true than false beliefs. I want to consider the original probabilistic formulation of Goldman's process reliabilism (Goldman 1979) and its problems before addressing the newer propensity account that he has introduced to address such problems (Goldman 1986, 2004). As I will show, there are conceptual features of both versions of the theory that can be used to understand the reliability of the experimental process.

On Goldman's account, cognitive processes come in different types and individual instances of these types are called tokens. When we assess the reliability of a *type* of belief producing process, we must base our assessments on a statistical analysis of all of the individual *tokens* or outputs of that process. So, ascriptions of reliability to a process type will be contingent on the ability to consistently identify tokens of that type and to calculate the ratio of true compared to false beliefs produced across these tokens. However, it is unlikely that each individual token will be identical with respect to the complete set of features ascribed to a process type. So, in order to calculate

the reliability of a given process type we need to specify a *range* of features that individual tokens must exhibit in order to qualify as viable instances of that type. However, we may disagree about which description of a process type to use, how general that description should be, or what the relevant range of features of a process type actually are. In fact, two problems arise that pertain to how narrowly or how widely we cast our descriptive nets with respect to process types.

The first problem is referred to as the generality problem (Goldman 1986, 49). Particularly when it comes to cognitive processes like perception, the reliability of a process under one description may differ from its reliability under another description. To take an example, suppose that I am near my window and I come to form the belief that I hear a cardinal singing in the tree just outside based on an auditory process alone. If we want to consider the reliability of the process by which this belief is produced in me, we need to describe that process. We have various options at our disposal for describing it. For example, consider two available options: (1) process of hearing a cardinal singing in a tree, (2) process of hearing a red bird singing. Now, suppose that I did hear a bird, but it was a bluejay mimicking cardinal song. Under description (1) we may be tempted to say that I formed a false belief based on an auditory process that was otherwise reliable–since the auditory process did convey to me correctly cardinal song, although audition alone is not reliable when it comes to discriminating a true song from a mimicked song. However, on description (2) the process is *not* reliable because even given that what I heard was cardinal song and cardinals are red, the song is being elicited by a blue and white colored bird. So, on the second description, the belief is false and the process is unreliable primarily because the description "red bird" is too general. Indeed, the generality problem seems insurmountable as in many cases we will have multiple options for describing belief-producing processes.

A second problem for Goldman's account that relates directly to the generality problem is the

Single Case problem (Goldman 1986, 49-50). In contrast to the generality problem, the single case problem emerges when we narrow our description of a cognitive process so extensively that it picks out only one instance of a single type of process of which there are and never will be other instances. For example, if I form a belief at noon on a windy day in October 2007, that my friend Julie is walking towards me on the Cathedral lawn, and I describe the visual process that produced this belief in a similar fashion, what are the chances that the conditions specified in that process type will ever be instantiated again? The answer is that the description of the type is so narrow that it is highly unlikely. This emerges as a problem for reliability ascriptions on Goldman's account because multiple tokens of a given process type are necessary for assessing the reliability of that type. Yet, if the description of this process is made more general, it pushes us head on into the generality problem. Again, if the description of a process type is too general, assessing the reliability of that process will fail.

The way in which Goldman (1986) attempts to accommodate the Single Case problem is by abandoning a probability interpretation of reliability for a propensity interpretation (1986, 2004). On this interpretation, the reliability of a type of belief-producing process is tied up with the *relative likelihood* of that process producing a higher ratio of true compared to false beliefs. The propensity account requires us to consider the reliability of a belief producing process without any real-world instantiations of that process. However, it seems unlikely that we will be able to specify how likely it is that a cognitive process will be prone to produce true or erroneous beliefs in the abstract. Furthermore, although the propensity interpretation may help us cope with the problem of single cases, it offers no solution to the generality problem.

I am interested in the aforementioned features of Goldman's account only insofar as they provide conceptual resources that can be used to begin to construct an account of the reliability of

experiment. Applying Goldman's process reliabilist account of cognitive processes directly to the evidence-producing processes we encounter in science, we can say that an evidence-producing process is *reliable* just so long as it produces a "high" ratio of good compared to erroneous data, where good data can be used to support true claims about *effects produced in a laboratory* and erroneous data cannot. Of course, this will require us to differentiate good data from bad data. I will return to this issue in Section 3.0. For now, on this account, we can say that evidence-producing processes come in types and each process type will have individual tokens or instances. In order to assess the reliability of a given process type on a probabilistic interpretation of reliability, the ratio of good compared to erroneous data produced must be assessed across all tokens of that type to determine if it is high or low. On a propensity interpretation, the reliability of the process type would instead be assessed in the abstract in terms of the *relative likelihood* that the type will produce a higher ratio of good compared to erroneous data. Additionally, the reliability of a type of evidence-producing process will be assessed in terms of both its context-dependent or local reliability in one instance of its use in one lab as well as its reliability across all instances in which it is used in laboratories (i.e., global reliability). I will further develop this account in the sections to follow, but for now I want to highlight some of its additional features.

Whereas the generality and single case problems may indeed be problems for assessing the reliability of the cognitive processes of individual cognizers, I do not view them as immediate threats to the enterprize of providing a process reliabilist account of the processes used to produce data in science. This is because experimentation requires that a given process type be fixed descriptively in an experimental design and protocol in advance of producing individual instantiations or tokens of that process type (i.e., individual experiments). This fixes the generality problem in so far as the general description of the process type can be regarded as fixed or constant. Similarly, a primary

way in which scientists attempt to ensure the reliability of the experimental process is by ensuring similarity across individual instantiations of a given experimental design and protocol both in the abstract and across individual uses of that process. This means that each token (i.e. trial) of a given experiment will be similar enough to the process type, that there will be many rather than merely a single instance of the type. So, in science, we rarely encounter situations in which we run into either the single case or generality problems. I further develop this point in Sections 3.2.2 and 3.3.3.

Experimental processes also differ from cognitive processes in so far as scientists believe they can exert control over them. In science we run across strategies for decreasing the error of experimental processes (e.g., Mayo 1996, 2000; Woodward 2000). Some philosophers of experiment have been concerned with characterizing these strategies in the context of developing both descriptive and normative accounts of the reliability of experiment (e.g., Hacking 1983; Mayo 1996, 2000; Woodward 2000). In the next section, I offer a brief description of two primary accounts of reliability on offer in this literature.

3.2.2 Reliability and the Philosophy of Experiment

The term "reliable" is applied to several different aspects of science and its activities by philosophers of science. Some authors have applied the term to data, evidence (Bogen and Woodward 1988; Hacking 1983; Mayo 2000), data interpretation (Woodward 2000) and knowledge claims. Others have used the term to refer to instruments, methods, estimates, measurement procedures, detection techniques and experimental arrangements (Bogen and Woodward 1988; Delahanty 2005; Mayo 1996, 2001; Woodward 1989, 2000). Some accounts have been directed at assessing the reliability of specific experimental techniques such as functional and other imaging techniques

(Bogen 2001a, 2001b, 2002a, 2002b; Delahanty 2005). Clearly, not all accounts of reliability will be equally applicable to all areas of science. In what follows, I want to briefly consider two philosophical accounts of reliability offered by Mayo (2000) and Woodward (2000), respectively. My aim here is not to engage in a critique of these accounts, but to provide two accounts of how reliability is achieved in science against which I can contrast my own account.

Mayo can be taken to ascribe reliability to a *complete procedure* for testing a hypothesis. This complete procedure may require a hypothesis to be put through a series of severe tests, from experimental design to data collection to data analysis and error correction. Such a complete testing procedure is *reliable*, on her account, if it satisfies a *criterion of severity*. A testing procedure satisfies the criterion of severity if and only if there is a high probability that a test T will not allow a hypothesis H to pass T with evidence e, if H is false (Mayo 1991, 529). As Mayo suggests, "requiring severity is similar to requiring that a test have a high power to detect the falsity of [a hypothesis] h" (Mayo 1991, 530). The relevant question, then, is how such severity is "built into" experimental testing procedures so as to ensure their reliability, or the likelihood that they will not yield erroneous results.

Mayo's suggestion is that error-statistical methods can be built into an experimental testing procedure at various stages and in various uses of that procedure. The precise details of her account are complex and a complete description of them is not relevant for my purposes here. What is relevant are some of the basic details of the primary case study from the history of science that she uses to illustrate how severe tests are constructed, namely, Eddington's experiment. The aim of Eddington's experiment was to collect data during a total eclipse of the sun in 1919 to discriminate between three competing hypotheses pertaining to how gravitation affects light: Einstein's, Newton's and the null hypothesis, which each contain a predicted estimate of "the deflection of

light at the limb of the sun" (Mayo 1996, 280). In Mayo's words (Mayo 1996, 278), a total eclipse provided a "highly severe test" of Einstein's hypothesis, originating from the theory of general relativity, because it provided the ideal conditions necessary to obtain the data required to falsify the hypothesis–conditions that did not occur often.

However, as Mayo points out and as physicists at the time acknowledged, the severity of the test itself could not be realized in practice given a variety of confounding variables that prevented the ability to obtain accurate measurements (e.g., bad weather conditions, lack of requisite observations due to technological limitations). Methods for reducing the errors in these measurements–error statistical methods such as the "statistical method of least squares (regression)" were used to overcome the limitations of the "observations", and to assign probabilities to the experimental results predicted by each of the three alternative hypotheses about the deflection of light (Mayo 1996, 282). According to Mayo, "this permit[ted] severity to be calculated" (Mayo 1996, 282), and Einstein's hypothesis passed this and several additional tests of severity.[3] On Mayo's account, then, the reliability or severity of a test is achieved, in part, by means of such error-statistical/error-correction methods as those we find in the Eddington's experiment.

While I agree with Mayo that subjecting an experiment to severe testing procedures of the kinds that she itemizes is an important way in which reliability is achieved in science, it is not the only way. While error-statistical methods provide one good way in which reliability can be achieved, as I will show below, they comprise only a subset of the kinds of strategies employed for ensuring the reliability of data production processes; there are other strategies that are equally as important.

Woodward (2000) also offers an account of the reliability of experiment. In contrast to Mayo,

[3]For example, it passed when a rival hypothesis was introduced and when alternative explanations for the observed results were provided.

he ascribes reliability to measurement or detection procedures. On his account, a measurement or detection procedure is *reliable* if it supports a systematic pattern of counterfactual dependence between "the data and the conclusions investigators reach" about phenomena on the basis of that data (Woodward 2000, S166). As Woodward puts it:

> We can think of cases of *detection* as a special case of this pattern in which there are just two possibilities–either the phenomenon is present in the experimental context (P_1), or it is not (P_2)– and two corresponding data outcomes or sets of data outcomes D_1 and D_2. Associated with these will be conditionals of the form [if D_j is produced conclude that P_i is true] telling us whether to infer P_1 or P_2 depending on whether D_1 or D_2 is produced.(Woodward 2000, S167)

Woodward adds that ultimately what we want to hold between individual data points and phenomena is a set of conditionals with the requisite "counterfactual sensitivity in both the production and interpretive phases of the detection and measurement procedure" (Woodward 2000, S167) such that the data produced instructs us to infer a phenomena claim P_1 iff P_1 is correct.[4] Woodward identifies a variety of different strategies for establishing the reliability of an experimental procedure including: (1) learning about "the error characteristics of repeatable methods" (Woodward 2000, S167), both by predicting and determining the kinds of errors to which such methods give rise in various applications (Woodward 2000, S168) and developing strategies (including statistical ones) to correct for them; (2) using repeated applications of an experimental procedure in order to detect enough instances to convince oneself and others that a phenomena claim is indeed genuine (Woodward 2000, S169); (3)"focus[ing] on highly specific local empirical facts about the causal characteristics of [a] detection or measurement process and investigat[ing] these by means of strategies (like calibration [. . .]) that need not involve explicit derivation." (Woodward 2000,

[4]The counterfactual condition that Woodward specifies is characteristic of Robert Nozick's (1981) tracking theory of knowledge. Nozick's account is directed exclusively at determining the *local reliability* of a belief producing process, where a reliability assessment is made on the basis of a one-trial use of a belief-producing process relative to a specific context.

S170); (4) determining the error characteristics of a method, and assuming that the method will have the same error characteristics when a novel phenomenon is investigated (Woodward 2000, S171) and (5) discarding data "when one has reason to believe that the process that has generated the data is unreliable" (Woodward 2000, S177).

I take each of the approaches that Woodward introduces to capture the ways in which reliability is achieved in practice in science. Furthermore, Woodward does not take the list that he provides to be an exhaustive one. Where his work on reliability leaves off, additional work on the issue of what makes a data production process reliable may begin. This is my goal in the next section. In what follows, I approach the issue first by means of investigating a case study from the neurobiology of learning and memory in which several important ways in which to insure the reliability of a *complete data production process* emerge. In order to set up this account, I will first begin with a general description of learning experiments in neurobiology in which whole organisms are trained in learning paradigms[5] and used for different types of molecular intervention experiments. Then, in Section 3.4, I will consider the nature of the relationship between reliability and repetition in neurobiological experiments.[6]

[5]I explain what I mean by "learning paradigm" in Section 3.3.

[6]One feature that differentiates my account of the experimental process from that of Mayo and Woodward is that I restrict applications of reliability to data production processes. Both Mayo and Woodward ascribe reliability variously to experimental techniques, procedures and arrangements. On their accounts, these aspects of the experimental process are deemed *reliable* when they yield data that results directly and exclusively from a discrete experimental manipulation. I think this is the essential idea captured when Mayo refers to an experimental arrangement or test as "severe" and when Woodward identifies a detection technique as "reliable". What ensures the direct connection between an experimental manipulation and data is not only, as scientists themselves have claimed, "the degree of exactness with which a measuring device measures a manifest [observable] variable" (Krauth 2000, 258), but also the human element, the scientist, who takes steps to secure and maintain the precision of the measuring devices and the stability of the system from which the measurements are being taken. In the biological sciences experimentalists play a prominent role in designing, implementing and refining experimental systems, arrangements and techniques. The fact that Woodward and Mayo suggest that measurement devices or experimental arrangements come to have precision via methods of error-reduction and error-elimination suggests that they take experimentalists to play a similar role in experimentation in other areas of science. If I am correct about the crucial role that experimentalists play in ensuring

3.3 DATA PRODUCTION: DESIGN AND IMPLEMENTATION

Many types of experiments undertaken in neurobiology have as their aim the identification of the molecular *mechanisms* of learning and memory.[7] I want to focus on here on one general type of experiment: (a) organism-level learning experiments.[8] The starting point for these experiments is

the reliability of their experiments and that they play this role in varying degrees at different stages in the process of producing data, then reliability is better ascribed to the *complete process* of producing data rather then to the objects (techniques, experimental arrangements) involved in this process. So what is needed is an account of reliability of experiment that both construes experimentation as a process and adequately appreciates the fundamental role of the experimentalist in this process. In the next two sections, I aim to develop such an account.

[7]The primary types of experiments that I am considering in this paper were undertaken in the lab of D.J. Sweatt and colleagues. See Atkins, et al. 1998. However, these experiments are representative experiments for similar types of learning experiments undertaken in neurobiology. I have in mind, in particular, the experimental work of Disterhoft and Thompson (nictitating-membrane response in the rabbit), LeDoux, Nader (fear-conditioning paradigms), Dudai (e.g., taste aversion paradigm), Morris (Morris water maze). My use of Sweatt's work is due primarily to having worked on the hippocampus and the extracellular signal-regulated kinase signaling cascade, the importance of which he established for both hippocampus-dependent learning and hippocampal synaptic plasticity (Sweatt and English 1996).

[8]A second type of experiment that will concern me in the remainder of this dissertation are *in vivo* plasticity experiments. In these experiments, in brief, animals are anesthetized, stimulating and recording electrodes are lowered into their brains to a synaptic location of interest in a brain area of interest *assumed* to be involved in some type of learning. For example, it is common to look at synapses in the hippocampus thought to be involved in associative learning (e.g., area CA1). The stimulating electrodes (1-2) are placed proximal to the presynaptic neurons at the synapse of interest and a recording electrode is placed proximal to the postsynaptic neurons to which the presynaptic neurons project. To take an example from a representative lab, a series (e.g., around 10) of test pulses of 0.1 Hz is delivered in 5 min intervals to pre-synaptic neurons and the average potential evoked in the post-synaptic neurons is recorded for each test pulse. The average response of the 10 test pulses is calculated and recorded. Once a stable response (and how this is defined varies from lab to lab, but in the lab I am considering, a stable response is indicated by an evoked potential that stays constant within a 15-20 percent range over 3 sets of 10 pulses applied in a series with 5 minute intervals in between them.) is achieved, it is taken as the baseline response and recorded by the experimenter. Five minutes elapse and then artificially induced trains or bursts of electrophysiological stimulation are delivered to pre-synaptic neurons. Five minutes after these trains or bursts of stimulation end, the investigator again delivers the same stimulation protocol as was applied prior to the train, namely a series of (10) test pulses of 0.1 Hz at 5 min intervals. This continues for as long as the experimenter decides to look at evoked potentials after the train. In these experiments, an increase a of 20-75 percent in the average amplitude of the evoked response of the post-synaptic neurons from the established pre-train baseline (average) evoked response that lasts for at least 30 minutes is taken to be indicative of LTP. It should be noted that once it has been established that a given stimulation paradigm can produce the desired kind of synaptic plasticity effect, then that paradigm is used to induce effects that are not LTP according to the aforementioned operational definition, but that are assumed to be LTP, even if there is no indication as to how

to assume, based on previous findings from neuropsychology and cognitive neuroscience, that a particular learning function (e.g., fear conditioning) is localized to a particular brain area (sp., the amygdala). In the type of experiments I am interested in, in basic terms, a group of intact rats (experimental group), is trained in a learning paradigm (e.g., classical or operant conditioning). Learning is defined operationally as a measurable change in a behavior or the frequency of a behavior over a designated baseline response measured prior to training. If a change in the rat's behavior is noted, its brain is removed and that brain area taken to subserve the type of learning

long-lasting they are.

Then, in order to get at those molecules that are implicated in the production of LTP, the investigator removes the brains of animals at different time points after the end of the trains or bursts of stimulation, and removes the tissue proximal to the recording electrode (i.e., proximal to the post-synaptic neuron)[9] as well as tissue from the same brain area (i.e., hippocampus) but distal from the location of the recording electrode (control = unanalyzed tissue) and then analyzes the tissue biochemically (e.g., by Western blot) for the molecule(s) of interest. If the molecule is shown to have a change in activity compared to animals that received only sets of test pulses, then another set of experiments is conducted. In these experiments, the experimenter injects the organism with or places in the brain area of interest by means of a cannula, an inhibitor of a molecule or molecule(s) hypothesized to be components of the mechanism productive of LTP. This is done at some time prior to delivering the trains or bursts of artificial stimulation to presynaptic neurons. After delivery of the train the effects of the inhibition on the post-train evoked response compared to the pre-train average evoked response on the production of LTP is recorded. If the potentiated effect is not observed, and biochemical analysis of the relevant tissue of the animal reveals that the activity of the molecule was indeed blocked, then the inhibition of LTP is attributed to the inhibition of the molecule. The combined data from these sets of experiments is used to answer mechanistic questions about LTP or to discriminate between competing mechanistic hypotheses of LTP.

Research studies of the kinds I have described contain several different kinds of experiments. In the first kind of experiment, an effect of interest is produced, a brain area of interest is removed and the activity (phosphorylation state) of some molecule of interest contained in the extracted tissue is investigated biochemically and measured. These experiments establish merely a *correlation* between the effect of interest and the activity of a molecule. Once this correlation has been established, then, the effect of interest is produced again while the investigator intervenes pharmacologically to alter the activity of the molecule of interest and determine the consequences of this intervention on the production of the effect. In conjunction with these experiments, in order to verify that the intervention was successful, tissue from the brain area of interest is removed, the activity of the target molecule is analyzed biochemically and measured to confirm that the pharmacological manipulation did in fact have the desired impact on its activity. The results of each of this second set of experiments in combination with the first are meant to yield data relevant to discriminating *mechanistic* claims about the involvement of the molecule of interest in the production of the effect of interest. When I refer to synaptic plasticity experiments in later sections of this dissertation, it is these experiments that I will have in mind.

the paradigm is supposed to produce is analyzed biochemically for activation and total levels of a molecule of interest, namely, some protein suspected to be one operative variable of the mechanism that produces the learning under study. These results are compared to results obtained from animals not trained in the learning paradigm (control group) or exposed to some condition different from the animals in the experimental group under study.[10] These experiments establish merely a *correlation* between the effect of interest (the form of learning under study) and the activity of a molecule.

In another set of experiments, a rat is trained in the same learning paradigm, while the activity of some molecule hypothesized to be a component of the mechanism that is productive of the change in behavior compared to baseline (i.e., learning) is blocked pharmacologically. The effects of the inhibition of this molecule on the behavior of the rat are studied. Verification that a molecule was indeed inhibited during training is obtained by decapitating the animal, removing its brain and analyzing the tissue of interest biochemically to verify that the activity of the molecule(s) was indeed blocked. This is achieved typically by comparing activation levels of the molecule to total levels of the molecule in Western blots prepared using brain tissue samples. The Western blots are visualized by means of chemiluminescence techniques that are applied directly to a membrane containing the proteins. The membrane is then exposed to a film, which results in the production of a film containing an image of the bands of protein that the membrane contains. The bands are then measured to determine if the activity of the molecule of interest contained in the brain area assumed to be involved in learning in the animal in which the activity of the molecule of interest was blocked, differs from tissue in the brain area of an animal not trained in the learning paradigm (control). If the activation level is similar in the two cases, then the investigator concludes that the activity of the molecule was indeed blocked and that the activation of the molecule of

[10]Depending upon the nature of the experiment and the dimensions of the variables studied, additional control groups may be requisite.

interest is "implicated" in the learning. The results of these intervention experiments combined with the results of the first type of experiment are meant to yield data relevant to determining the truth or falsity of *mechanistic* hypotheses about the involvement of the molecule of interest in the production of learning.

With this general understanding of organism-level learning experiments in neurobiology, I want to begin to characterize the process involved in producing data to substantiate mechanistic claims about organism-level learning. I take experimentation to be best construed as a process– the *experimental process*, which is represented in *Figure 7*. To characterize the stages of this process, designated in part by the arrows in the figure, I will borrow and slightly refine Woodward's distinction between *data production* and *data interpretation* and show how the process of data production can be understood in terms of Goldman's description of process types and individual tokens of those types.

On my account, data production may be divided into two discrete stages: (1) design and (2) implementation. The design stage, in basic terms, involves the development of an experimental design and protocol. In organism-level learning experiments, the design stage is set in motion by a question about the mechanisms involved in the production of an effect of interest, namely *learning* (Figure 7). In order to address such questions, an investigator must produce an effect in the laboratory that is similar to the original effect of interest. This requires the development of an experimental design that includes a standardized protocol that is to be implemented in the controlled environment of the laboratory. In conjunction with the development of an experimental design and protocol, a set of competing hypotheses (in *Figure 7* these hypotheses are represented by H1, H2, H3) about the effect of interest are identified. For example, in the experiment that I am considering, one set of preliminary hypotheses included: (H1) Extracellular signal-regulated

kinase/mitogen-activated protein kinase (ERK/MAPK) activation is increased during mammalian associative learning, (H2) activation of ERK/MAPK is decreased during mammalian associative learning and (H3) there is no change in ERK/MAPK activation during learning.

In the design stage, conceptual distinctions are drawn, step-by-step protocols are written up, model organisms, techniques, equipment and methods are selected. In effect, the design stage is the conceptualization stage of data production. It typically proceeds in discrete stages as questions are posed, suggestions about how to address them are provided, projections are made about potential problems that might be encountered in the course of implementing the design, tentative solutions to these problems are offered, and the combined considerations are worked into the design and protocol that contains a set of rules that must be followed to ensure the reliability of each implementation of the design. The design stage may be regarded as descriptively specifying a *process type* that consists of a range of features that must be instantiated in each individual instance of the type. These features are contained in the experimental design and protocol, which are, in essence, the process type. The description of the process type will be specific enough to allow for the identification of individual tokens of the process type, and to exclude those tokens that do not conform to the process type in terms of those features taken to ensure the reliability of the type. Ultimately, a process type will have a built-in propensity to be reliable before it is implemented in the laboratory. Using Woodward's (2000) terminology, an experimentalist wants the likelihood that the type will be reliable when it is implemented in practice to be "high". However, problems pertaining to the reliability of the process type may arise during its implementation in the laboratory. In such cases, changes will be made as needed to the design and protocol in order to re-establish the reliability of the process type.

A complete description of the ways in which concerns about reliability constrain the develop-

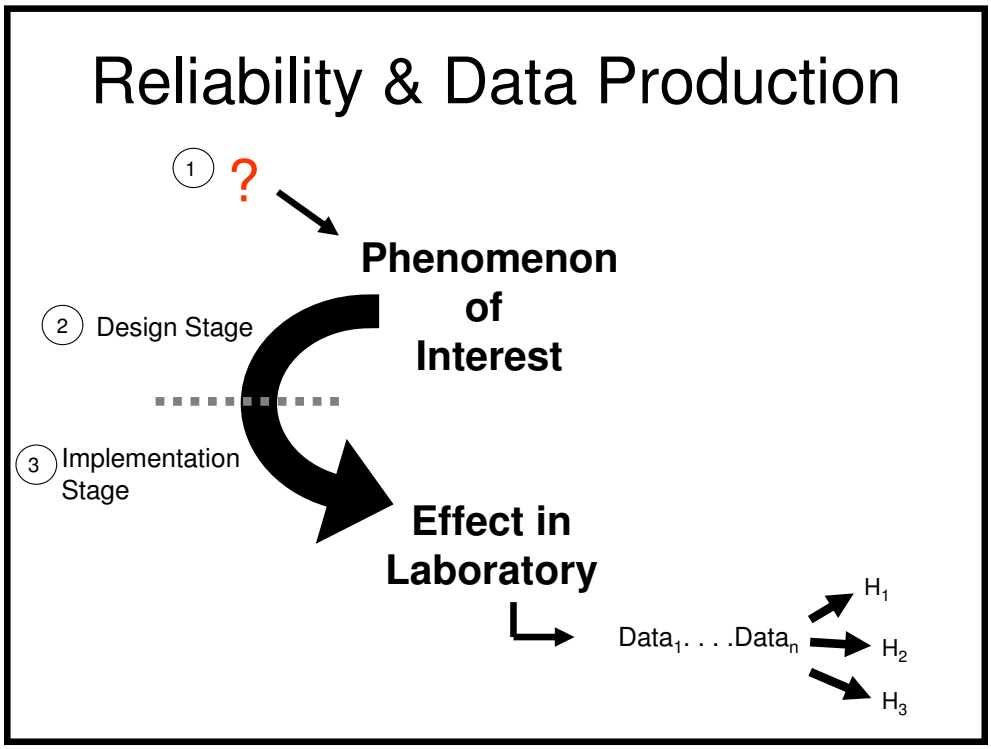


Figure 7: Reliability and Data Production

ment of experimental designs in the neurobiology of learning and memory is beyond the scope of this dissertation. However, in order to provide the reader with an idea of the kinds of measures taken to ensure reliability at this stage, I want to offer several examples. The primary issue addressed at the outset of the *design stage* in neurobiological organism-level learning experiments is how to produce learning in a way that retains those features that set it apart as a unique phenomenon but eliminates those factors that impede the discovery of the variables that contribute to its production. It is in the move from the world outside to the world inside the laboratory that crucial assumptions are made and where concerns about the reliability of the data production process begin to be operative (*Figure 7*). If an investigator wants to describe some feature of the effect of interest or to isolate what causes it, she necessarily has to direct the question at an effect that can be accompanied feasibly by a definitive set of competing hypotheses or claims about the effect and what produces it. Effects in the world are not like that. So, an investigator must essentially translate the effect into her own lab.[11] Part of securing reliability in the process of data production involves simplifying the effect of interest just enough to narrow down the range of possible hypothetical claims concerning its features or the candidate mechanisms that produce it. This narrowing down is rarely, if ever exhaustive–some potential hypothetical claims may never be formulated in the itemization process because they the investigator conducting the study may overlook them. An investigator strives in the design stage to be as inclusive as possible, yet (in the ideal case) remains open to the possibility that during the implementation stage, a novel discovery may require the formulation of additional hypotheses to be added to the current list of candidates, and modifications to the experimental design in light of such discoveries might be necessary. In addition, it should

[11] I want to add here that my description of how this process unfolds is at best a rational reconstruction. In contemporary neurobiology, I think it is rarely the case that anyone starts with some effect "in the world outside the laboratory".

be noted that strict adherence to the aim of narrowing down the range of potential hypotheses in the development of the experimental design, which is a normative prescription of reliability, often has negative ramifications relative to extending claims about the form of learning produced in the laboratory back to the original effect of interest outside the laboratory.

The process of narrowing down the range of competing hypotheses is achieved in part by developing and designing a highly controlled set of procedures or testing conditions by which to produce and detect a simplified version of the effect of interest. It is in the process of providing an *operational definition*[12] of the effect, that the effect is simplified. In the neurobiology of learning and memory an operational definition of an effect like learning is built directly into the design of an experimental paradigm. An experimental *paradigm* is *a standard method or procedure for producing an effect of a specific type*. The following features are typically included in the design of a learning paradigm: (1) *production procedures*, namely, a specification of the stimuli (independent or input variables) to be presented to the organism, how those stimuli are to be arranged (e.g., spatially, temporally) and how many times they are to be presented during phases of (a) pre-training (b) training and (c) post-training; (2) *measurement procedures* that specify the response variables to be measured in the (a) pre-training and (b) post-training phases of the experiment and how to measure them using apparati designed for such measurement; (3) *detection procedures* that specify what the comparative measurements of the response variables from the different phases of the experiment must equal in order to be able to ascribe learning to the organism. This detection procedure is simply an operational definition that specifies the measurable"change" in response variables from pre- to post-training as a result of the stimuli presented that must be observed in order to say that learning has been occurred. It is a comparison of the data points produced pre-

[12] I define "operational definition" later in this paragraph.

and post- training that is used to make such a determination.

A protocol is that part of the production procedures of an experimental paradigm that will specify, for example (1) the duration of time of the presentation of each stimulus to be used in the experiment (2) the duration of time that is to elapse between presentation of the stimuli used in an experiment: inter-stimulus interval (ISI), (4) the amount of time that is to elapse between individual trials (the inter-trial interval (ITI)) and (5) the amount of time that is to elapse after the last trial before the brain is removed. A protocol is selected in the design phase for the express purpose that it can be used to produce the effect of interest and the effect is measurably "marked" or "robust". Sometimes a determination of the effectiveness of a given method for inducing learning will require preliminary trials using different protocols using the same paradigm to determine which protocol is best suited for the purposes of the study. When the aim is to identify the mechanisms of learning, a protocol that produces robust learning always is taken to be superior to one that produces a less marked form, since one can always determine when a robust effect is present. This aides in the operationalization of the effect, as the features will be more easily detectable and measurable, and there will be less error in detecting when the effect occurs.

What an experimental learning paradigm in essence does is describe a controlled context that has attached to it a finite set of possibilities for producing an effect, and hence a finite set of hypotheses that correspond to those possibilities. So, what must be mapped out in the design stage are specific guidelines to guarantee that (1) the paradigm used to produce the effect is exclusively what produces the effect, (2) the procedures for measuring the effect are discrete and appropriate for capturing it, and (3) the methods used for detecting the effect are aptly designed for such detection.

Organism-level learning experiments also require the use of biochemical techniques to deter-

mine the activity of molecules at the purported time that learning was produced or its production was interrupted. Time constraints that must be adhered to by an investigator during an experiment are included in the design of a *biochemical protocol*. The protocol specifies that the brain of an animal must be removed and kept cold during dissection and frozen as soon as possible following tissue removal in order to guarantee that the tissue does not degrade and the phosphorylation status of the molecules does not change as a result of a decrease in temperature. If the tissue is compromised as a result of changes in temperature, then it cannot be used to produce data that can be used to make knowledge claims about the molecules that are productive of learning. Ultimately the steps of the biochemical protocol are developed so that the blots produced will capture directly the activity of those molecules at the time that the brain was removed. So the precise timing of steps and the requisite temperature at which to keep samples are specified in the experimental design and are necessary for ensuring the reliability of the process type and its tokens. In addition, the precise concentrations of buffers and antibodies to be used in all biochemical processes are also designated in the experimental design. If any step of a biochemical protocol is overlooked or performed incorrectly, this may result in data that does not accurately reflect the status of the molecules of interest in the brain at either the time of the production of the effect or shortly thereafter.

The implementation stage of data production begins after an experimental design and protocol has been completed. It involves individual instantiations of the experimental design by means of the systematic following of the experimental protocol using the equipment, materials, and techniques laid out during the design stage. At this point an investigator takes individual subjects or groups of subjects/samples/groups of samples at a time, and runs them through the steps of the protocol, following those steps as precisely as possible. The *immediate* output of each individual implementation of the design is a data point or set of data points. A given implementation of the

design will only be accepted as a token of the process type if it conforms to the type in terms of the relevant range of features. If any of those features are missing data obtained by means of that implementation of the type will be excluded from the data pool.

During the implementation stage in neurobiological experiments an investigator plays an active role in ensuring that each instantiation of the process type conforms to the features relevant for reliability. In implementing the design an investigator must be alert to any extraneous variables that may affect a given instantiation of the design and compromise its reliability. As is often the case with biological organisms, especially in intervention experiments in which pharmacological manipulations are used to block molecular activity, the stability or normal behavioral function of experimental subjects (rodents) is sometimes compromised. Such events are sometimes unforseen and so, measures to address them are not necessarily specified in the experimental protocol. For example, one specific issue that arises in every experiment with an organism is keeping the organism stable, whether it is under anesthesia or has been injected with a drug that compromises its normal functioning. Although rough estimates can be used to gauge the viability of the animal–or the stability of its vital signs, sometimes changes occur for which a reaction must be decided on the spot. To take another example, sometimes the equipment simply will not work and problem solving strategies must come on-line that are not necessarily contained in the protocol. So local concerns about reliability will be present in each instantiation of an experimental design. The aim will be to attain the reliability that was supposed to be guaranteed by the design and protocol. However, as long as these events can be dealt with in a way determined (sometimes post-hoc) not to compromise the overall reliability of the individual implementation of the experiment in which the event has occurred, the data obtained from such experiments is included in the statistical analysis of the data.

With respect to biochemical experiments, both accuracy and timing are crucial to ensuring that a given instantiation of the design and protocol match the process type exactly. For example, a biochemist must continuously maintain the temperature of protein samples produced during the implementation of a biochemical experiment and must monitor each of the processes involved in the production of the membranes to which the proteins are blotted. This includes the careful following of those steps contained in recipes for making fresh solutions and buffers. If any step is skipped or performed incorrectly, the experimentalist will not be able to determine with any degree of precision whether or not the data obtained was the product of a process that was accurate to the process type and therefore, reliable.

Changes to an experimental design and protocol may result for reasons other than a recognition that an experimental method is prone to error. For example, it might be decided that one procedure may be more economical than another (e.g., a type of drug being used) or more easy to perform than another. A change needed in the implementation of the experimental design will typically result in a modification to whatever aspect or component of the design is discovered to be problematic.

Once enough data points for each type of experimental manipulation have been collected the data points taken together form a complete data set and each data set is analyzed statistically. This brings us to the issue of where statistical methods fit into the data production process. Ultimately the aim of selecting a statistical method is to guarantee that the use of the method will accurately reflect actual relationships contained in each data set. The application of a statistical method to a given data set is a one-trial process, so it does not fit neatly into the account of data production that I have provided. However, what can be said about statistical methods is that they do not derive their reliability from their use in analyzing the data produced in the context of one set of experiments. Rather, they attain a public designation of reliability only when they have been shown repeatedly to

have a high probability of accurately capturing information contained in many data sets. This only comes from their repeated successful use in different areas of science or from better theoretical analysis.

In neurobiological learning experiments, typically data collected from six to ten subjects for each manipulation is taken to be sufficient for analyzing the data statistically and determining its statistical significance. The statistical methods will be selected with the aim of maintaining the reliability that has been built into the experimental design and protocol and secured during the implementation stage of data production.

Statistically analyzed data is used to identify a true hypothesis from a set of competing hypotheses about the effect of interest produced in the laboratory. Once a true mechanistic claim has been identified, it is first interpreted as true of the effect produced in the laboratory. This initiates what I identify as the stage of data interpretation. I describe this stage of the experimental process in detail in Chapter 4.0. [13]

3.4 WHAT DOES REPETITION DO FOR RELIABILITY?

Some of the claims that I have made up to this point in the chapter suggest that *repetition*, i.e., more than one trial of a given experiment, is important for the reliability of a data production process. The obvious question is, then, what does repetition do for/with respect to reliability? In

[13] I describe this stage in detail in Chapter 4.0, with reference to Figure 8. Then, the same mechanistic claim is extended back to the original effect of interest in the world that prompted the empirical question about the mechanisms productive of learning in organisms in the first place. This is represented by arrow (3) in Figure 1. As I mentioned in my critique of Woodward (2000), interpretation is a one-shot process, and so, neither of these interpretive steps can be characterized as reliable or unreliable. Rather, philosophers of science should concern themselves with the *legitimacy* of each interpretive step.

this section, I will investigate the relationship between reliability and repetition with respect to neurobiological experiments in order to provide one answer to this question.

A fruitful way to begin to think about the role of repetition with respect to reliability in experimentation is to think about what the consequences for reliability would be in the absence of it. The goal of those organism-level learning experiments that I have described in this chapter is ultimately to establish a causal link between ERK activation and learning, i.e., that a change in ERK activation must occur for learning to occur and if it does not occur, then learning will not occur.[14] One set of experiments is designed to detect changes in ERK activation during learning. A second set of experiments is designed to inhibit ERK activation during a learning event by putting an animal that has been injected with a drug through a learning paradigm to see what effect such inhibition has on learning (i.e., they perform an "intervention" experiment cf. Woodward 2003). In what follows, I want to consider what would be gained if the experiments that comprise these two sets were all single trial experiments.

Suppose an investigator takes 1 laboratory rat and trains it in a cued fear conditioning paradigm, with the aim of determining if the activation of the molecule ERK was increased, decreased or unchanged during learning. These are the three hypotheses. Let us say for the sake of argument that he has followed the directions for running the experiment set out in the experimental design and protocol. He runs the learning experiment and then investigates ERK activation by means of

[14]This is why neurobiologists who work on learning and memory typically describe their goals as one of identifying the cellular and molecular mechanisms of learning and memory (e.g., Sweatt 2003). The ultimate conclusion of the Atkins et al. 1998 study is that MAPK/ERK is "required" for mammalian associative learning (Atkins et al. 1998, 602). One aim of recent work in the philosophy of science is to understand the nature of mechanisms and when they are explanatory (e.g. Bectel and Richardson 1993; Machamer, Darden and Craver 2000, 2007). Oftentimes such descriptions of mechanisms are meant to imply causal links between variables even in cases in which the words "implicated" or "involved" are the words employed–in other words, that a molecule plays *some* role in the production of an effect such as learning.

biochemical analysis of tissue removed from area CA1 of the hippocampus of the animal's brain. He then runs a second experiment, in which he removes the brain of a laboratory rat that has not been trained in a learning paradigm and he analyzes tissue from area CA1 via biochemical methods (e.g., Western blot). Suppose at the end of these experiments, he arrives at a set of data points[15] and these data points discriminate the hypothesis that ERK activation is increased during learning from the other two competing hypotheses about ERK activation during learning.[16]

In a second set of experiments, which include the "intervention" experiment, the investigator takes another laboratory rat, injects it with an inhibitor of the ERK cascade (sp., a drug, e.g., SL-327), and trains it in the same learning paradigm as the rat that received no such injection. He then compares the data points that he obtains with respect to learning and the ERK signal from these experiments to those data points from the first two experiments mentioned above. Together, the data indicate that the rat does not learn and that ERK activation was blocked in the brain of the rat that received the SL-327 injection. Furthermore, they indicate that ERK activation is increased during learning and that when ERK activation is blocked, learning does not occur.

Given these sets of experiments alone, can we say that this complete data production process was reliable? In other words, can we say that it results in data that can be used to discriminate among competing hypotheses? In one sense, we can say that *some* steps have been taken that establish reliability. For example, even though these were all single trial experiments, the investigator followed those rules of running these types of experiment laid out in the experimental protocol.[17] We can take this rule following as one indicator of the reliability of the data production process.

[15] I say "a set of" instead of "two" data points because he is interested in (a) whether the animal learned and (b) whether there is a change in ERK activation

[16] In other words, the ERK signal observed with respect to the rat put through the learning paradigm compared to that observed with respect to the animal not but through a learning paradigm is of a greater magnitude.

[17] And, I am assuming here, for the sake of argument, that this is an otherwise "good" protocol for the purpose of hypothesis discrimination.

The question is, then, is this indicator enough given that the data is supposed to serve the function of discriminating between the three competing hypotheses?

One way to answer this question is to consider the wealth of background information that there is about the behavioral, molecular, and physiological triggers of ERK activation. ERK activation can occur in response to a *stressful stimulus* presented to a behaving organism (Gioia et al. 2001; Shen et al. 2004). It can also result from direct stress to brain tissue. For example, damage to brain tissue during tissue cutting and changes in temperature that may occur as a result of large time lapses from removal of an animal's brain to freezing the relevant brain tissue may also be accompanied by changes in ERK activation (Belelovsky, Maroun and Rosenblum 2007; Ho, Delgado, and ODell 2004).[18] This suggests at the very least that the standard for ERK activation against which all the other experiments in the study are compared, namely, the baseline animal, may not be a good standard, because there is some chance that the ERK activation observed in this animal may be the result of human error. The investigator may have managed to create unnecessarily a stressful situation for the animal. He may have unknowingly "nicked" the brain during its removal from the skull, which may have triggered molecular activity that affected ERK activation (increasing or decreasing it from "basal" levels). He may also have taken just slightly too long to put the tissue samples on ice. Since the ERK signal observed in those "experimental" animals that were trained in a learning paradigm, or injected with a drug and then trained, must be compared to the ERK signal observed in this one "baseline" animal, if the detected signal fails to reflect ERK activation levels under "basal" conditions, this will prevent the detection of changes in ERK activation from baseline. Yet this is ultimately what the investigator needs to know in order to determine

[18] Specifically, stress to the tissue, such as oxidative stress or Ca2+ influx can trigger ERK activation. Furthermore, tissue slicing alone leads to ERK activation (Ho, Delgado, and O'Dell 2004). When brain tissue is sliced, it results in cell death, that can trigger, for instance, an increase in Ca2+ and activation of a number of signaling cascades, including the ERK cascade.

if there are actual changes in ERK activation during learning. This means that it will be highly unlikely that the investigator will be able to discriminate between the three competing hypotheses about ERK activation during learning.

A primary option that an investigator has at his disposal for combatting such threats to the reliability of the data production process is to run additional experiments in which the brains of animals not trained in learning paradigms are removed and the relevant tissue is subjected to biochemical analysis. The basic idea is that the investigator must be able to rule out that the observed ERK signal is not the result of any of the identified confounding variables. So he must engage in an effort to eliminate any and all confounding variables not only in the baseline experiments, but across each type of experiment. Consistency with respect to following the rules of the protocol across multiple trials is one way to achieve this. The desired goal is to identify the mean and standard deviation from the mean of ERK activation across animals who have not learned, as a comparative measure for investigations of ERK activation in animals who have learned or who have not learned possibly as a consequence of the actual inhibition of ERK activation. Repetition in such cases, then, may regarded as one important indicator of reliability.

However, repetition alone may not necessarily eliminate systematic human errors, because consistency across trials cannot be taken to mean that no confounding variables were present. Perhaps the investigator follows the same protocol multiple times, but manages to unknowingly damage the brain each time–so that the ERK signal across non-trained animals (i.e. across trials) is taken to be within a particular range, but it fails to be an accurate reflection of basal levels of ERK in rat. Clearly, this is a problem, because we are again left with insufficient grounds for discriminating among the competing hypotheses.

If following the rules of an experimental protocol and repetition of that protocol over multiple

trials will not guarantee the reliability of a data production process, what will? This is where additional "data" must come to bear on the data production process. The simple answer is that an investigator will read other research papers and get a sense of what basal levels of ERK activation look like. For the sake of having multiple indicators of reliability, they will get a sense of what ERK activation looks like under many varying kinds of conditions by consulting the available literature in the area–which is considered to be good scholarship and is relevant for publication in reputable journals.[19] If an investigator determines that his results for ERK activation under varying conditions differ from that of other investigators, he will attempt to address what might be going wrong in his experiments by comparing his experimental design and protocol to theirs.[20] In addition, if the investigator moves to submit his paper for publication, the reviewers might arrive at the conclusion that there is a systematic error in his results, given other results in the field and suggest changes to the protocol that might eliminate the error in a way not currently achieved by the protocol. These are other ways in which reliability is achieved. [21] Again, repetition is used as just one indicator of the reliability of a data production process. In such contexts, the main aim of repetition is to prevent experimental errors or artifacts. However, there are clearly other indicators of reliability, which are equally as important as repetition, as I have suggested above.

[19]It may be possible that there is a lack of consistency with respect to the ERK signal under varying conditions across trials–this, if true, will not be a problem that the investigator himself must solve, but one that the entire community should try to solve.

[20]In the design of the experiment, he may have been aiming to replicate the results of another study undertaken previously in another lab, but then found out that there were subtle differences between the published protocol and the actual protocol. For example, that the animals or tissue samples had been handled slightly differently in the that lab than was conveyed in the published protocol.

[21]I made this claim earlier in the chapter when I said that a given investigator must pass the community test of reliability–that he must go beyond what he takes to be a reliable process and be judged by what the scientific community at large takes to be a reliable process. Now, it is quite *possible* that an entire scientific community may be in error, but given the potential idiosyncratic nature of the types of human errors (e.g., "nicking" the brain) that may be involved in neurobiological experiments, I think it is unlikely that systematic errors will be established across an entire community of neurobiologists.

3.5 CONCLUSION

In this chapter, I have sought to provide an account of the reliability of experiment that is based on the idea that experimentation is a process. It is also meant to capture the essential role that experimentalists play in ensuring the reliability of this process and to be accurate to the evidence-producing processes of neurobiology. In order to develop my account, I restricted my application of "reliability" to the process of data production. I described a data production process as reliable insofar as it could provide the data requisite to support knowledge claims about effects like learning that are produced in a laboratory. On my account, the reliability of data production is insured across its two stages: (1) design and (2) implementation. I explained that in the design stage experimentalists take a variety of steps to ensure the reliability of their experimental designs and protocols in the abstract, before they go on to run individual experiments in the laboratory. As I showed with respect to learning experiments undertaken in neurobiology, numerous decisions are made for the sake of reliability in the design stage such as: the development of learning paradigms and protocols, animal models, contexts in which to run experiments, drug concentrations and an experimental protocol with step-by-step instructions for running experiments. Then, in the implementation stage, the reliability of the data production process is achieved by strict adherence to the experimental design and protocol across individual instantiations of that protocol. I suggested, in agreement with Woodward (2000) and Mayo (1996, 2000) that information about errors exhibited across individual instantiations of the design may be eliminated by making modifications to the experimental design and protocol ("the process type") in an effort to reestablish its reliability and the reliability of the complete process of data production.

This conceptual framework also provides a novel way to avoid specific problems for process

reliabilist accounts of cognitive processes. One solution to the generality problem when we consider data production in science is to view experimental designs and their adjoining protocols as types of processes to which we can make reliability ascriptions, with each individual instantiation of the design in the implementation stage comprising an individual token of the type. Since controls on reliability will be operative in the construction of the type (i.e., the experimental design), we can make reliability assessments with respect to the features of the design, and we can compare the individual tokens of the design to the design type in order to determine whether they possess the relevant features that are to make a given implementation of the design reliable. We do not encounter the problem of how to describe or identify tokens of a given type as we can use the type to discriminate the range of features that a token most possess in order for it to qualify as an instance of the type.

Clearly, on my account of reliability, it must be acknowledged that the whole enterprize of ensuring reliability in science is a circular one: one itemizes the features of what one takes to be a reliable process in the abstract and tries to realize them in practice. The type is deemed reliable if individual instantiations of it each point in the same direction when it comes to identifying one claim as true from a set of competing claims. If in the process of instantiating the type, problems such as gross inconsistencies between measured data points arise, then the type is revised and the process begins once again. In this way, the entire enterprize of ensuring the reliability of data production processes in the laboratory sciences is, as Hacking (1992) aptly described it, a *self-vindicating* one.

4.0 VALIDITY OF EXPERIMENT IN THE NEUROBIOLOGY OF LEARNING AND MEMORY

4.1 INTRODUCTION

The validity of experiment has not been widely discussed in recent philosophy of science literature. Nor has it been an issue with which contemporary laboratory scientists in various areas of science have been exclusively concerned. This was, however, not always the case. In experimental psychology in the 1940's, 50's, and 60's, the issue of validity received its most extensive treatment. For example, numerous psychologists became concerned with the validity of experimentation and with what they viewed as different discrete types of validity that pervaded the experimental context. Most notably, work in developing intelligence tests and educational tests for the college boards (SATs) directly concerned the issue of validity–the most general issue being the question of whether such tests could be used to assess intelligence across different ethnic populations not exposed to the same overall educations (Messick 1989). Today the issue of validity and the distinction between reliability and validity has been reinvoked to capture discrepancies between psychiatric categories (e.g., schizophrenia, manic depression) and "the natural world"–the worry being that perhaps current taxonomies of mental disorders fail to cut nature at its joints. However, despite this history, and despite current concerns about validity in areas like psychiatry, little has been

said by scientists and philosophers of science with respect to the issue of validity in contemporary laboratory science. The aim of this chapter is to provide a starting point for bringing the issue of validity back to the philosophy of science and placing it on a par with the issue of reliability.

Given how vast the literature on validity is, however, this chapter and my discussion of validity throughout the dissertation can only be considered a first pass to get a handle on the primary issues. To adequately provide an account of validity and its dimensions would involve me in a far larger project than can be accommodated in this dissertation. So, the conclusions at which I arrive with respect to different types of validity in this chapter are tentative; they are meant to serve more as a foundation for future work in the validity of experiment than as the final word on the issue.

I also want to say at the outset that although the structure of this chapter will to some extent parallel that of the preceding one on reliability, concerns about validity do not explicitly emerge during the process of producing data in neurobiological experiments, so they cannot be said to operate as a constraint over data production. Yet, I deduce such constraints from potential assessments of the validity of interpretive claims that we encounter in the context of data interpretation. Using this strategy, I arrive at certain conclusions with respect to those neurobiological experiments with which I am concerned, about what the prescriptions of validity may be when experiments are being designed and implemented.

The structure of this chapter is as follows. In section 4.2, I consider the ordinary language definition of *validity*. I use it as a means to get clear on how we should use the term specifically with respect to the experimental process. I then proceed to an analysis of those definitions of validity and its types that we encounter in experimental and cognitive psychology. Using this analysis as a basis, I attempt to develop a coherent account of validity and its parameters that specifies the dimensions across which the validity of interpretive claims should be assessed and

how validity ought to operate as a constraint on data production. I end with a consideration of the role of validity in neurobiological learning and memory experiments.

4.2 VALIDITY

It is worthwhile given that I am writing in the context of the philosophy of science, to differentiate the type of *validity* with which I will be concerned in this section from that of logical validity that is common in the philosophical literature. In the context of philosophical logic, "validity" is a feature ascribed to arguments that contain conclusions that follow logically from a set of premises. This notion of *logical validity* is indeed related to the kind of validity that will be the focus of this chapter, although I will not spend a lot of time discussing it. I will only say that in the history of science, particularly in the history of psychology, validity can be conceived of as a logical relationship between experimental results and knowledge claims. On this understanding of validity, an experiment or test is valid if it supports the conclusions that are drawn from its results. I will, however, in this chapter, understand the term "validity" as a feature that can be ascribed exclusively to interpretative claims.[1]

To develop a foundation for my own account of validity, I want to begin with the ordinary language definition of validity. On the OED definition, *validity* is ascribed to claims or assertions that are "well-founded on fact or sound principles." An additional qualification is that such assertions are "thoroughly applicable to the case and circumstances" to which they are applied. This definition suggests that validity involves a relationship between two things: a foundation or a basis

[1]I include this distinction at the suggestion of John Norton, who rightly pointed out that I should acknowledge common uses of validity in philosophy before introducing my own notion of the concept.

("facts"), and that which (e.g., the conclusions) the foundation or basis supports. Second, when we say that a relationship between two things is valid, we are making a judgment as to how well those conclusions based on the facts apply to a given case or circumstances.[2]

What is worthy of note with respect to this definition of "validity" is how it differs from that of "reliability". With respect to reliability ascriptions our worry is about inconsistency in the future, despite similarities in the past; but with respect to ascriptions of validity, our worry is instead that either we lack a sound foundation for our claims (i.e., if the process by which those facts upon which we ground our claims is unreliable) and are therefore unentitled to lay claim to the validity of our assertions; or our assertions are not applicable given the contexts to which we apply them.

So, on my understanding, the OED definition provides a good starting point for beginning the process of sifting through the philosophical and scientific literature on validity. I want to say at the outset, that on this definition, "validity" will be a term best ascribed to a process such as *data interpretation* where a claim that pertains to an effect produced in a laboratory, which is substantiated by data, is extended and applied somewhere else. A determination of whether or not this extension is valid will require the identification of criteria for assessing the validity of a claim that seeks to conjoin two contexts and their contents.[3]

The term *validity* is rarely mentioned by name in the philosophy of science literature. Instead, focus on the nature of interpretation in science has been on the *generalizability* of scientific claims (e.g., Glymour (1980), Latour (1988), Cartwright (1983, 1999), Hacking (1983)) rather than on

[2]In some systems of philosophical logic, *validity* is defined as a feature of arguments in which the conclusion or set of conclusions follows directly from a set of premises. Note that in neither of these cases is "validity" equated with "truth". Instead, if a claim z follows directly from a set of premises x and y, then that claim may be said to be "well-founded", on those premises, or to be a direct consequence of them even if neither the premises nor the conclusions are true.

[3]This idea is in principle contained in the work of the American psychologist Samuel Messick who worked for the Educational Testing Service in the United States. I encountered his work after I had arrived at my own conclusions about how we ought to understand validity, but my understanding is very similar to his.

the validity of those claims. As Francesco Guala (2003, 2005) has recently pointed out, because philosophers of science generally take science to be in the business of providing nomic generalizations, the issue of validity has been traditionally equated with that of generalizability. Guala (2003, 2005) has recently sought to generate interest in the issue of validity among philosophers of experiment by importing a distinction drawn by experimental psychologists (Campbell and Stanley 1963) between *internal* and *external validity*. While I will not here analyze Guala's account of validity, I take my project of informing analyses of experimentation in philosophy of science with conceptual tools available in the scientific literature to be in the same spirit as his project.

4.2.1 Validity in Experimental Psychology

I will begin this section by considering some general definitions of validity that have been offered throughout the history of psychology to determine what features, if any, they have in common and whether they amount to some overarching general conception. The difficulty here is separating out general definitions of validity from notions of validity that are very context specific (e.g., put forward with respect to intelligence testing or perceptual studies). However, definitions of validity have evolved over time and more recent definitions of validity better accomplish this task.[4] I will then go on to consider a subset of the different parameters of validity that have been identified historically by psychologists who have sought to alter the validity of their experiments in the context of experimental design. I will extract from this subset three parameters of validity that I take to comprise a good starting point for evaluating neurobiological experiments.

[4] I have in mind here specifically the work of the psychologist Donald Campbell, who, over the years has attempted independently and in collaboration with other scholars to isolate a working notion of validity that is generally applicable to all laboratory sciences.

4.2.2 Scientific Understandings of Validity

One of the first rigorous discussions of the issue of validity began to emerge in the 1940's in psychology (e.g., Jenkins 1946; Goodenough 1950; Guilford 1946) in the context of educational and intelligence testing in psychology. This discussion culminated in the publication of the American psychologists Lee Cronbach and Paul Meehl's (1955) paper on "construct validity"[5], one type of validity that had been identified out of a host of other types, but which prior to this point lacked a rigorous definition. In their paper, Cronbach and Meehl (1955) refer to "the validation" of psychological tests. By this they mean that when an investigator asks whether a test is valid, he is asking a question about the inferences that result from using that test. They claim:

> "In one sense it is naive to inquire "Is this test valid?" One does not validate a test, but only a principle for making inferences. If a test yields different types of inferences, some of them can be valid and others invalid." (Cronbach and Meehl 1955, 297)

So, on Cronbach and Meehl's account, validity is a feature that is ascribed to inferences that result from using a particular test. Yet, while Cronbach and Meehl tell us to what validity can be ascribed, they do not say explicitly what it is. On my interpretation they take "valid" to mean something like "true" and invalid to mean "false". This definition lacks a certain degree of precision, but moving forward in the history of psychology, definitions of validity become slightly more clear.

In their book, *Quasi-Experimentation*, Cook and Campbell (1979) "use the concepts of *validity* and *invalidity* to refer to the best available approximation to the truth or falsity of propositions". In so far as they take validity to be a feature best ascribed to propositions that are the result of experimentation, their account of validity does not differ substantially from that of Cronbach and Meehl. Where the two accounts differ is in the types of validity that are the targets of interest. I

[5] As I devote a section to construct validity below, I will not define it here

discuss these in detail in section 4.5 below.

The American psychologist Samuel Messick (1989) offers a far more detailed and context-specific definition of "validity" in claiming that it is:

> an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.

Notice that Messick's definition does not deviate far from the ordinary language definition of validity, yet it contains concerns that are relevant specifically to interpretations that are based on measurements undertaken in the context of intelligence testing. It is interpretations about intelligence based on such measurements, which can, according to Messick, be valid or invalid–and according to different degrees or dimensions. As he states:

> Validity always refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores.[6]

On Messick's account, ascriptions of validity will only be appropriate if they have adequate evidential support.

On a much more recent account of validity, William Shadish, Thomas Cook and Donald Campbell (2002) define validity in a way that closely resembles Messick's definition. They use the term "to refer to the approximate truth of an inference" (Shadish, Cook, Campbell (SCC) 2002, 34). So,

[6] As an aside, clearly assessments of the validity of the interpretations of test scores like the SAT and the GRE have a profound impact on the lives of students who want to get into a good college or go on to graduate school. Test scores dictate the quality of the institutions to which a student can enter. The student not only has to worry about the test designer's assessment of the validity of the inferences based on the test design (those inferences that inform the structure of the test and its contents), but also those assessments of the committees in charge of entrance evaluations. Recently the legitimacy of using test scores to infer aptitude for college entry has been put into question, because the questions are thought not to transcend cultures and educational differences that exist in the United States. As a result, some colleges do not take the scores as seriously as they used to, because of the worry that inferences about scholastic aptitude made on the basis of such scores could be invalid.

they take validity to be "a property of inferences" rather than "a property of designs or methods". More specifically they state:

> When we say something is valid, we make a judgment about the extent to which relevant evidence supports that inference being true or correct. [. . .] validity judgments are not absolute; various degrees of validity can be invoked. (SCC, 2002, 34)

So, similar to Messick, contained in their understanding of validity is the issue of degrees–that inferences can be more or less valid. The interesting question, then, is how assessments of more and less valid can be made, and this brings us to the issue of different types of validity that factor into making such assessments.

Before moving on to identifying and defining several types of validity, I want to criticize a general feature that is common to each of the aforementioned definitions of *validity*. There is a tendency in the scientific literature to fail to keep the concepts of reliability and validity distinct. I think this is due to the shared recognition that validity requires reliability. Yet, I attribute the conflation of the two terms to the result of a failure to acknowledge that assessments of validity require two separate kinds of judgments. The first is a judgment about the reliability of the data production process in general. The second is a judgment about the legitimacy of applying a claim arrived at in one context to another context. Assessments of legitimacy have two parts; we not only require of a claim that it was produced by a reliable process, but we also want to be certain that their is adequate "translatability" between the two contexts, with respect to the claim. Yet, if the context of the lab and the context to which the claim is applied are not identical, wholly dissimilar or even slightly dissimilar, then the second condition required for legitimacy, i.e., similarity or identity between the context (or its aspects) in which a claim is produced and that to which it is applied, is not met. While I think this is implicit in general definitions of validity–or it is at least

what psychologists want to mean, as a result of the emphasis on the importance of reliability in science, I think it is often obscured. One advantage of my general account of validity is that it keeps these two conditions conceptually distinct.

This brings me to the aim of the next section. If we want to determine the similarity between a laboratory and its contents and the world and its contents, or an identity relationship between something under study in the lab and some phenomenon in the world, we have to engage in comparative analysis. Yet each context has complex dimensions, and the assessment of the validity of an interpretive claim that connects the two contexts will require us to engage in analyses across multiple relevant dimensions. It would be helpful to begin with some sort of foundation for such analysis. Such foundations may be located in the history of psychology and contemporary theoretical work on experimental design. In these contexts, different types of validity have been itemized; each type pertaining to a specific set of aspects of experiments. This list includes the following types: "intrinsic validity" (Gulliksen 1950), "content validity", "concurrent validity" (Cronbach and Meehl 1955), "statistical conclusion validity" (Cook and Campbell (1979), "ecological validity" (Brunswick 1956, Lewin 1943, Gibson 1941, 1960), "construct validity" (Cronbach and Meehl 1955), convergent and discriminant validity in addition to "internal" and "external" validity.

4.3 THREE TYPES OF VALIDITY

In this section, I will only consider definitions of those types of validity that relate directly to the issues that I am interested in with respect to neurobiological experiments. So, in what follows, I consider only a subset of analyses that have been provided for three types of validity: *ecological validity*, *external validity*, and *construct validity*. One problem with which I am faced here, is that

these different types of validity do not necessarily map directly onto the neurobiological experiments. This is because the meaning of these concepts has evolved in particular areas of science having experiments that differ from those found in neurobiology. So my additional aim here is to clarify the types of validity with respect to my project and to explain how they are relevant.

4.3.1 Ecological Validity

A primary context in which concerns about *validity* arose historically in experimental psychology was in a debate between the Hungarian-born psychologist Egon Brunswik (1943) and the German-born psychologist Kurt Lewin (1943). At issue was the dissimilarity between the environmental context of the laboratory and the world outside the lab and the potential negative impact that studying effects (sp., perception) in isolated contexts had on the scope of psychological explanations. This type of validity came to be known as "ecological validity" (Bronfenbrenner 1979) and it instigated "ecological approaches" to the experimental study of perception in psychology (e.g., Brunswick 1949; Gibson 1960, 1979). Although different authors since this time have provided different interpretations of the concept of ecological validity (e.g., Bronfenbrenner 1977, 1979), some even equating it with "external validity" (e.g., Cook and Campbell 1979), in every treatment it concerns essentially some type of relationship between those artificial experimental contexts and their features (e.g., presented stimulus patterns, size, responses required of an organism) and the natural environments in which those effects typically occur. It may also involve the relationship between how an experimenter himself understands or interprets the relationship between the experimental context and the world.

Lewin and Brunswick were interested in developing methods that were appropriate for the study of perception in psychology. Lewin (1943) defined one such approach, the *field theoretical*

approach, as: "a method of analyzing causal relations and of building scientific constructs" (Lewin 1943, 294).[7] He and other field theorists defined a *field* as simply the environment of an organism. However, there are various options for conceptually (sp., temporally and spatially) constraining or widening that environment (or "field"). For example, a field could include the environment proximal to a subject, the environment distal from the subject, or both. Furthermore, that environment could be temporally circumscribed in addition to spatially circumscribed. Lewin, wanting to secure for psychology the kind of rigor of the constructs (the idea of a "field") that he took to be the hallmark of physics, conceived of a field as temporally and spatially isolated vis-a-vis the "life space" of an organism. A life space consisted of the environment proximal to the organism in a situation S at a time t ($dx/dt = F(St)$ (where F= field)). In other words, "a field at a given time" was defined exclusively in the experimental context relative to the immediate and proximal situation in which an experimental subject was placed, without regard for other variables that may impact that subject's perception at that given moment. In essence, the concept of a field excluded "all those processes in the physical and social world" that might effect the life space of an individual at any given time (Lewin 1943, 306). So the concept of a "field at a given time" essentially abstracted away from both (1) the actual complex environment of the laboratory in which an organism was placed and (2) of the actual complex environment of an organism perceiving in the natural world.

Brunswick (1943) took issue with Lewin's, what he termed, "molar approach" (262) to addressing psychological questions (e.g. about perception and intelligence), deeming it restrictive in terms of the environmental setting captured by the "field at a given time" construct. For Brunswick, the natural environmental setting of an organism contains a whole host of variables such as: illumination, background and foreground, object size, color, distal as well as proximal objects–aspects of

[7] By "construct" Lewin means simply a theoretical concept. I will come back to this point when I discuss the issue of construct validity, below in Section 4.4.3

an organism's environment not captured by Lewin's "field at a given time" construct. Brunswick took the ramifications of the invalidity of Lewin's construct–i.e., its failure to capture the true ecological circumstance– to be severe with respect to perception experiments. Specifically, he claimed that the construct invalidity that may result from positing a "field" could result in drawing false conclusions about, for example, the causal variables present in the context of an experiment that play a role in the production of the effect under study.[8] Furthermore, it would leave psychology in the business of solving "narrow-spanning problems of artificially isolated proximal or peripheral technicalities of mediation which are not representative of larger patterns of life" (Brunswick 1943, 262).

So there are two aspects of ecological validity that are apparent in Brunswick's critique of Lewin. First, an investigator should worry about whether or not a construct captures something true vis-á-vis the world outside of the laboratory. Second, an investigator should worry that he/she might fail to represent, as part of their constructs, variables that are relevant to the production of the phenomenon that they are trying to study in the lab. However, this second type of construct failure is secondary and a direct result of an investigator abstracting away from the natural environment in which the organism resides and behaves, neglecting features of that environment to the extent that

[8]Brunswick (1943) proved this point himself by following a human subject around one day and documenting her response to various objects in her environment, which suggested that she paid more attention to distal as opposed to proximal stimuli, which flew in the face of a major assumption upon which both perceptual psychology and stimulus-response psychology were based. Furthermore, you can find the incorporation of more distal stimuli in more modern psychological and neuroscientific experiments. For example, most of the mazes that are used to study "spatial learning" in the rat capitalize on a rat's perception of distal stimuli (e.g., objects on the wall surrounding the maze) relative to the maze. Also this attention to distal stimuli has resulted in such experiments normally being conducted in environments that are barren except for those distal stimuli, to avoid the possibility that the rat would pay attention to distal stimuli other than those designated by the experimenter. So in some cases, shedding light on the "ecology" of an organism does not lead in the direction of the kind of validity Brunswick wanted to secure for psychology. It merely feeds into how to make experiments more reliable.

they do not factor into the conception of the laboratory environment. [9]

Both Brunswick (1947) and J.J. Gibson (1960, 1979) went on to develop elaborate, what they took to be *ecologically valid* approaches to the study of perception and other psychological phenomena that contained ecologically valid constructs. The strategies that I provide for striking a balance between reliability and validity with respect to neurobiological experiments on learning are in the same spirit as the kinds of solutions that they provided for the study of perception.

I want to consider a more recent treatment of ecological validity in the scientific literature that analyzes the concept into a variety of different dimensions. Schmuckler (2001) provides a brief historical review of the literature on *ecological validity* and indicates that concerns about it remain prevalent in modern cognitive psychology. However, he suggests that over time the meaning of the concept has been lost, primarily because psychological experiments contain different elements (e.g., stimuli, subjects with histories, environments), and psychologists who have been concerned with ecological validity have attended to the issue with respect to different elements of these experiments instead of the whole group at once. Today, Schmuckler claims, ecological validity is a broad category that can be viewed as having multiple dimensions related to the nature of the setting, stimuli and response of an experiment. He itemizes a set of different dimensions with which an investigator must concern himself. These dimensions fall into roughly two distinct categories. First, an experimentalist ought to be privy to differences and similarities between his abstract representation of the laboratory environment and how the organism perceives and experiences it and *actual* way in which the organism perceives that environment, i.e., the organism's representation of that environment. Second, he should understand how the artificial setting and circumstances of the laboratory differ from or are similar to the actual circumstances in the world. If the experi-

[9] In other words, there are two contexts, the context of the lab and the context of the natural world, and the complexity of both may be overlooked by the experimentalist.

menter is truly worried about the validity of his interpretive claims along ecological lines, he will attempt to maintain a balance between replicating the real-life situation in the lab while capturing the nature of the social and cultural context in which that situation is found in the world and ensure that the laboratory environment contains relevant features of naturalistic settings. Furthermore he will attempt to determine whether stimuli used in an experiment "are representative, natural and stable occurrences in the world" and whether they "continue to be relevant when removed from that world and placed into an artificial setting" (Schmuckler 2001 422-423). He should also think about whether a task that an organism engages in the experimental context is natural, and how it compares to other tasks that the organism may engage in, in its natural environment (e.g., lever pressing), and what the implications of such differences are. In addition, he should be conscientious that his description of the variables intervening between stimulus and response may not map directly onto the actual intervening variables present. As Schmuckler rightly points out, given the complex dimensions of ecological validity, one of the most difficult tasks faced by the experimentalist is which dimension should take precedence over the others; on which factor one should focus on first.

I have described validity as a feature of interpretive claims. So when we ask of an interpretive claim whether or not it is ecologically valid, we are asking if the ecological features between the lab and the world are similar such that we can legitimately apply a claim arrived at in one context to the other. A comprehensive analysis of the ecological validity of a claim would include an analysis of all of the aforementioned dimensions. Such a comprehensive analysis of the neurobiological experiments at issue in this dissertation far exceeds its current scope, so I will be forced to focus my attention on a subset of these dimensions.[10] I will engage in such an analysis of interpretive

[10]This is one interesting direction for future research.

claims in neurobiology in Section 4.4.2. First, I want to consider two other parameters of validity that are related to but different from ecological validity.

4.3.2 External Validity

Cook and Campbell (1979) define *external validity* as "the approximate validity with which conclusions are drawn about the generalizability of a causal relationship to and across populations of persons, settings, and times" (Cook and Campbell 1979, 39). This definition is essentially identical to that provided by Shadish, Cook and Campbell, 23 years later (2002, 83). A primary focus of both works is the issue of validity with respect to experiments run in human populations. A classic example that instantiates concerns about external validity is a study published in 1954 aimed at determining a causal link between smoking and lung cancer in middle-aged white men from 9 states (Hammond and Horn 1954). In evaluating the external validity of that study, one is asking whether the causal relationship found to exist in the study group holds true for other populations (e.g., women, adolescents, African Americans, populations in other countries), in other settings (e.g., individual-related contexts–settings in which subjects lived for extended periods (e.g., with heavy air pollution))[11], and at other times (e.g., will the result be valid with respect to future populations). For this study, and similar types of studies, one can imagine a variety of complex features of persons, environmental settings, and times that could potentially "impact" on the validity of the

[11] It should be noted here that there tends to be some overlap between definitions of external validity and definitions of ecological validity, but clearly given the kinds of things itemized in the context of ecological validity it is not identical to the concept of external validity, despite the fact that the extension of a causal claim beyond the laboratory will require that certain conditions of ecological validity be met. But, for example, the experimenter's representation of how an organism perceives a lab environment and how that organism actually perceives that environment is an issue of validity that is with respect to the relationship between the lab environment and an abstract model rather than the lab and the world. So, since there are these kinds of exceptions, I take these categories of validity to be distinct, as experimentalists throughout the history of psychology have.

extension of a causal claim beyond that experimental context in which it was detected. So, this understanding of external validity in terms of pointing at the dimensions across which we should investigate the generalizability of external claims beyond the laboratory is quite helpful for directing my analysis of neurobiological experiments. Clearly the specific details that will emerge in such analyses will be idiosyncratic to a given study, but Cook and Campbell (1979) and Shadish, Cook and Campbell (2002) provide a good foundation in proposing broad general guidelines. I now want to look at one last type of validity that also has been of interest to psychologists since the 1940's: *construct validity*.

4.3.3 Construct Validity

The category of "construct validity" generated interest among American experimental psychologists in the post-WWII era at a time when it became fashionable to scientifically study abstract features of the population, such as *intelligence*, *scholastic aptitude*, *mental depravity* and *mental illness*. As concerns mounted as to whether test designs actually captured the features they were intended to capture, the psychologists Lee Cronbach and Paul Meehl (1955) were called upon to provide an account of the corresponding concept of "construct validity" and the challenges that investigators faced in the so-called "validating" of constructs. As a result, their account is the first and most comprehensive treatment of "construct validity" in the literature. Here, I am interested primarily in how they define "constructs" and their ideas about how constructs come to be, in their words, "validated".

Cronbach and Meehl define a *construct* as "a postulated attribute of people, assumed to be reflected in test performance" or which can "account for" test performance. (Cronbach and Meehl 1955, 283, 282). Such attributes are ascribed to organisms, but cannot be directly measured. How-

ever, an investigator can use indirect measures that he takes to be representative of a construct to "capture" that construct. For example, an investigator may design an IQ test with what she takes to be a representative set of questions that probe a person's general intelligence. She will then assign a point value to each question and then decide on scores that she takes to reflect such categories as "giftedness", "high intelligence", and "low intelligence". In using the test to determine intelligence, she is assuming that test performance *is the result of* a person's intelligence–that intelligence is the causal variable responsible for test performance. The problem of construct validity arises, then, when the assumption is shown to be problematic. It may be problematic for two reasons. First, the variable that the test actually measures may not be intelligence but some other feature.[12] Second, intelligence itself may not be causally responsible for performance on the test.

One aim of Cronbach and Meehl's (1955) article is to explain why the problem of construct validity arises at all. They claim that it results primarily from the fact that constructs are vaguely defined theoretical categories. They are theoretical postulates that are supposed to causally account for a set of data and to be "reasonably" related to a 'real world' concept (e.g., intelligence). When an attribute cannot be measured directly, an investigator first will attempt to define it generally and then will try to identify the conglomeration of features or behavioral traits that they take to comprise it (e.g., anxiety may include increased heart rate or blood pressure, proneness to ulcers). They point out that this results in two distinct kinds of problems. First, a construct when it is defined generally might miss certain features that ought to be included in the construct category and investigators are typically not in agreement about what features should be included, nor is there an objective way of resolving such disagreements. Secondly, it may be that the suggested means of measurement may not be an *adequate* measure of construct at all–it may only capture

[12]This interpretation would ultimately be linked to "test" validity–namely, does the test measure what it is supposed to measure.

certain features rather than all of the features that have been ascribed by the experimenter herself to the construct. One example that Cronbach and Meehl provide to illustrate this latter problem is the instance in which an intelligence test reveals intelligence differences that appear to correspond to cultural differences rather than to some "culture-free" construct of intelligence. This raises the question of whether such tests are valid with respect to intelligence generally or are measuring cultural differences in education levels. "Construct validity", when ascribed to tests, then, is taken to mean something like "accuracy to the construct" on Cronbach and Meehl's account.

Given how vague constructs in general are, Cronbach and Meehl admit that a variety of variables may impact the assessment of the validity of a test for accurately capturing a construct. They point out that when the Binet intelligence scale came out, it was determined valid in part because the results coincided with teacher's assessments of student's intelligence. However, by the 1950's, the Binet test was taken as the only valid indication of intelligence. One sees this in other contexts as well, that confidence in the validity of one test may alter views about the validity of another if their assessments do or do not coincide.[13]

Another problem that results from the fact that constructs are "open-ended" concepts is that they are, in the context of the broader theoretical networks of concepts in a given area of science, always in the process of validation:

> if the network [of concepts] is very incomplete, having many strands missing entirely and some constructs tied in only by tenuous threads, then the "implicit definition" of these constructs is disturbingly loose; one might say that the meaning of the constructs is *underdetermined*. Since the meaning of theoretical constructs is set forth by stating the laws in which they occur, our incomplete knowledge of the laws of nature produces a vagueness in our constructs. And if it is the case that we are still in the process of discovering the laws that involve a given construct, then we do not

[13]This pertains to the issues of "concurrent validity" and "discriminant validity". An example of concurrent validity on Cronbach and Meehl's account would be two tests converging on the same construct, whereas discriminant validity would be the opposite idea–that two tests diverge. Both types of assessments could impact views about the construct validity of the other test.

yet know precisely what that construct is. (Cronbach and Meehl, 1955, 294)

A final feature of Cronbach and Meehl's account is that they take assessments of construct validity to apply variously to such things as tests, measurements themselves as well as interpretations of data. I, on the other hand think that we must investigate features of tests and measurements when we are assessing the validity of interpretive claims that are made on the basis of data obtained from these tests. Clearly, we do not want to say of an interpretive claim that it can be construct valid. However, when we ask whether an interpretive claim is valid, we are concerned whether there is an identity relationship or not between what is purportedly measured by a test and that abstract feature of a person that the test is supposed to capture. Similarly, we want to know if the causal variable taken to be responsible for test performance is the actual causal variable that is responsible. Such assessments will necessarily be problematic given the vagueness of the construct category and disagreements across researchers in terms of how to measure abstract features of persons like intelligence.

I want to consider another notion of construct validity that extends the concept in a way relevant for my aims with respect to understanding neurobiological experiments. Shadish, Campbell and Cook (2002) claim that constructs can be any kind of category or concept that refers to some unit included in an experiment, such as "people", "treatments", "outcomes", and "settings" (SCC 2002, 69). In addition, there may also be a construct–a theoretical representation of the relationships that bear between all of these elements of the laboratory (SCC 2002, 72). Then, on their account, construct validity will "involve making inferences from the sampling particulars of a study to the higher-order constructs they represent (SCC 2002, 65, 70)." As an example they point out that a study that includes schizophrenic subjects may include subjects who were diagnosed by a psychiatrist referencing the *Diagnostic and Statistical Manual of Mental Disorders* (1994) or a

Schizophrenia subscale. These subjects would be considered different types of schizophrenics, which may impede extending claims about them to the schizophrenic population generally, when that population may be comprised of both types of subjects. The types of settings investigated in a study, may also not map onto the concept of those settings outside the laboratory.[14]

Shadish, Cook and Campbell's treatment of construct validity obscures the original understanding provided by Cronbach and Meehl. Yet, there is something interesting in their notion that I would like, for the purposes of the subsequent discussion, to develop. If in our analyses of the validity of an interpretive claim we could investigate the relationships between the elements of an experiment (organisms, effect under study) and the actual models or concepts of those elements, then assessments of construct validity will be far more complex than what Cronbach and Meehl suggest. The category will capture relationships between the laboratory and what is outside it, including the theoretical concepts and models that drive research, in a way that none of the aforementioned types of validity that I mentioned captures. Either we should broaden the notion of construct validity in the way suggested by Shadish, Cook and Campbell or we should develop a new category of validity entirely. Here, I will do the former. From hereon in, when I use the term "construct validity" I will have the broader notion in mind.[15]

[14]While this differs from what was originally meant by Lewin (1943) with respect to "ecological validity", it is clearly related.

[15]One aim of this dissertation is to get a handle on what validity is and what some of its dimensions are–those dimensions that correspond to those aspects of experiments with which I am most interested. However, this means that what I offer in this dissertation is clearly only a first pass. These issues deserve a far more detailed treatment, and that will be the aim of a future project.

4.4 VALIDITY: A PARAMETRIC ACCOUNT

Based on the aforementioned section, there are at least three types of validity with which an investigator should be concerned in the data interpretation stage of the experimental process: (1) ecological, (2) external and (3) construct validity. These three types of validity are represented in *Figure 8*. Underneath each type of validity, I have itemized a subset of those questions that capture what I will refer to as the *parameters of validity*, which will be assessed during data interpretation. In order to explain how such parameters come to bear on the experimental process with respect to analyzing the validity of interpretative claims, I turn now to the task of presenting my account of the data interpretation process, which is represented in *Figure 10*.

4.4.1 Validity and Data Interpretation: General Account

On the account of the data production process that I provided in the previous chapter, if that process is reliable overall, it enables the discrimination of one claim from a set of competing claims about an effect produced in the laboratory. According to my account, the data interpretation stage of the experimental process involves two discrete steps. The first step is simply a reflexive step, namely, to take that claim that has been indicated as true of the effect produced in the lab via the data production process and to *interpret* it as true of that effect as it is produced in the lab. The second step is to extend the claim beyond the context of the laboratory, to apply it to the original phenomenon of interest in the natural world, and to take it as true of that phenomenon.

As I mentioned briefly at an earlier point in this chapter, two conditions must be met in order for an interpretive claim to be deemed valid. First, that claim must be the outcome of a reliable data production process. Second, since interpretive claims in science often relate two contexts, namely

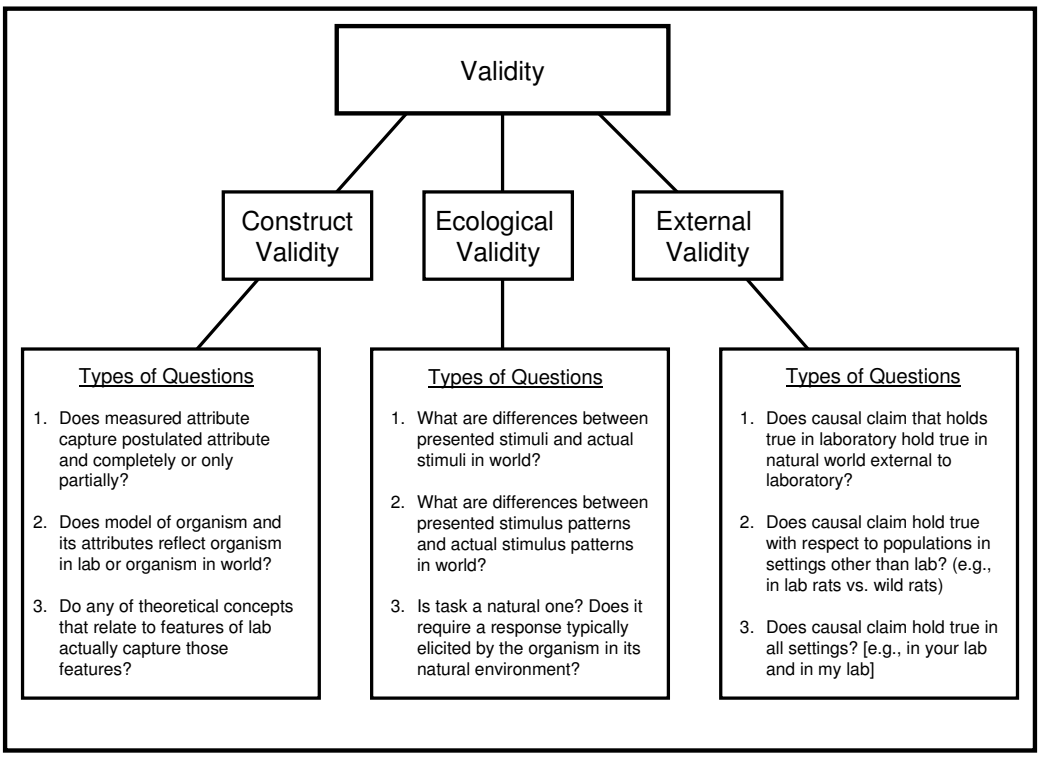


Figure 8: Validity: Taxonomy of Partial Types

the context out of which a claim is produced and that context to which it is applied, the extension of the claim from the one context to the other must be *appropriate*. Assessments of "appropriateness" hinge on the two contexts and their aspects being *sufficiently* similar (ecological, external validity) or identical (construct validity). The sufficiency of the similarity or identity is to be determined by comparing the two contexts or the variable measured in the laboratory to one in the world across relevant *parameters of validity*. The more similar the contexts are across the various identified parameters, then the more likely it is that the interpretive claim will be valid, or valid across some parameters and not others.

What these conditions entail is that reliability can occur without validity, but not vice versa. Validity is contingent on both the reliability of the data production process and a tight similarity between the two contexts brought together by means of an interpretive claim. Or it is contingent on the identity between what a test measures and what it is supposed to measure; whether the proposed variable is the cause of the observed effect or not in the lab and in the world. With respect to the application of interpretive claims beyond the context of the laboratory, the condition of reliability may be met, but the dissimilarity between the lab and its elements and the world and its components may make the extension of the interpretive claim inappropriate or illegitimate. I will discuss the relationship between reliability and validity in more detail in the next chapter.

4.4.2 Application of Conceptual Framework to Case Study

With respect to the Atkins et al. (1998) study that was the focus of the last chapter, the primary claim is that "mitogen-activated protein kinase (MAPK) activation is necessary for contextual and cue fear conditioning" (Atkins et al. 1998, 606) and the broader claim is that "MAPK is necessary for the consolidation of associative memories in the mammalian nervous system" (Atkins et al.

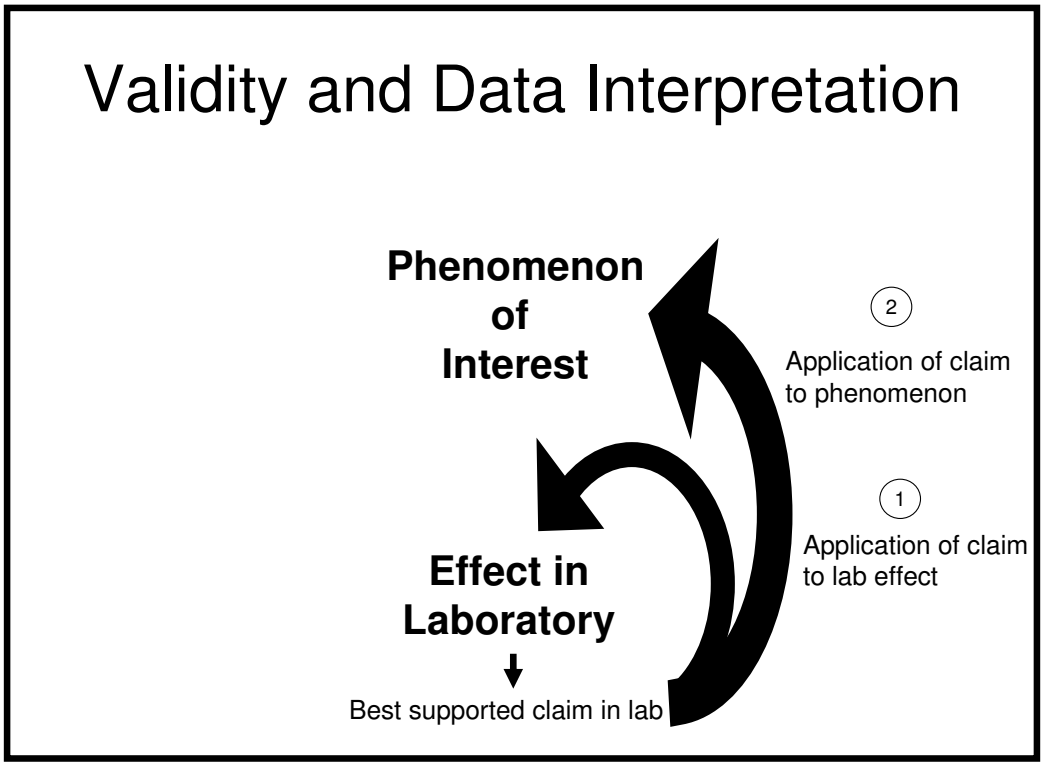


Figure 9: Validity and Data Interpretation

1998, 607).

The general structure and presentation of these claims about the role of the MAPK molecule in learning instantiates what I take to be the steps involved in the data interpretation stage of the experimental process (please refer to *Figure 9* and *Figure 10*). First, the claim arrived at via a (presumed) reliable data production process is applied to the effect produced in the laboratory. So, in this case, the claim is that the molecule of interest, MAPK, is involved in the forms of learning under study in the laboratory, namely contextual and cue fear conditioning, in the rat subjects used in the study.[16] Next, that same claim is *extended beyond* the context of the laboratory to what I take to be the original phenomenon of interest in the natural world–namely: *mammalian associative learning.*[17] This claim can be interpreted to have at least two distinct meanings (captured in *Figure 10*). First, activation of MAPK is requisite for associative learning in rats generally. Second, activation of MAPK is requisite for associative learning in all *mammals*, including human beings. We can assess the validity of each of these claims with respect to the two conditions I have specified for validity and those parameters of validity that I have identified (*Figure 8*). This is my aim in the next section.

4.4.3 Investigating Validity Across Multiple Parameters

To assess the validity of the first interpretive claim arising out of the Atkins et al. (1998) study, namely the reflexive extension of the claim to the effect produced in the laboratory, we need to first ask whether the data production process was indeed reliable. The immediate aim of the study was

[16] I just want to assume for the sake of argument here that this is the basic claim, even though the claim lacks a precise referent and thus could be taken to mean that *all* forms of contextual and cued fear conditioning in *all* mammals require activation of MAPK.

[17] This is what I take them to mean by "consolidation of associative memories in the mammalian system", given the title of their paper, which is "The MAPK cascade is required for mammalian associative learning" (Atkins et al. 1998, 602).

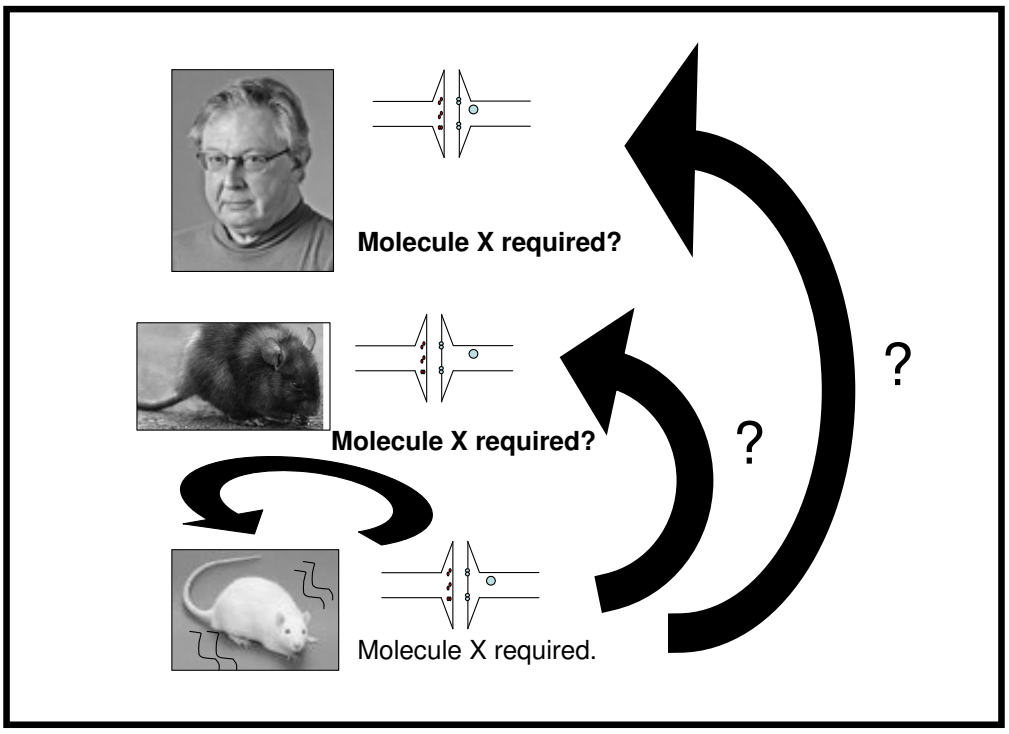


Figure 10: Validity and Extrapolation

to determine whether activation of MAPK was necessary for contextual and cued fear conditioning in laboratory rats. In Chapter 3.0, I identified each measure taken in both the design and implementation stages to ensure the reliability of the data production process in cued fear conditioning experiments. I want to assume for the sake of argument that the entire data production process was reliable and indicated a causal role for activation of MAPK in cued fear conditioning in rat subjects.[18] The first question then is: (1) Is the extension of the causal claim to populations of *rats* in the natural world *valid*?

Providing an answer to this question in part requires that we engage in a similarity/identity analysis of the two contexts and their features across the suggested parameters identified in *Figure 8*. I only want to consider a subset of these parameters. The first question I want to ask pertains to construct validity. Does the measured form of learning, namely, cued fear conditioning, capture the postulated attribute of associative learning completely or only partially? To this question, I think we are forced to answer that whatever associative learning is in the broad sense, cued fear conditioning is only one type of that broader category, and the cellular and molecular mechanisms that are likely to be involved in one form of associative learning, are probably unlikely to be the same ones involved in all forms.[19]. A second question pertains to differences between laboratory rats and rats in the world. Laboratory rats lack a rich behavioral history. A rich behavioral history might indicate experiential differences that may contribute to learning. Freezing responses may be a byproduct of the closed environment of the laboratory. A third question pertains to stimulus patterns in the world, which do not necessarily reflect those patterns of stimuli that are presented in the context of classical conditioning paradigms. So, there might not be any real-world correlate to

[18]The only reason to doubt the *internal validity* of the interpretive claim is if there is a discrepancy in the way that the investigator understands the causal relationships holding between elements in the laboratory and what the relationships between those elements actually are. (Schmuckler takes this to be one dimension of ecological validity.)

[19]That this is acknowledged to be true by some neurobiologists is captured in Sweatt's taxonomy in *Figure 11*.

which such changes pertain.[20] Also, it is probably an infrequent occurrence that rats get shocked in the world.

A fourth question we may ask pertains to the pattern of stimuli used to produce learning in laboratory rats and the pattern of stimuli that a rat may encounter in the natural world. There is an artificiality to the stimulus patterns in the cued-fear conditioning paradigm in part because the aim is to produce a measurable increase in MAPK. However, it is unlikely that natural stimulus patterns in the world would occur at the same rate and have the same duration in such a limited time frame. We might anticipate, in light of such differences that there would be molecular differences between the two cases as well. A final related point pertains to whether there is anything like cued fear conditioning in the world and if there is, whether the form of it studied in the laboratory is at all similar to the phenomenon as it occurs in the world.

If there are such differences between the case of the lab rat and a rat born and raised in a natural environment, then one can anticipate the extreme difficulty that we confront when we ask about the validity of the second and stronger claim made by Atkins et al.: Activation of MAPK is necessary for mammalian associative learning more generally. Yet, the way that some neurobiologists speak about the ease of using model systems to understand the molecular mechanisms of learning and memory in human beings leads one to think that the validity of such claims is not at stake. Yet, we may ask the same kind of questions as those identified above with respect to the lab rat and the human being. Is there something like cued conditioning in human beings? Does it require similar stimulus patterns in order for it to occur? Will the same brain structures be involved, and the same molecules? Will it present in the same way that it presents in laboratory rats or will the types of responses elicited differ? (For additional problems that arise at the molecular level with respect to

[20] A neurobiologist might admit this, but then, how would the causal claim be tailored to accommodate such differences?

extrapolation from model systems to human beings see Ankeny 2000 and Schaffner 2001. For an alternative view see Weber 2001). [21]

4.4.4 Building Models for Validity Assessments

Notice that in essence what we have to do to answer questions about the validity of interpretive claims that connect two different contexts is develop working models of the two contexts and their various elements. In other words, we have to build a model of an organism in its natural environment, the features of that environment, how the organism behaves in that environment, what happens at the various levels of organization in response to environmental stimuli, what form learning takes. To some extent these are the questions that are asked prior to running an experiment. But they reemerge when we attempt to take claims arrived at in the laboratory back to the world. Whereas when an experiment is being run these questions are asked for the sake of designing an experiment, in the interpretive stage the aim is a critical one. That is the prescription of validity at the stage of data interpretation: be critical about the relationship between the lab and the world when extending an interpretive claim.

The fact that the assessments of the validity of interpretive claims require model building is in essence why I want to keep some notion of construct validity distinct from ecological validity. I discuss in detail what I mean by "model building" in Chapter 6.0. It may be that in our models we miss certain features of the world or certain features of the lab and this can disrupt our ability to

[21] Notice that the types of validity and the questions that represent different parameters may either guide an analysis of validity or be used to situate questions that arise as an investigator engages in the analytic process of itemizing the parameters of validity assessed when considering the validity of an interpretive claim. I think that this taxonomy will constantly be revised in light of the different kinds of questions that each individual study will present for an investigator. So I look at the taxonomy that I have extracted from experimental and cognitive psychology as a rough guide. It does not have to be a starting point, but some such sort of taxonomy and itemization of parameters should emerge when an investigator adequately engages in this process. I will return to this point in Chapter 6.0.

make objective validity assessments, which is something an experimentalist needs to be conscious of, too. In particular if our models are informed by our pre-experimental commitments we may "misrepresent" the lab and/or the world.

On a final note, I want to point out that the claims I have made with respect to the Atkins et al. (1998) study are not unique to that study. This is the common form in which interpretive claims are formulated in the neurobiological literature. I have focused on this study in particular because it is one with which I am familiar, it combines the methods of both synapse-level plasticity experiments and organism-level learning experiments and the principal investigator, J.D. Sweatt, is concerned in his more theoretical work with broader issues about experimentation on learning in contemporary neurobiology. For a lot of reasons, it made good sense to analyze this study. I want to add that neurobiology is certainly not the only field in which the issues of validity have not received adequate treatment, it is a common feature of other areas of science as well. It is in part because neurobiology is a newer field, that discussions of the validity of experiment have not yet emerged.[22]

4.5 CONCLUSION

In this chapter, I have suggested that two specific conditions must be met in order for an interpretive claim to be determined valid. First, the claim must have arisen from a reliable data production process. Second, the context out of which the claim originates must be *sufficiently similar* to the context to which it is applied in order to be deemed *legitimately applicable* to that other context.

[22] But I take this to be surprising, given the critical history of classical and operant conditioning in experimental psychology in the latter part of the 20th century.

If the two contexts are significantly dissimilar then it is unlikely that the claim will satisfy the second condition of validity. The prescription of validity then, which I will consider in the subsequent chapters, is to increase the similarity between the two contexts: namely the context of the laboratory and that of the natural world. In the case of construct validity, the prescription is to be certain that the postulated causal variable is actually responsible for the observed effect in both the laboratory context and the external world.

With respect to the second condition of validity, I identified a simple taxonomy of different types of validity. Within each category I identified a set of questions that I took to correspond to different parameters of validity. This taxonomy is at best a first approximation to a broader taxonomy with more refined kinds. Each parameter corresponds to features of the laboratory and features of the natural world, or representations of the features of the two contexts. I claimed that assessments of the validity of interpretive claims should proceed across these different parameters of validity, and that if at any point the differences between the two contexts was determined to be significant with respect to one of the parameters, then the interpretive claim should be deemed invalid. Finally, I suggested that adequate and detailed parametric analyses of those contexts that interpretive claims bring together involve the building of models of those contexts and their features.

Up to this point in the dissertation I have claimed that reliability constrains data production processes whereas concerns about validity often do not arise until the stage of data interpretation. Yet, what I have suggested with respect to neurobiology is that concerns about validity with respect to applicability of interpretive claims to contexts outside the laboratory do not arise in either stage of the experimental process. In the next chapter, I will consider what I take to be the serious and negative implications of this lack of concern for validity for the entire field of the contemporary

neurobiology of learning and memory. Specifically, I will show how ensuring reliability may only lead to very uninteresting claims about learning and its mechanisms. The aim of the penultimate chapter, then, will be to use the parametric analysis of validity offered in this chapter to bring back validity to neurobiological experiments on learning.

5.0 CONFLICT BETWEEN RELIABILITY AND VALIDITY OF EXPERIMENT

5.1 INTRODUCTION

In Chapters 3.0 and 4.0 I claimed that the experimental process can be divided into two stages: (1) a data production stage and (2) a data interpretation stage. The data production stage consists of both (a) a design stage, in which all features of the sets of experiments that comprise a research study (e.g., operationalized concepts, models, techniques, methods, protocol) are articulated and (b) an implementation stage in which those steps laid out in the design stages are instantiated in individual experiments. The data production stage is followed by the *data interpretation* stage, which consists of two steps (a) the application of a claim arrived at via the data production process to the effect produced in a laboratory and (b) the application of that same claim to the original effect of interest, which is located somewhere outside of the laboratory. Despite the fact that data production is (or ought to be) temporally prior to data interpretation, it is not uncommon for concerns about data interpretation to influence the development of experimental designs, and so play a role in the design stage.

In Chapter 3.0, I claimed that *reliability* is a feature best ascribed to the process of data production; that in science investigators aim to ensure the reliability of the designs and implementation of their experiments, primarily because only a reliable data production process will yield data suf-

ficient to discriminate a true claim from a set of competing claims about an effect produced in the laboratory. I considered two types of neurobiological experiments: organism-level learning and synaptic-level plasticity experiments, and I provided concrete examples from these experiments in which a given investigator typically takes steps to ensure the reliability of the data production process. I suggested that concerns about reliability in such neurobiological experiments began and ended with the data production process. And, whereas the data production process instantiates those features necessary for ascriptions of reliability (e.g., one being "repeatability of the process"), the process of data interpretation does not. In turn, reliability will be silent with respect to the process of data interpretation–it is not operative in this process, nor should it be ascribed there.

However, in light of the fact that I had divided the experimental process into two stages, the second stage being data interpretation, I suggested that some criterion must be operative in this process–some criterion other than reliability. I pointed out that I was not alone in thinking so, that in the history of experimental psychology various experimentalists (e.g., Brunswick 1943; Lewin 1943; Cook and Campbell 1979) have been concerned with another feature of experimentation in science that differs from reliability, seems to be (at least for some) equally as important, and is discussed most commonly in contexts concerning how far beyond the laboratory scientific claims extend, namely, *validity*.

In Chapter 4.0, I suggested that ascriptions of validity are most appropriately made to claims that attempt to establish a relationship between a foundation (e.g., one or more substantiated claims) and the context or circumstances to which that claim is applied. Interpretations exhibit these features in so far as they are used to establish such relationships between a claim discriminated as true by data obtained from a data production process and effects that occur in specific contexts. Since, data interpretation involves two steps as I described above, validity can be as-

cribed at either step: an interpretation may be valid or invalid with respect to the effect produced in the laboratory, or valid or invalid with respect to the original effect of interest. The first type of validity best correlates with what I take experimentalists to mean when they talk about "internal validity". The second type is clearly what is meant by "external validity".

In Chapter 4.0, I identified three types of validity that have been identified in the literature: (1) *construct*, (2) *ecological*, and (3) *external* validity. I suggested, in agreement with Cook and Campbell (1979) that concerns about these three types of validity, if they are operative at all in the experimental process, ought to be operative in the design stage of data production when empirical questions and explanatory goals are conceptualized. However, the role of each type of validity was difficult to establish with respect to those organism-level learning and synaptic-level plasticity experiments that are my target case studies, primarily because none of the measures taken in the design stage were ones that satisfied the prescriptions attached to these three types of validity. In fact, even if some of the measures taken to ensure the reliability of these experiments suggested an awareness and appreciation of these validity types, that the prescriptions of these three types of validity were taken seriously in the data production process was not apparent in these experiments.

This leaves me now with several questions to answer: Why do concerns about reliability play such a prominent role in experimentation in neurobiology, whereas concerns about validity play such a limited role? Are they not both equally important to neurobiologists? Yet, based on the explanatory goals of neurobiologists should they not be equally as important? If reliability is taken to be more important, are there any specific consequences of this nearly exclusive commitment to reliability? What are they? The aim of this chapter is to provide answers to these questions.

Specifically, I aim to show by means of the same case studies that when reliability is pursued at all costs, or to a significant degree, *validity*, is sacrificed. In such cases, though the data production

process is constrained so as to guarantee that the one true or *most* true hypothetical mechanistic claim is discriminated, the generality of that claim, i.e., its application to the original effect of interest, is severely limited. It is not merely that these claims are not generalizable outside the laboratory; rather, it is not a legitimate move to try to apply them outside the laboratory.

An acknowledgement of the limited generality of claims emanating from the laboratory sciences is not novel in the philosophy of science. For example, both Cartwright (1983, 1999) and Hacking (1983, 1992) have emphasized this lack of generality, and have sought to pinpoint its source(s). Latour (1988) takes an even more radical stance on the generalizability of scientific claims, in suggesting that it is an irremediable predicament of science that it arrives at *local* as opposed to *global* truths. More recently, Guala (2003) has focused on a distinction between what he terms the *local* and *global* validity of scientific claims in order to determine whether the situation of laboratory science is really as bleak as Latour and others have suggested.

In this chapter I will focus on two consequences of insuring reliability. The first concerns those taxonomies of learning that we find in neurobiology and how they compare to other taxonomies of learning in neuroscience and in general to those types of learning captured by other experimental methods. The second and what I take to be more pressing issue is whether the causal mechanisms that are arrived at in the context of such experiments can be generalized to cases beyond the laboratory. I think the answer to both of these questions is no, and I will show it in this chapter.

5.2 CONFLICTING PRESCRIPTIONS OF RELIABILITY AND VALIDITY

Before proceeding to the case studies I want to consider the prescriptions of reliability and validity in the abstract with respect to the framework for experimentation that I have provided. A relia-

bility criterion is used primarily during the design and implementation stages of data production. It prescribes that an experimentalist take simplifying steps in the move from asking an empirical question about an effect of interest, to translating that question into the context of the laboratory so that the empirical question is *redirected* at an effect to be produced in the laboratory. Reliability is not concerned with how well the match is between the original effect of interest and the effect produced in the laboratory. Instead, it prescribes simplifying measures in order to constrain or narrow the range and number of competing hypotheses pertaining to the causal mechanisms that are productive of an effect of interest–measures that will most likely result, at the end of the data production process, in the discrimination of one claim from of a set of competing claims (mechanistic or descriptive) about the effect produced in the laboratory. Reliability concerns the immediacy of the lab and how to produce an effect in this artificial context and get results–data that can be used to discriminate knowledge claims. It is not concerned with the legitimacy of extending these knowledge claims beyond the context of the laboratory. As it operates to constrain the process of data production, the consequence is often a limiting of the scope of data interpretation.

If concerns about validity are operative at all at this stage it may enable an investigator to appreciate the complexity of an organism and its environment and how it responds to that environment, but I will show that when such information is known, it is used typically for the purposes of increasing reliability by taking measures to control these complex factors. In cases in which validity is operative and taken seriously during the design stage of data production, its prescriptions are dictated with respect to the *goals* of data interpretation, which are essentially *explanatory goals*. Validity is forward-looking; the prescriptions that it makes in the context of the design stage of data production are with an eye towards the legitimacy of applying the claims discriminated by the data production process both to the (a) effect produced in the laboratory and (b) the original effect

of interest. Ensuring the validity of the move from a true claim about the effect produced in the laboratory to a claim about the original effect of interest requires, in part, a mapping between the original effect and its context and the effect produced in the laboratory in an artificial setting.

5.3 TENSION BETWEEN RELIABILITY AND VALIDITY: EXPERIMENTATION AND MAKING THE LTP-LEARNING LINK

Since my aim is to understand the nature of the link between synaptic plasticity and learning by analyzing the relationships between the two primary types of experiments that have been used to make the case for such a link, I want to propose the following set of steps for this analysis. First, I will consider reliability and validity in the context of those synapse-level LTP experiments in which the first explanatory aim is to test a hypotheses and discriminate a true claim about the molecular mechanisms of LTP generally. Then, I will consider the issues of reliability and validity with respect to extending such claims beyond the laboratory to: (1) LTP as those effects that are produced in all laboratories that study it–(i.e., those phenomena combined that neurobiological investigators all classify as LTP), (2) learning as an effect that is produced in neurobiological laboratories and (3) learning as an effect that occurs in the world. Third, I will consider reliability and validity in the context of those learning experiments in which the aim is to discriminate a true claim about the molecular mechanisms of organism-level learning. Then I will consider the issues of reliability and validity with respect to extending such claims beyond the laboratory to (1) learning as an effect that is studied in other laboratories and (2) learning as an effect that occurs in the world.

Despite the fact that most neurobiologists will claim that LTP is simply a model of learning,

and that studying LTP as well as other forms of synaptic plasticity (e.g., LTD) is undertaken only to provide us with insight into the mechanisms of learning, these studies are intimately linked and they have been so historically. It was a particular conception of learning, namely Donald Hebb's conception, that learning involved primarily associations between stimuli and responses (S-R learning) or between two or more stimuli (S-S learning) that informed his thinking about the neural (i.e., synaptic) mechanism that must underlie learning:

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that As efficiency, as one of the cells firing B, is increased (Hebb 1949, 62).

And, what informed Hebb's postulate was in part the Pavlovian idea that repetition and contiguity of stimuli are sufficient for learning to take place–so, too, Hebb assumed that the contiguous presentation of stimuli in proximity to an organism would result in the repetitive and contiguous firing of the two cells carrying information about the stimuli, and if these cells were near enough to each other, they would fire together repeated and contiguously and so their connections would be strengthened. That it was a growth process that was involved in such strengthening was suggested by the anatomical findings of Santiago Ramón y Cajal and Lorente de Ńo. So, associative learning, as it had been studied in classical and operant conditioning laboratories was Hebb's model for a neural mechanism of all learning.

For years (nearly 25) neurophysiologists wondered if neurons were capable of exhibiting the kind of effect that Hebb postulated, so there were a variety of attempts to stimulate cells artificially, causing them to fire repeatedly and simultaneously without success until in (1973) Bliss and Lomo produced the effect, namely a long-lasting potentiation in the perforant pathway of the hippocampus of an adult anesthetized rabbit that later became known as LTP. Although I do not want to go

into the history of experiments that led up to modern day research on the molecular mechanisms of LTP,[1], this digression into the history of LTP is in part to suggest, that synaptic plasticity research has been from the start constrained by a particular theory of organism-level learning–how it is produced, those features it exhibits as an effect–namely a theory that grew up and was shaped in an experimental context that was itself very much removed from learning as an effect in the world–a theory and an experimental approach that were criticized for their lack of *validity*.[2]

5.3.1 Synapse-Level Plasticity Experiments

In Chapter 3, I gave a brief description of *in vivo synapse-level plasticity experiments* undertaken in the rat hippocampus. I selected to highlight this type of plasticity experiment because I am most familiar with the stages of data production (design and implementation) and data interpretation with respect to these experiments (cf. Sullivan, Kanterwicz, Thiels (in prep)). However, these experiments are representative types of plasticity experiment, sharing some general features in common with plasticity experiments undertaken in other areas of the brain (e.g., the amygdala, cerebellum) in the anesthetized and awake behaving rat. The aspects of these experiments that I will highlight also hold true for *in vitro* electrophysiological plasticity experiments undertaken in the rat hippocampus and the rat amygdala *in vitro* (i.e., in hippocampal or amygdala slices).

Synaptic-level plasticity experiments begin with an empirical question about an effect of in-

[1] I refer the reader to my M.S. Master's Thesis, in which I provide such an historical account, which is available through the Pitt ETD website.

[2] To date, while LTP is assumed to be the neural correlate of learning and memory, no conclusive evidence has been produced to substantiate that when an organism learns LTP occurs in that organism's brain or brain correlate.However, artificially induced synaptic plasticity experiments offer some of our best insights into the cellular and molecular mechanisms of organism-level learning experiments–so much so that it is what we know about the molecular mechanisms of plasticity that drive our ideas about what mechanisms of learning to search for in organism-level learning experiments as well as the designs of these experiments, as is exemplified in recent work (Whitlock et al. 2006). Typically synaptic plasticity experiments are included in research studies in which learning is also studied.

terest. This is a general question typically formulated as: What is the role of molecule x (e.g., ERK) in the production of effect of interest y (e.g., LTP) in area z (e.g.,the hippocampus)?[3] This question is then translated into a question about the role of that molecule x in the production of the effect to be produced in the laboratory. It is at this stage of the experimental design that an investigator can select from a whole host of different methods for producing LTP as an effect in the laboratory. In Chapter 2.0, I identified these methods as *stimulation paradigms*. One can use a stimulation paradigm with differences with respect to any of following parameters: the stimuli could be delivered at high-frequency (HFS) or at the frequency of theta rhythms: theta-burst stimulation (TBS), with different numbers of pulses, different frequencies at which these pulses are delivered, different numbers of trains of HFS or bursts of theta stimulation, different duration of stimuli, different inter-pulse and inter-train intervals, essentially any number of protocols.[4]

Different types of considerations may inform the selection of the electrophysiological stimulus/stimuli to be used in such experiments. An investigator may be concerned with how the form of LTP that she is studying in the laboratory relates to other forms of LTP being studied in other laboratories. She may be concerned instead with how the form of LTP under study is related to a form of learning under study in her laboratory or another laboratory. She may also (although this would be rare with respect to plasticity experiments) be concerned with how the form of LTP under study relates to learning as an effect in the world.[5] Clearly, she probably will not be able to answer

[3] A variety of molecules/brain areas could be substituted in for the first and last variable.

[4] And, I am only considering electrophysiological stimulation paradigms here, but there are pharmacological manipulations that can alter the activity of protein kinases (e.g., protein kinase A) and others that can alter the activity of ion channels and receptors (e.g., NMDA, TrkB) that may also produce a change that looks like LTP, without direct stimulation. In some cases we may be hard pressed to want to identify all of these as instances of LTP, because if we admit so many varying cases, the meaning of LTP as it is captured in Hebb's postulate is lost.

[5] Although instances of this last concern driving LTP experiments may be rare, typically in the introduction and discussion sections of published research studies, the term "learning" is used generally–that LTP is thought to underlie "learning", not just one form, but all forms. Although this might not be the meaning that the authors intend, if we were to interpret what is written literally, everyone seems to be claiming that LTP underlies learning generally.

all of these questions simultaneously, but concerns about extending the claims that she arrives at in the laboratory to any one of these three other types of effects prescribe that she at least try to answer one of them, it is at her discretion to decide which question is most important, that is, if she takes any one of them to be important. And, even if she may take one to be important, this might not be reflected in the choices that she makes during the design stage of data production, even if it is reflected with respect to how she interprets her claim in relationship to these other effects. If she addresses none of these questions, then the extension of that claim she arrives at in the lab about the role of a molecule of interest in the effect that she has produced in the lab to these other effects will be essentially invalid (since, as I have said previously, an extension is an interpretation, and interpretations can be valid or invalid).

In the design stage of data production, given an investigator's interpretive/explanatory interests, various options exist for producing the effect of interest. An investigator can select that stimulation paradigm with parameters that she deems best able to produce a robust change in synaptic potential that is long-lasting, or long-lasting enough to study the activity of a molecule over an extended period of time (at least up to 1.5 hours) after the end of the application of the LTP-inducing stimulus. The longer an effect is studied the more readily the precise role of a given molecule in the production of an effect can be investigated, because the temporal profile of the activity of a molecule compared with the temporal profile of the produced effect, i.e., LTP, may yield clues as to during which phase of LTP out of the three that have been identified, induction, maintenance or expression[6], the molecule is operative. This also holds true for intervention experiments–if a

[6]LTP induction essentially pertains to those molecular events at the synapse that are requisite for LTP to occur; LTP expression is thought to be underlied by a different set of molecular events that result in the sustained potentiation; and LTP maintenance is thought to include all of the downstream changes that subserve the effect long-after induction. These phases have been teased out by pharmacological methods that suggest the activity of certain molecules is necessary at specific times following stimulation.(Sweatt et al. 1998)

long effect is observed without pharmacological manipulation, but with it, the effect cannot be induced at all, or it can be induced but it lasts only for 15 minutes–these can provide clues as to whether the molecule is necessary for the induction, expression or maintenance of LTP. Clearly the prescriptions of reliability are operative in the choice to use such stimulation paradigms, because it increases the likelihood of addressing the empirical question about the effect produced in the laboratory: what is the *role* of ERK in the production of LTP?

However, once a stimulation paradigm has been selected, the empirical question is transformed into: What is the role of ERK in the production of *high-frequency induced long-term potentiation* (not just any LTP, but HFS-LTP–and not just any HFS-LTP, but HFS-LTP with the HFS having these specific parameters and the LTP having this specific electrophysiological profile)? In other words, once the investigator moves to study LTP in the laboratory, and picks a stimulation paradigm that gives her a form of LTP to study. If that stimulation paradigm that she selects for the sake of reliability bears no resemblance to stimulation paradigms used to induce LTP in other labs, even if the effect in principle does bear such a resemblance, can she be said to be studying the same effect that investigators in other labs who also claim to be studying LTP, or even LTP in area CA1 of the hippocampus are studying?

I think the answer to this question is no, and so I think that any claims that are discriminated in this laboratory and that may be taken as valid with respect to the form of HFS-LTP in area CA1 of the hippocampus under study in this laboratory, so long as the overall data production process was reliable, will not be valid with respect to other forms of LTP under study in other labs. This is so because the causal variables used to produce the effect of interest in the one case, do not map onto those causal variables used to produce the effect of interest in another case, and this leaves open the possibility that the variables that intervene between one form of LTP-inducing stimulation and the

effect produced by it may differ from another form.[7] This may explain why different laboratories have obtained different findings with respect to the roles of different molecules in the production of LTP. For example, Eric Kandel and colleagues have suggested that CAMKII is necessary for LTP, Sweatt and colleagues that ERK is necessary, other labs have suggested it is PKA (e.g., Winder) that is necessary. If the causal variables used to produce LTP in these labs differ, then this might account for the differences in claims from one lab to another about which molecules are essential for the production of LTP and what role they play.

That those neurobiologists who study LTP have very little concern with validity when it comes to the issue of whether the form of LTP that they are studying is valid with respect to any other form that has been studied is suggested by the fact that it is common to find any number of stimulation paradigms described in the LTP literature that are being used to produce LTP in various regions of the brain or in brain slices. This is because the selection of stimulation paradigms is often driven by idiosyncratic views about which stimulation paradigm will ensure that the data production process will be most reliable[8]–and there seem to be many options for this purpose. This is clearly a problematic situation. If one explanatory goal of neurobiologists who work on LTP is to understand the mechanisms of LTP, then it would seem that LTP can be produced in many ways, not just one way, and maybe not every form of LTP is equal to another. If that is true, then just what are neurobiologists identifying the molecular mechanisms of? LTP generally, or a multitude of different types of LTP? The fact that there are so many different ways to produce LTP

[7]Note, the only way to determine if there are similarities or differences between these different forms would be to undertake parametric studies. Yet given that there are so many different stimulation paradigms with different protocols attached to them in existence, it may be more appropriate at this point to try to coordinate efforts across labs instead of continuing to produce more and more different forms of LTP. It is not clear what the benefits will be if research continues on its current course.

[8]I do not want to exclude the possibility of other motivating factors, however, reliability does seem to be fundamental.

and different identified types of LTP (e.g., HFS-LTP, TBS-LTP) will clearly make the second aim of synaptic plasticity experiments, namely identifying the cellular and molecular mechanisms of learning and memory, even more difficult. And right now we do not have a clear division of labor with respect to different forms of synaptic plasticity, just a mishmash of different effects all placed under the same heading, with any host of molecules differentially activated and implicated in their production.[9]

The second explanatory aim of synapse-level plasticity experiments is to understand the cellular and molecular mechanisms productive of *organism-level learning*. As I mentioned above, we can divide this into two separate empirical questions: How do claims arrived at in the laboratory about a form of LTP relate to (1) a form of learning produced in the laboratory (2) learning generally as an effect in the world or even (3) the same form of learning in the world? The desire to make the link between the mechanisms of synaptic plasticity and the mechanisms of organism-level learning is at least implicit in published research papers in the synaptic plasticity field. In addition, it has become commonplace for research studies to contain both organism-level learning studies and synaptic plasticity studies (typically LTP studies) both involving biochemical investigations of the same brain area, in which the findings of the two studies are compared to lend credence to the idea that the same molecular mechanisms are operative in both types of effects (e.g., Liu et al., 2007; Grenado et al., 2007; Atkins et al., 1998). So, a second tension between reliability and validity arises with respect to the primary goal of synaptic plasticity experiments that has to do primarily with the relationship between LTP and learning.

If there are a variety of stimulation paradigms that can be used to induce LTP in the brain, some

[9]It seems as though some neurobiologists and philosophers regard this "multiple realization" status of LTP as compatible with the "reductionist" or "mechanistic" goals that have been attributed to neurobiology. (e.g., Bickle 2003)

of these paradigms must bear a closer resemblance to those stimuli that the cells being stimulated would be most like when a whole intact organism is presented with a stimulus, or rather multiple stimuli, or simply a learning situation. In fact in the early 1980's investigators interested in inducing LTP did have concerns about how closely the stimuli used to induce LTP ought to resemble the electrophysiological activity invoked by natural environmental stimuli. The stimulus paradigm taken to bear the closest resemblance to such natural stimulation is theta-burst stimulation (TBS). However, HFS is the more widely employed stimulation paradigm, suggesting that many investigators are not concerned with whether the stimuli used in LTP experiments bear any similarity to natural stimuli.

There are several additional and important features of synaptic plasticity experiments that are relevant to my analysis. First, the model of the organism that is used in LTP experiments is an simple reduced model compared to the input-output model of the organism (Figure 1, Chapter 1) that may be drawn based on what is known about the relay of information from sensory inputs to motor outputs in an intact whole mammalian organism. Investigators that work on LTP treat the synapses at which they produce the effect as deserted islands in the brain–they do not consider events upstream from the location at which they are attempting to produce the effect, nor do they consider the consequences downstream. The input-output model that is employed typically includes only the pre-synaptic neurons to which stimulation (input) is applied to the post-synaptic neurons from which recordings are taken (output). The intervening variables of interest are selected events (depending on one's interest) that take place in either the pre- or post-synaptic neurons. The reason such a simplified model can be used in *in vivo* plasticity experiments of the type I am considering is that the rats are anesthetized, and under anesthesia, the assumption is that since there are no sensory inputs and no behavioral outputs, the brain is essentially silent; the synapse under study

"naive". Of course, the obvious differences between these two models suggest that any claims about the molecular mechanisms arrived at in an experiment with an anesthetized rat will be incredibly difficult to relate to learning in awake behaving rats. In this case, the measures taken to isolate the synapse in the rat's brain are in part undertaken to simplify the model of the organism to insure that the effect produced is actually the result of the experimental manipulation and that a true claim is discriminated with respect to the role of a molecule of interest in the production of the effect. So, again, concerns about reliability are operative in these experiments rather than concerns about the possibility of extending the claim to organism-level learning.

A related issue is how claims about the molecular mechanisms of LTP arrived at in LTP experiments can be extended to all the common forms of learning under study in neurobiological laboratories, such as fear-conditioning, classical conditioning of the eye-blink reflex (rabbit) and spatial learning in the Morris water maze to name a few. It may be claimed that LTP is most similar to specific forms of associative learning than non-associative forms of learning that we might identify (e.g., even conditioned taste aversion). As I mentioned before, the use of different stimulation paradigms for inducing plasticity makes it difficult to know how to relate findings from one lab about the molecular mechanisms of LTP to findings from another. Even if the same lab (as was the case in the research study I consider in the next section) conducts LTP and organism-level learning experiments, studies the activity of the molecule of interest in the same brain area in both types of experiments, the stimulation in the two cases in not identical, the input-output models are not identical, and the correlational claims that are made on the basis of the same molecule being operative in both cases seem to invoke an extension of the results of one study to the other, which is not a valid interpretation given how truly different the two produced effects are.

Just so long as a stimulation paradigm results in the contiguous and repetitive firing of pre- and

post-synaptic neurons at the synapse under study, much like how associative learning is assumed to require such contiguity and repetition, and just so long as the observed effect is as a long-lasting behavioral change in response to a stimulus–which at brain synapses is simply a change in synaptic response, then this is taken to be sufficient for the LTP-learning link–or at least is used in support of the idea that LTP is a likely candidate for learning. However, I do not think that a satisfaction of these conditions alone could ever be enough to even get at the mechanisms of *associative learning*. That it is not enough has been suggested by cognitive and neuroscientists (e.g., Shors and Matzel 1998).

Yet another tension between reliability and validity that emerges with respect to synaptic plasticity experiments concerns the choice of stimulation paradigms. Reliability prescribes selecting a paradigm with stimulation parameters that produces data that can be used to discriminate between competing hypotheses about the role of the molecule of interest in the effect. Validity with respect to LTP experiments would prescribe utilizing stimuli that more appropriately matched those electrical stimuli that would be produced in the brain when an organism learns, if it is indeed the case that one wants to make specific claims about the molecular mechanisms involved in learning. This is in fact the specific prescription of ecological validity, if we are to take seriously either the immediate environment (i.e., the brain region) or the global environment (the brain) in which the synapses that are under study in plasticity experiments exist. However, neurobiologists would claim that our knowledge of the brain is too limited to undertake such studies, that the benefits of changing current approaches do not outweigh the costs, and that even if the claims arrived at are not applicable beyond the lab, it is better for our data production processes with respect to LTP to be reliable and yield some insight into general plasticity and learning mechanisms, rather than make a move to increase validity, which would result in the data production process being less reliable and

not yielding a clear benefit. For, if stimulation parameters are used in such experiments that resemble more closely organism-level learning, there are too many additional variables that would need to be considered, and in turn, more control that would need to exerted in the experimental context over all of these additional variables. Validity would prescribe incorporating such complexity into the experimental design. Reliability prescribes simplification. In general investigators go the route of reliability. Of course, the trade-off is an inability to extend the claims that they arrive at in the lab to learning.

In summary, in LTP experiments, what is clear is a commitment to the reliability of the data production process even at the cost of sacrificing (a) construct validity (with respect to both the model of the organism, and the match between the construct of LTP in one laboratory compared to the effect as it occurs in various laboratories, and the match between the construct and learning as it is studied in laboratories or as it occurs as an effect in the world); (b) ecological validity in so far as it does not take into consideration the variables in the local and global environment in which a synapse resides and (c) external validity, which would prescribe taking measures in the experimental design that would guarantee the legitimate application of the claims that are arrived at via the data production process at least to LTP generally.

5.3.2 Organism-level learning experiments

The second types of experiment that I introduced in Chapter 2, Section 2.3.1. are organism-level learning experiments. The aim of these experiments is broadly to identify the cellular and molecular mechanisms of learning and memory. A handful of experimental paradigms currently exist for studying learning in the rat[10], the five that are used most widely are (1) cued fear-conditioning,

[10]I identified several of these by name in Chapter 1 and have included a descriptive list in the appendices.

(2) contextual fear conditioning, (3) cue and contextual fear conditioning, (4) the Morris water maze and (5) conditioned taste aversion. Here I only want to consider a neurobiological study (Atkins et al. 1998) in which the effect that was produced in the laboratory was fear-conditioning and the paradigm used was one type of cued and contextual fear-conditioning paradigm. While the study contains additional experiments, including contextual fear conditioning experiments and LTP experiments undertaken in slices of the hippocampus of the rat in which LTP was produced in area CA1, I will not here consider these other experiments. The claims that I have already made with respect to synaptic plasticity experiments are applicable to the experiments contained in this study, and the claims that I will make about the relationship between reliability and validity in the organism-level learning experiment that I will consider will hold for the other organism-level learning experiments contained in the study as well.

In these experiments the effect of interest produced is a form of learning, sp., a form of mammalian associative learning and the aim was determine if a protein kinase, the extracellular signal-regulated kinase (ERK), plays a role in this type of learning. I am selecting this study for several reasons. First, the lab in which the study was conducted is one of the most prominent neurobiological labs in the country studying the role of a variety of protein molecules in both learning and synaptic plasticity events. Second, the study has been widely cited in those research studies and review papers that are concerned with the role of molecular signaling cascades in learning and synaptic plasticity. Third, it is representative of experiments that are still being undertaken currently in cellular and molecular neurobiology. Fourth, the primary investigator, David Sweatt, has done extensive theoretical as well as experimental work in the area of learning and synaptic plasticity–in addition to review papers, he has published the current state of the art textbook in cellular and molecular neurobiology: *Mechanisms of Memory* (2004). So what we can obtain looking

at the corpus of Sweatt's theoretical work in addition to his experimental work is the different roles that reliability and validity can and often do play in the experimental study of learning and synaptic plasticity.

The published research paper (Atkins et al. 1998), contains descriptions of several types of organism-level learning experiments that instantiate those features that I described for such experiments in Chapter 2.0. The research study begins with a question about an effect of interest, namely: What is the role of mitogen-activated kinase (MAPK)/extracellular signal-regulated kinase (ERK) in *mammalian associative learning*? Associative learning can be construed as both an effect in the world and a pre-operationalized concept. Clearly we can detect instances of this in the world: for example, when a child learns that an object with four legs that barks in the world and the word "dog" somehow go together–we say that the child has learned to *associate* the word with the object.[11] This would fit the general definition that neuroscientists have given associative learning: when an organism learns a relationship between two or more stimuli, or a stimulus and a response. However, associative learning cannot be measured directly in the world, and in order to study it in the experimental context it must be defined operationally.

In the Atkins et al. 1998 study, the effects to be produced in the lab that are taken both to be operationalizations of associative learning, which are supposed to be produced in the context of two different experimental learning paradigms and measured indirectly by means of a comparison of measures taken pre- and post-training are two instances of *fear conditioning*, (which is itself identified as an associative form of learning): (1) *cued-fear conditioning* and (2) *cued-and-contextual fear conditioning*. Although I am not opposed to the idea that fear conditioning may occur in the world all of the time[12], it is not clear in what ways that phenomenon will be similar to the effect

[11] I do not mean to say, however, that language learning can be understood purely in terms of associative learning.

[12] As Floh Thiels pointed out during my defense, a lot of fear conditioning similar to the kind studied in the labora-

produced in Atkins and colleagues' lab (i.e., whether it will have the same features or be produced by similar mechanisms) or will bear resemblance to associative learning construed more generally.

The protocol for producing contextual fear conditioning consists of the following steps:

> The animal was placed in the fear conditioning apparatus for 2 min and then a 1-second shock (1 milliampere) was delivered to the floor grid. The animal remained in the apparatus for 10 s, then was removed for 45 seconds into a neutral holding box. The animal was placed back into the fear conditioning apparatus, then shocked again after 2 min. This protocol was repeated for a total of 5 shocks. [. . .] The stimulus strength and number of training pairs were chosen based on pilot experiments to optimize learning. To assess contextual learning, the animals were placed back into the training context 24 hours post-training and scored for freezing for 5 minutes. (Atkins et al. 1998, 608)

As Atkins and colleagues claim, this protocol for contextual conditioning was selected strictly to "optimize learning", a prescription of reliability. Reliability urges that measures be selected to produce robust effects, because robust behavioral effects are more easily detectable and more easily attributable to the learning situation. Such measures also have the potential to produce robust molecular effects that can be used to discriminate between competing causal hypotheses about the role of MAPK/ERK in the production of the effect. An alternative protocol with different parameters may not result in the production of as robust behavioral and molecular effects, and would thereby make it difficult to arrive at a conclusion about competing hypotheses as to the role of the molecule in the effect.

Learning is operationalized with respect to the contextual conditioning paradigm as an increase in freezing to context after training compared to freezing response to context alone prior to training. So, fear, which also cannot be studied directly in the rat, is operationalized by means of measuring extent and duration of freezing for 5 minutes after training. There may be other ways to assess fear,

---

tory is probably occurring today in Iraq.

in fact, paradigms that provide a more reliable assessment of fear in the rat are being developed, which suggests that current methods of detecting fear are not necessarily comprehensive. Again the operationalization of learning is contingent upon reliable assessments of fear.

In addition to selecting a learning paradigm attached to a protocol that produces a robust effect, the model of the organism that informed these studies was extremely simplified compared to that input-output model of the organism represented in *Figure 1*. The investigators assume no intervening variables to be operative between the stimuli presented to rat subjects and their behavioral outputs. Rat subjects are raised in impoverished environments in the laboratory in order to increase the reliability of the learning paradigm and its protocol to be the exclusive causal variables operative in the production of learning as an effect in the laboratory. So animals are assumed to be naive coming into these experiments. On their model, the stimulus inputs (context+shock) presented to the organism are assumed to be followed by ERK activation in the hippocampus and the downstream consequence of this activation is the behavioral response.

Using this contextual conditioning paradigm, Atkins and colleagues found that activation of the p42 isoform of ERK was increased significantly in the hippocampus of animals 1 hr after they were trained in the contextual fear conditioning paradigm compared to naive animals. This increase correlated with learning, i.e., freezing or fear conditioning to context, taken to be indicative of the animal forming a predictive association between the context and shock (or that the context predicts a forthcoming shock), which was determined by assessing the duration and extent of the rat's freezing to the context after being trained in the learning paradigm. In addition, in experiments in which an inhibitor of ERK activation, SL327 was used to block activation of ERK by injecting this drug into animals at least 90 minutes prior to training, this blockade was correlated with a lack of freezing to context following training.

Atkins and colleagues' (1998) study is motivated primarily by concerns about the reliability of the data production process rather than by concerns about the validity of data interpretation. This is apparent with respect to the choices made in the design stage of data production, but it is precisely these choices that affect the validity of data interpretation. First, Atkins and colleagues use a protocol to induce fear conditioning that is particular to their laboratory, which means that the causal variables operative in the production of the effect may be different from those contained in fear conditioning paradigms employed in other laboratories. This raises the question of whether the claim about the role of ERK in the form of fear conditioning under study in that laboratory can be extended to contextual fear conditioning studies undertaken in other laboratories. Furthermore, the claim that Atkins and colleagues (1998) make in the title and in the discussion section of their paper is that the profile of ERK activation that they show to be required for contextual fear conditioning (as well as cued and contextual fear conditioning) provides adequate evidential support for the claim that ERK activation plays this same role in all "associative learning". Here we might ask many questions, the primary one being is this what this study actually demonstrates?

I think that the claim arrived at with respect to the role of ERK in contextual fear conditioning begins and ends with the data production process that uses this paradigm. However, I do think Atkins and colleagues want to extend the claim to all forms of associative learning–not merely those forms of learning studied in conditioning experiments, but any learning that involves the establishment of associations (which for Sweatt appears to be all of it). This is evident in Sweatt's (2004) textbook in which he claims:

> It is likely that most declarative learning occurs as learning something within a variety of contexts: other facts or places with which the fact is associated. I stress this point because it is important to keep it in mind as we begin to explore the molecular basis for declarative learning. It is certainly possible that many of the molecular mechanisms that are discovered as subserving what we have defined as associative conditioning may translate directly as mechanisms for declarative learn-

> ing. Stated more strongly, at this point it is appropriate to hypothesize that associative molecular mechanisms will be part of the molecular infrastructure of declarative learning. (Sweatt 2003,)

This is indeed a strong claim, but based on how many different forms of associative learning Sweatt himself identifies in *Figure 6*, the likelihood that these different forms of learning share molecules in common may be true, but that the mechanisms are conserved across all types seems implausible given how many different options there are for producing associative learning. Furthermore, if we compare the taxonomy of learning arising from neurobiology and compare it to a taxonomy that is more theoretically driven, as Squire's model in *Figure 2* is, then what becomes clear is that out of neurobiology there are as many different learning types as there are experimental paradigms for its study, and the fact that these types of learning are themselves differentiated operationally suggests that they are different types and current molecular analyses may lack the sophistication to get at what mechanistically differentiates these different types. In addition, I think we might expect a much more cumbersome taxonomic scheme than that provided by Squire and it may become increasingly unclear, how to relate the one taxonomy to the other.

5.4 CONCLUSION: CAN WE GET RELIABILITY WITHOUT VALIDITY AND VICE VERSA?

In *Figure 12*, I have attempted to pictorially represent the relationship between reliability and the 2nd condition of validity, namely the condition pertaining to the idea that the context to which an interpretive claim is applied must be similar to that context in which the claim is produced. In this graph, the line for validity is designated with a "V" and the line for reliability is designated with an "R". Interpreting the graph from left to right, we can anticipate that when the overall reliability

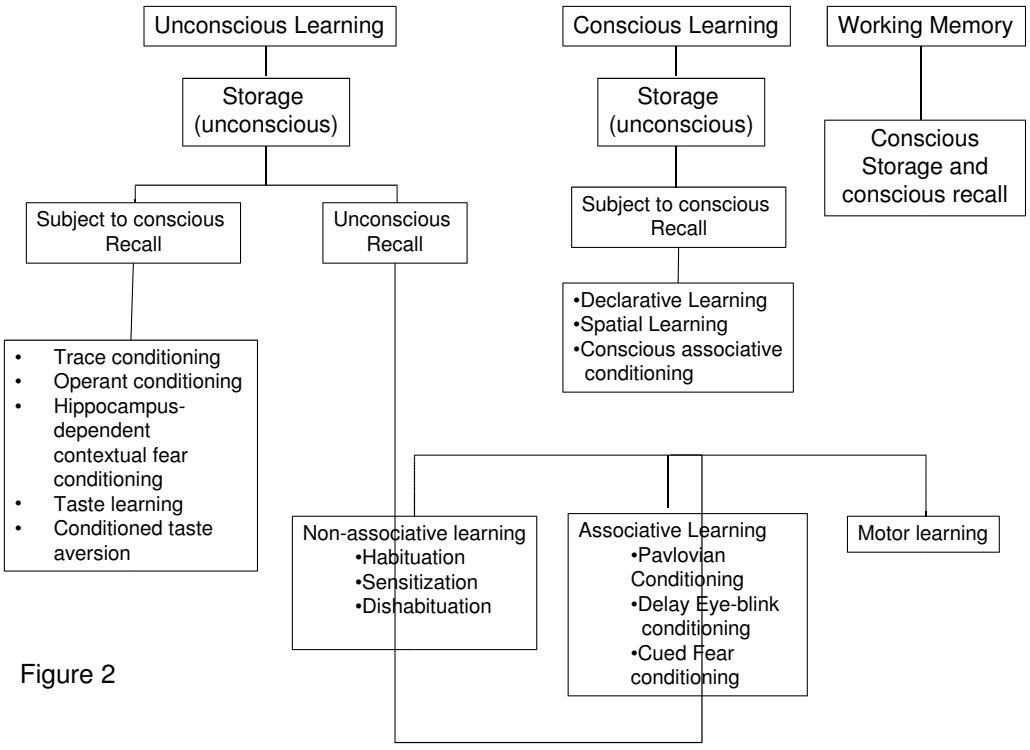


Figure 11: Sweatt's Taxonomy (Adapted from Sweatt 2004)

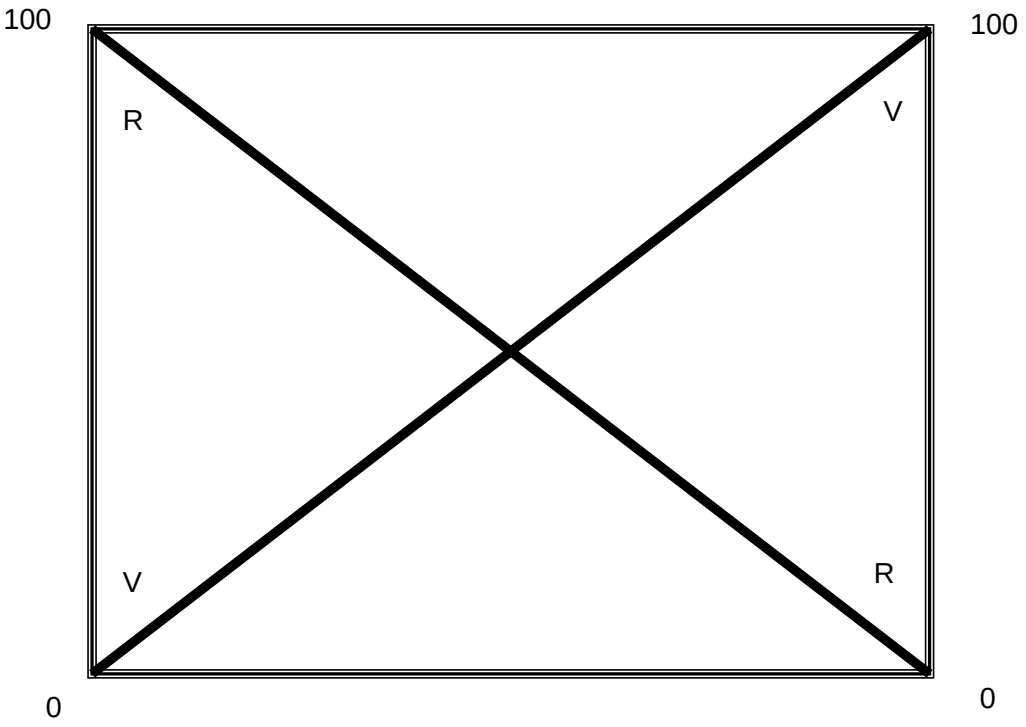


Figure 12: Validity-Reliability Graph

of a data production process is high, it corresponds directly to the low validity of the interpretive claims with respect to their possible extension beyond the context of the laboratory. However, when the similarity between the world and the lab is high, we can expect that the reliability of the data production process will be low.

This graph is clearly an idealized representation of the relationship between reliability and validity. The relationship between the two is not likely to be linear.[13] However, this graph is just meant to conceptually distinguish the 2nd condition of validity from reliability. It is the second condition that prescribes in the context of data production to increase the similarity between the lab and the world, but the consequences will clearly be negative for reliability. They will also be negative with respect to the first condition of validity, reliability. For, when we aim to meet the 2nd condition of validity, and increase the similarity between the lab and the world, we introduce too much complexity–too many variables, into the experimental context. The likely result is that neither reliability nor validity is achieved.

If considering the parameters of validity during the design stage of data production would increase the scope of our knowledge of possible variables that may confound our experiments, and as a result, point us in the direction of where we need to exert more control in order to ensure the reliability of the data production process, can we then construe validity as a prerequisite for reliability? Yes and No. It would appear that concern about controlling confounding variables that may jeopardize the reliability of the data production process might prompt an investigator to itemize a list of possible confounding variables, but a knowledge of these variables could either be used to increase the validity of the data interpretation, which may or may not decrease reliability, or it could be used to increase reliability by exerting more control over the experimental system

[13] Paolo Palmieri correctly pointed this out.

and the data production process.

If validity ascriptions are best made to data interpretation, and such ascriptions require that the claims that are part and parcel of data interpretation have a solid foundation and this solid foundation is (essentially) the reliability of the data production process, then can we construe reliability as prerequisite for validity? I think we can, but clearly, if pursuing reliability at all costs negatively impacts the validity of our experiments, we have to inquire if there is some way to preserve the reliability of the data production process while at the same time increasing the validity of the claims that arise from this process with respect to the world outside of the laboratory. I present one possible way in the next chapter.

The predicament in contemporary neurobiology may be explained by several factors. First, the neurobiology of learning and memory research is still a field in its infancy and the neurobiological study of learning and memory has exploded over the past 30 years in various labs all over the world. A concerted effort to appreciate the complexity of learning has only just begun over the past 5 years, with a publication of two lexicons of learning and memory (Dudai 2002; Roediger, Dudai, Fitzpatrick 2007) and with meetings (e.g. two, James S. McDowell sponsored meetings on learning and memory in January of 2004 (La Jolla, CA, organized by philosopher Carl Craver) and November of 2005 (NYU, organized by neurobiologists LeDoux and Dudai)) that have brought together experts in the field to identify those fundamental challenges for the neurobiological study of learning and memory.[14] Second, as my characterization of the relationship between reliability and validity suggests, reliability should be a goal of data production and it is rational that investigators should aspire to promote the reliability of their experiments, since if a claim discriminated in the process of data production is shown to be false, in part because the process used to produce

[14]Of course, this does not mean that neurobiologists have addressed explicitly the conflict between reliability and validity with respect to their experiments or the specific drawbacks of their near exclusive commitment to reliability.

it was unreliable, extending a false claim to other contexts will trump validity as well. Third, in general in the laboratory sciences, the designated hallmark of good research is the reliability of data production. The general community, including peer reviewers of the top ranked journals in the field, seems to be less concerned with how interpretations of results are framed. This is evident especially in neurobiology, since in the methods section, researchers tend to make it a point to emphasize the reliability of those experimental arrangements that they use to produce data, but in the titles and discussion sections, the interpretative claims made are bold, often extending far beyond what is substantiated by the data. Fourth, there is an additional possibility that neurobiologists who study learning and memory do not want their experiments to be valid–that they do not take the application of their results outside the lab to be their ultimate goal. For example, Mook (1983) suggests of psychological experiments, that sometimes external validity is not a goal of scientific research, that sometimes it has a more modest aim:

> But even where findings cannot possibly generalize and are not supposed to, they can contribute to an understanding of the processes going on. Once again, it is that understanding which has external validity (if it does)–not the findings themselves, much less the setting and the sample. And this implies in turn that we cannot assess that kind of validity by examining the experiment itself. (Mook 1983, 382)

If neurobiologists want to make the same kind of claim with respect to their experimental findings, then I would only suggest that they change how they frame their claims and be realistic in print about how far the claims about the cellular and molecular mechanisms of learning and memory can be extended beyond laboratory produced effects.[15] However, even if this is true, then neurobiology

[15]I want to add here that Mook (1983) suggested in response to Campbell and Stanley's (1967) paper, which suggested that both external and internal validity should be goals of experiments in the social sciences, that, afterall, psychologists are not really after external validity construed as generalization–that many psychological experiments are not designed to be generalizable, and even if it is the case that sometimes sweeping generalizations are made based on experimental results, this is typically just to stress the importance of the research, rather than lay specific claim to external validity.

is not really addressing those interesting empirical questions about plasticity and learning that they claim to be addressing–and then maybe the findings should not be construed as having the impact that they have been purported to have. Yet, then the question is, what IS really gained as to our understandings of learning and memory by the undertaking of such research?

Based on the purported goals of neurobiologists with respect to explaining the mechanisms of learning and memory, I see two options for the future direction of neurobiological research. Either neurobiologists can continue in the direction that they are headed, and continue to produce reliable findings about effects produced in laboratories that result in claims that extend no further than the laboratory effects that they produce, or they can take specific steps to increase validity when it comes to extending these claims beyond the laboratory context. As a proponent of the second option, in the next chapter, I want to sketch a set of rough guidelines for achieving this goal.

6.0 STRATEGIES FOR ENHANCING VALIDITY WITH RESPECT TO NEUROBIOLOGICAL EXPERIMENTS

6.1 INTRODUCTION

The previous chapters have culminated in the final project of this dissertation. What I have shown up to this point is that neurobiologists often make global claims about the cellular and molecular mechanisms of learning and memory. Yet, in most if not all cases of experimentation in this area, these interpretive claims are at best only *internally valid* claims–i.e., not extendable beyond the context of that laboratory that gave rise to them. This is because experimentalists in cellular and molecular neurobiology are more concerned with securing the reliability of their experiments than ensuring their validity. This aim makes sense, given the complexity of the brain and the complexity of learning phenomena. It is also rational, given that reliability is requisite for validity. And yet, the global explanatory aims of neurobiology, namely, to understand the cellular and molecular mechanisms of learning and memory across organisms, are not being realized. I take this to be the explanatory crisis in neurobiology. The question then, is, how do we overcome the limitations of current experimental approaches in neurobiology? Laying the groundwork for a solution is the primary aim of this chapter.

The basic idea behind the solution that I provide is to increase the similarity between the

learning situations in neurobiological laboratories and learning situations *broadly construed* in the world. Up until now, I have not said anything explicit about how we come to isolate learning situations when we step outside the laboratory. Nor have I provided an account of what the features of those situations are. This is in part due to the fact that throughout the dissertation I have sought to track the experimental process as it occurs in neurobiology. Although I claimed that the experimental process begins with an empirical question about a phenomenon in the world, in neurobiology this process moves swiftly away from the world. But that is where we need to return to in the context of interpretation. The question is, then, what are the features of those learning situations that we want to get back to and how can we come to *build* those features into the lab to make the move from lab to world more tenable? I provide one answer to this question in this chapter.

A primary problem that needs to be overcome is that making the neurobiological laboratory more similar to the world will most likely negatively impact the ability to detect molecular effects and to attribute their causes to actual learning situations. The proposal that I offer in this chapter is based on the assumption that this is not an insurmountable problem. I think it is important to entertain this assumption in order to move the field towards talking explicitly about how to overcome the limitations of current experimental approaches. It is an open question whether the strategy I am proposing will have a definite pay-off down the road. However, as I will show, very recent work on learning and synaptic plasticity suggests that it could. And I take that as enough of an indication that the direction I am suggesting for neurobiology is a viable path to explore.

The structure of this chapter is as follows. Since the solution that I provide is based in part on the idea that experiments can be understood as physical models[1], in Section 6.2, I try to relate

[1] I will explain "of what" in the next section.

the experimental process in neurobiology to a small framework of models. I then attempt to get a handle on precisely how learning situations in the world may differ from learning situations in the lab and the "hard problem" –the challenges that confront us in trying to extrapolate from learning in model systems (sp. rodent) to learning in human beings. In Section 6.3, I propose a framework for building worldly learning situations gradually into the neurobiological laboratory. In Section 6.4, I go on to consider some recent attempts in neurobiology to increase the validity of interpretive explanatory claims about learning and synaptic plasticity that instantiate part of the proposal that I outline. I conclude by considering problems with the proposed framework and the potential negative implications for reliability and whether these problems have a solution.

6.2 MODELS AND THE EXPERIMENTAL PROCESS

Hacking (1983; 1992) emphasizes the idea that scientific practice involves not only intervention, but also *representation*. In science, *models* represent. I take a model simply to be something that represents an idea, an entity (e.g., a thing, a phenomenon or system (organism, synapse, neuron, brain, hippocampus)), an activity (e.g., transmitter release) or a process (e.g., learning) and that may or may not be similarity to that idea, entity, activity or process that it represents. Although this definition is very general, it suffices for my purposes here.

A number of different kinds of scientific models have been identified in the philosophical literature (e.g., Duhem 1954, Hesse 1962, Cartwright 1983, 1999; Harré 1998; Aronson, Harré, Way 2000; Morgan 1999). I will treat exclusively here of two kinds of general types of models: (1) *descriptive* models and (2) *physical* models. To further restrict my focus, I only want to consider 2 subtypes of each type of model (sp., those that are relevant to my analysis of neurobiological

experiments). I take a *descriptive* model to represent an idea, entity or process either by means of a description or a pictorial representation. Often descriptive models are idealized models that are designed to represent something complex in a simplified way. There are two types of descriptive models that I have in mind. The first type of model is a description of an entity or process that is contained in a scientific theory. For example, the description of learning as requiring the repetition and contiguity of stimuli is essentially the general model of learning at the heart of neurobiological learning theory. I suggested in Chapter 2.0, that the theoretical framework in neurobiology consists of a set of different models: of learning, the organism, the synapse and synaptic change. These are all descriptive models.

The second type of descriptive model that I have in mind is that captured in an experimental design and protocol, which is essentially a description of the experimental situation and its overall features from the beginning to completion of data production. I will refer to this type of descriptive model simply as an "experimental model".

In contrast to descriptive models, a *physical* model is constructed out of material components, and is often intended to look similar to the situation, entity, process or activity that it is intended to represent, or to the situation, entity, process or activity itself. A primary feature of a physical model is that its material components can be physically manipulated or altered. There are two types of physical models that I am interested in with respect to neurobiological experiments. The first type of model instantiates a descriptive model in physical components. I will refer to this type of model simply as an "instantiated experimental model". On my view, an experiment is simply an instantiation of an experimental model in material components. For example, in neurobiological experiments the experimental design and protocol, which I am claiming comprise an experimental model, are physically instantiated in the overall "learning situation" in the lab. A learning situation

includes the entire set of components relevant to that "learning situation" including: the experimental subject or "model system" (e.g., a rat or mouse), the different contexts to which the subject is exposed (e.g., the colony room, lab, Skinner box), the stimuli and stimulus patterns occurring within those contexts (e.g., other rats, tone, shock), and the behavioral responses of an organism. These can all be understood as features of a learning situation.

The descriptive model that is contained in the experimental design and protocol of a synapse-level plasticity experiment essentially describes what I will refer to as a "plasticity situation", which is instantiated in the physical model. A plasticity situation includes all the material components of the experiment including: the rat, the brain, a synapse in that brain, the electrophysiological stimuli and stimulus patterns. In part, when we think of plasticity situations, the synapse itself can be regarded as in the context of the brain–the cells that surround it, the environment, and that environment can change in light of rat-level experiences or internal changes, such as increased hormone levels, oxidation or ischemia.

An instantiated model is judged not only in terms of how closely it conforms to the details of an experimental model.[2] In validity assessments, it is also judged with respect to the likeness that it bears to that system, entity, activity or process *in the world* that it is supposed to be "standing in for" (e.g., Morgan 1999, Guala 2005). I want to use a small-scale example as starting point for making a broader claim about the nature of the physical model at the heart of neurobiological learning experiments. "Model systems" of the type we find in neurobiology: rats, sea mollusks and fruit flies, may be viewed, in the first place, as "standing in for" rats, sea mollusks and fruit flies in the natural world. However, more importantly, they are supposed to be "standing in for" the human population. They are meant to be physical models of human beings. Neurobiologists who

[2] And I take this to correspond to reliability.

work on learning are not really interested in rat learning per se, they are interested in rat learning as a model of human learning. [3]

When model systems like rats are used in experiments in order to gain information about human learning, we can think of the instantiated experimental model as containing all of the material components specified in the experimental design and protocol: the environment of the rat pre- and post- experiment and all of those features of the artificially produced learning situation, which I will consider in the next section. In this way we can say that it is the entire model that is "standing-in for" learning in the human case. So, in validity assessments with respect to the interpretive mechanistic claims arising from neurobiological experiments, we should be concerned with a comparative analysis of all of these features of the learning situation and their counterparts in the world.[4] My aim in the next section is to get a handle on these features.

6.2.1 Differences in Learning: Between the Lab and the World

Learning situations in the world do not come as neatly packaged as we find them in the neurobiological laboratory. Even if we consider a relatively circumscribed example, that of students in the educational context of a classroom, the context has multiple stimuli (the subject matter, the instructor, other students, the world outside the classroom), with multiple different stimulus parameters

[3] Yet, you would not guess this given Sweatt's taxonomy of learning (*Figure 10*), which looks more like a taxonomy of learning in the rat.

[4] A similar claim can be made with respect to plasticity experiments. Those plasticity situations and their material contents that we encounter in the neurobiology laboratory function as "stand-ins" for plasticity as it occurs in the human brain during learning. As Hesse (1962) has claimed, models can also be simulations of natural processes. There is a sense in which learning experiments are supposed to be simulations of learning in the world and plasticity experiments are simulations of a natural plasticity processes that result from experience. However, for the sake of restricting the focus of this chapter, I will not address this sense of model here. Neurobiologists are interested in the features of synaptic plasticity in the rat brain, because they want to know something about synaptic plasticity in the human brain.

(e.g., auditory, visual; e.g., differences in frequency of light or intensity of sound). If the learning stimuli proper can be discretely identified at all, they do not unfold over an obviously specified range of time; learning of the subject matter most likely occurs across different contexts in light of different experiences with the subject matter (e.g., reading the texts outside of class, talking to friends about the material). In addition, we have to consider all of those factors that an individual student brings to a given learning situation–including his/her experiential history, attention, motivation and purpose to learn. It is a daunting task to begin even with this case of a human learning situation with which we are familiar, to get a handle on its specific features.

In light of the problems that will immediately greet us if we try to get a handle on human learning in the world, I want to begin in a different place. It is far easier to look first at an experimental system: a rat learning situation in a laboratory, and to itemize those features that human learning situations do not have, than to begin with human learning situations and move in the other direction. I have been engaged in this project to some extent throughout the dissertation with respect to neurobiology experiments, but I want to engage more systematically in an itemization of those features characteristic of rat learning and plasticity situations in neurobiological laboratories here.[5] In what follows, I will itemize a general set of features that I take to comprise the instantiated experimental models of rat learning situations and to identify those features we either do not or are unlikely to find with respect to *rat learning situations* in the natural world. I also itemize those features of the models of synaptic plasticity situations that we encounter in the lab. I identify those features that will most likely not be characteristic of plasticity situations with respect to rats in the

[5]A second, indirect way, which was part of the original formulation of this document was to look at criticisms of the learning paradigms of classical and operant conditioning arising in the history of science in the areas of gestalt psychology (Kohler 1947), ecological psychology (Gibson 1960, 1961; Brunswick) and ethology (Breland and Breland 1961; Hinde 1973; Lorenz 1965; Shettleworth, 1972). This remains the aim of a future project, but one that I will not engage in here.

natural world. So, I will begin by comparing learning and plasticity situations with respect to the lab rat and the wild rat first and will identify the challenges in extending claims arrived at in the one context to the other. Then, I will consider problems with extending the same claims to human learning situations.

The first features that I want to consider in combination are the environment of the laboratory rat and its experiences prior to a learning situation. Rats used in laboratory experiments are raised in a highly controlled and compared to the world, highly impoverished environment. Colony rooms consist typically of multiple cages in which individual rats are housed alone prior to being "run" in an experiment. It is unlikely that they engage in social interactions with other rats. They are raised on a light-dark cycle to mimic a natural circadian rhythm. They receive daily rat chow and water and are visited by veterinary staff who clean their cages. Every now and then there is the semi-random (for rat) event of a human being coming in and taking other rats away for an experiment. Then there is the event of a human being coming in and taking them away in a cage-like container for an experiment, and all the subsequent events that unfold before their being used in either a learning or plasticity experiment (e.g., anesthetization).

While a lab rat is not completely devoid of experiences, given the types of experiences that it may have, it has qualitatively different and quantitatively fewer than that corresponding organism of a similar age raised in the natural world.[6] Rats in the natural world are free. They burrow and forage. They live in social colonies in burrow systems. The males of the species have a polygamous mating system. Many rats in the wild live in large metropolitan areas and encounter a wide range of different stimuli in such contexts including auditory, visual, tactile, olfactory and gustatory stimuli (e.g., see Barnett 1975; Calhoun 1963 on behavior of Norway rat in the wild; Galef 1985 on social

[6]One issue for the reliability of the data production process is that even laboratory rats have had some experiences prior to training in a learning paradigm or being used for a synaptic plasticity experiment.

learning in wild Norway rats). In light of differences in the environment resulting in differences in the overall experiences of the rat, we can say that rats in the wild have the potential to bring a history of experience to bear in each individual learning situation.

What does this mean for extending an interpretive claim about a molecule of interest and its role in learning from the learning situation of the lab rat to those learning situations of rats in the wild? If brains are plastic and undergo experience-dependent modifications of some kind is true, then we can anticipate that the brains of laboratory rats are significantly different from the brains of animals that grow up in the wild. After all, this is the aim of impoverishment for the neurobiological laboratory–to ensure that any observed electrophysiological or molecular changes are the exclusive result of the experimental manipulations. Because animals in the wild have a broader range of different types of experiences, it is quite possible that their synapses are different from those of the lab rat. We could speculate on the differences – perhaps different molecules are operative in learning in rats in the wild, or the activation profiles of those molecules differ. Perhaps the electrophysiological changes in synaptic strength during a single learning event are not displayed in as marked an increase or decrease, given previous alterations at a given synapse. Most likely in the laboratory we are studying an extremely unnatural case of plasticity.[7]

Another feature of the learning situations contained in instantiated experimental models in neurobiology is that the forms of learning are primarily if not exclusively Hebbian. The learning paradigms that are used are primarily classical and operant conditioning. The assumption that learning requires the repetition and contiguity of stimuli presented within a narrowly circumscribed temporal window and spatial location is physically "built into" the learning situation in the

[7]And we can say this, even if it is the case that the learning paradigm used in the lab is similar to some form of learning in the world (e.g., that fear conditioning in rats is like fear conditioning in human soldiers in combat). And I do not mean to suggest here that these synapses are actually "naive", because laboratory organisms do have some experiences. Yet, we can still imagine that synapses in the brains of organisms in the world are far less naive.

instantiated experimental model.

Here, too, with respect to this feature of the learning environment, there are differences between the experimental model and the world. First, stimuli having the exact spatial and temporal parameters found in these learning experiments simply do not occur in the world or occur very infrequently. This does not mean that classical and operant conditioning do not occur outside the laboratory at all. In fact, one of the primary challenges to Pavlov's study of conditioned reflexes consisted in developing methods to prevent his canine subjects from becoming conditioned to the "wrong" stimuli (e.g., the presence of the experimenter could elicit a salivation response, or other features of the environment proximal to the delivery of a food stimulus). Even if we grant that the temporal and spatial parameters of stimuli occurring in the world are different than those that we find in the laboratory, I think it is unlikely that repetition and contiguity alone is always necessary for learning even in the case of the rat. Even in the lab rat we encounter instances of one-trial learning (e.g., rats trained in a taste-aversion paradigm will avoid the stimulus that they believe made them ill).

That repetition and contiguity would be necessary for learning in lab rats may simply be a by-product of their being dull.[8] First of all, lab rats have a limited range of learning experiences, so it may be that in order to produce learning they require repetition and contiguity that may not be required with respect to organisms in the world. To put it another way, rats in the wild are probably more "advanced" than rats in the laboratory, and they may not need repetition and contiguity of stimuli every time they learn that a context is harmful or that running across glass hurts. We may

[8]However, as Whitlock et al. 2006 claim, a primary reason that iterative paradigms are used in neurobiology is to produce a detectable molecular effect, so it may not be the case that all neurobiologists take learning to require repetition and contiguity of stimuli. Still, either they need to recast their interpretive claims about the mechanisms of learning so that they are not invalid, or they need to enhance the validity of such claims by making changes to current learning paradigms or developing new ones.

anticipate that given these differences, the profile of the activation of molecular signaling cascades in such cases will not be identical to those we observe in the lab (e.g., they may be different in so far as where they occur in the brain, the magnitude of the activation, the targets downstream of such activation). And plastic changes, if they do coincide with a learning event, may also be dissimilar.

The instantiated experimental models in synaptic plasticity experiments are based on a similar assumption at a different level; namely, that changes in synaptic strength, like learning, require the repetitive and contiguous delivery of electrophysiological stimuli to neuronal synapses. If it is correct, as I am arguing, that learning situations in the natural world can occur in the absence of the such stimulus patterns, and if synaptic plasticity actually underlies or is responsible for learning, then it may be the case that changes in synaptic strength can occur in the absence of the repetition and contiguity of electrophysiological stimuli.

Another aspect that is built into the instantiated model of "learning situations" is that neurobiological experiments involve the use of stimuli that an organism is not likely to encounter in the wild. Rats in the wild probably rarely have the experience of being shocked on the foot or tail–so these noxious stimuli used in the laboratory environment are not identical or even similar to noxious stimuli common in the natural world. Nor is the overall context of a Skinner box anything like a natural environment–so learning in such cases is not taking place in a natural setting with everyday stimuli as it does when organisms in the world, learn.

A fourth feature of the laboratory "learning situation" is the kind of behavioral responses that are taken as indicative of learning in the lab rat. One may imagine that animals raised in natural environments that have had a history of learning experiences have developed a complex repertoire of behaviors. This is in part because animals raised in confinement have behaviors that are in part

by-products of their environments. Wild rats may or may not exhibit freezing behavior when they encounter a noxious stimulus in the world. Because they are in an open space they can flee rather than freeze. Neurobiologists in general, study behavior to get at molecular changes that intervene between stimuli and responses. But clearly given the repertoire of behaviors that we encounter with respect to rats in the wild, we can imagine that there are interesting differences downstream of molecular changes between laboratory and wild rats.[9]

This issue is similar in "synaptic plasticity situations" in the laboratory where an increase in synaptic strength at the synapse that was stimulated is what is investigated, rather than changes in synaptic strength downstream of that synapse. Clearly, the relevant changes would have to be exhibited downstream of that synapse on the Pavlovian model–there would have to be a change in the behavior of the series of synapses from stimulus to response. However, investigators do not look downstream from the one synapse.

6.2.2 The Issue of Extrapolation to the Human Case

Philosophers of science have become increasingly concerned with the specific epistemological problems that arise in extrapolating from model organisms ("model systems") to human beings (e.g., Ankeny 2000; Schaffner 2001). For example, Schaffner (2001) has argued that problems with extrapolation from model systems will greet us at every level of analysis from the gene to behavior, particularly if *behavior*, *learning* or consciousness are our explanatory targets. Other philosophers have expressed a less pessimistic view in light of the conservation of low-level molecular mechanisms across different species (Weber 2001). In the case of learning, I agree with Schaffner (2001).

[9] And, with all of this said, it is quite possible that learning has taken place in the absence of a behavioral response, which is captured in the idea that there may be a "potential" for behavior but no elicited behavior (e.g., Dudai [1989 1994])

The issue of extrapolation appears nigh intractable when we realize that model systems are situated in experimental contexts. It is not just that neurobiologists want to extrapolate from one organism to another organism; they want to extrapolate from an entire experimental system to a target system in the world.

I have essentially shown that we have good reason to believe that the interpretive claims about the cellular and molecular mechanisms of learning and memory based on the experimental study of learning in rats are unlikely to be valid with respect to rats in the wild. Neurobiologists are not interested in this type of extension. They want their claims to be applicable to learning in human beings. Yet, when we realize that the interpretive claims arrived at on the basis of neurobiological learning and plasticity experiments are not readily extendable to rats in the wild, which are closer to lab rats biologically than we are, this extension to the human case seems illegitimate and misplaced. This is the *hard problem* of extrapolation. What is the solution?

There is a two-fold answer. First, if you want an organism to be a true model system of another more complex organism, you should make the model look similar to the organism it is supposed to be a model of. Human beings (in the civilized world) are not raised in cages, they are exposed to stimuli from fetal development to death, they exhibit complex behaviors and are capable of higher-order forms of reasoning and different cognitive processes including all those that we classify as learning. This is not only the result of biology; on a very basic level it is the result of biological and environment interactions. This is why I take it to be important that we regard model systems as physically situated in the environment of an experiment. It is the whole thing that functions as the model, not just the organism.

It is probably the case that we can learn very little about ourselves by studying organisms that are lower than us on the totem pole. But it seems to be true that we should be more inclined to want

to extrapolate from instances of lower-level organisms that have learning and plasticity situations similar to our own, than from organisms that in the end bear little likeness to members of their own species. Wild rats do more so than laboratory rats. Yet, neurobiologists cannot import rats from the world into the laboratory and run experiments on them. Even if they could, the brains of these organisms would be *terra incognita*– they probably would not know what to look for at the cellular and molecular levels–because what they would be looking for would be shaped exclusively by what they now know. But clearly more importantly, they would not be able to exert the control over such systems necessary for detecting the cellular and molecular mechanisms of learning and memory.

There seems to be only one option then to improve validity in neurobiology: to increase the similarity between the entire learning or plasticity situation in the laboratory by making the instantiated physical models of which they are a part *more similar* to those learning and plasticity situations that organisms encounter in the world. After all, the target system to which neurobiology wishes to extend its claims is not just the human organism, but a target process with respect to a target system, namely human learning. So, neurobiologists need to increase the similarity between the entire experimental system and the target system. They should not only want a better model organism–one that more closely approximates human beings–but also a better experimental model that is similar to those learning or plasticity situations in the world.

By conceptually differentiating the aforementioned basic set of features–the environment, experiential history, stimulus types, stimulus parameters, stimulus presentations/occurrances, history of experiences–that are part and parcel of those physically instantiated models of learning and plasticity situations, we arrive at a general framework of manipulable (or in the worldly case mutable) parameters that we can modify to make those learning and plasticity situations in the lab more like

those in the real world. These parameters can be fit into the classification of types of validity that I identified in Chapter 4.0. The general prescription then, for neurobiology, is to manipulate these parameters in ways so as to increase the structural similarity between the experimental and world models. The primary challenge then, is to introduce these parameters to increase validity while striking a balance with reliability.

I want to end this section by pointing out that the aforementioned general features of learning and plasticity situations in the laboratory, i.e., those features of the instantiated experimental model, correspond roughly to the parameters of validity that I identified in Chapter 4.0. What I have done is a general assessment of the very basic differences that we will find between the lab and the world with respect to learning and plasticity experiments. Yet, any one of the given general features that I have identified can be further divided into sub-features and even a single stimulus can have complex dimensions (cf. Gibson 1960). For the moment, given those features of learning situations that I have treated of in this section, I want to go forward with a positive proposal for increasing validity in the future in neurobiology. In the remainder of this chapter, I will present a case study from contemporary neurobiology that instantiates the basic idea of the proposal I have in mind. I will then conclude by articulating the basic framework of my proposal.

6.3 A RECENT ATTEMPT TO INCREASE VALIDITY IN NEUROBIOLOGY

As I mentioned in Chapter 5.0, concerns about validity particularly with respect to synaptic plasticity experiments have been surfacing in neurobiology over the past 15 years.[10] Yet only several

[10]I want to qualify this statement. I think that initially when approaches to the neurobiological and neurophysiological studies of learning and memory were first being developed, investigators were interested in more complex questions about learning, which involved the use of more complex learning paradigms (e.g., Sharp, McNaughton and

studies in the field (e.g., Sharp, McNaughton and Barnes 1985; Moser 1993; Whitlock et al., 2006) have sought to establish what is to date the best evidence that a causal relationship between a learning event and LTP exists. I want to consider some relevant features of one of these studies, that of Whitlock and colleagues (2006), because they correspond to the idea I put forward that increases in validity may occur as a result of a "radical change" from a current learning paradigm.[11]

In the experiments undertaken by Whitlock and colleagues (2007), male rats were implanted with a multi-electrode recording array positioned in area CA1 of the hippocampus. This position was selected, because the form of learning investigated was inhibitory avoidance learning, which "depends on the hippocampus" (Whitlock et al. 2006, 1083) . The rats were then trained in an inhibitory avoidance (IA) paradigm. In this paradigm a rat is placed on the illuminated side of a 2-compartment chamber. When the rat crosses over to the unilluminated side of the chamber it receives a foot-shock. If the rat has learned the association between the dark context and the foot shock, it will subsequently avoid entering the dark part of the chamber. Immediately following training in this paradigm recordings of extracellular field potentials of hippocampal CA1 pyramidal neurons were recorded in order to determine if training in the learning paradigm induced LTP.

Whitlock and colleagues used the IA paradigm for several reasons that corresponded directly to concerns about reliability. First, they wanted to overcome the challenges faced by using iterative

Barnes 1985 and Morris (review 2003). However, I think such studies had a tendency to lead investigators to appreciate the challenges to studying learning and arriving at definitive claims about its mechanisms. So, I take the neurobiology of learning and memory as it currently stands to be a response to problems of complexity–namely, a shift in the field away from complex to more simple experimental approaches of the kind that Eric Kandel has advocated.

[11] I want to note that the Whitlock et al. study, while it may have some problems of reliability is at least an attempt to move in the direction of a satisfaction of the 2nd condition of validity–that is, making the lab more similar to the world. In turn, I will not be concerned with the methodological limitations of problems with this study. In addition, there is reason to believe that despite the fact that Whitlock and colleagues (2007) claim to be offering a "one-trial" learning paradigm, it may not be physiologically different from a fear-conditioning paradigm. Yet, it is true that when cue and conditioning paradigms are used for molecular studies, they always involve multiple trials of running an experimental subject through a paradigm.

paradigms like classical conditioning "that require many trials to form a memory" (1093) because differences in the rate of learning across rat subjects might disrupt the ability to detect LTP. So, one reason why the IA paradigm was selected was to increase the likelihood of obtaining a detectable, measurable result. However, selecting to use a one-trial stimulus as opposed to a set of repetitive stimuli was a move that also enhanced validity in so far as it corresponds to the idea that both learning and long-term potentiation could result from a one-trial learning situation rather than a multi-trial one. One-trial learning (e.g., Guthrie 1935), is not captured by current iterative paradigms, so the findings from these experiments are not readily extendable to instances of one-trial learning in the world. So, with respect to this experiment, it is not an "incremental change" that has been made with respect to the experimental model but a substantial change that increases the representation of different forms of learning across neurobiological experiments in a way that better reflects the *general* kinds of learning that we encounter in the world.

One of the biggest challenges to establishing a link between learning and LTP in the mammalian brain is showing that LTP is the result of an organism's experience in an actual learning situation rather than that of artificial stimulation applied directly to neurons. Whitlock and colleagues found that training in the IA learning paradigm not only elicits inhibitory avoidance at the level of the behaving organism, but it also elicits a form of LTP that more closely approximates what it might look like when it is generated in the brain of a behaving organism by natural stimulus patterns (one-shot patterns). So, we can say at the very least that the validity of the overarching interpretive claim that LTP is the neural correlate of learning and memory is increased in light of the findings of the Whitlock study.

A complementary approach with respect to increasing the learning-LTP link would be to develop LTP stimulation paradigms that more closely mimicked the patterns of activity present in

learning paradigms such as the IA paradigm. Historically, neurobiologists did try to develop such stimulation paradigms. An alternative stimulation paradigm, theta-burst stimulation, was developed by Larson and Lynch (1986; 2003) to reflect the naturalistic stimulation of "electric odors" at the level of the behaving organism. However, the inability to detect measurable synaptic changes has lead to the use of more reliable as opposed to more naturalistic stimulation patterns. Perhaps efforts could be focused in the direction of overcoming such challenges, even by making a stimulation paradigm look more like the IA learning paradigm.

In light of my proposal, a prescription to increase the validity of this study could take the following form. An investigator runs a second experiment in which the IA paradigm is used and recordings are taken from area CA1. Yet, in this experiment, the rat could be raised in an environment in which it is exposed to a condition that a natural rat may experience in the world. For example, living in a social group with other rats. A third experiment could be undertaken in which individual rats were exposed to an at least minimally enriched environment prior to training in the IA paradigm. In a fourth experiment a rat would be raised in an enriched environment with a social group.

Now, it may be that negative results will be obtained. That once the experiences of a rat are modified in any way above those experiences that a normal lab rat encounters, suddenly no LTP will be detected in area CA1 following IA training. However, such negative results could be interesting in so far as they may lead us in the direction of where to look next to get a grasp on what the neural correlates of learning look like in organisms in the world. Another set of experiments could have as a goal the detection of differences between the brains of rats raised in social groups and rats raised in a more isolated state (with fewer rats proximal to it). Another set could be used to detect differences between the brains of rats raised in enriched compared to impoverished environments.

A third set of studies could combine both approaches and investigate the differences. Clearly with respect to any of these studies, it would be at the discretion of the investigator to weigh the pros and cons of which investigatory avenues with respect to validity to explore first. This is in part why I think that the collaboration across neurobiological laboratories and different areas of neuroscience (anatomy, physiology, behavior) will be essential to determining the order in which such experiments should be pursued and what features they should have. For example, how should we understand what it means for something to be an "enriched" environment? Is it an environment that is more like the natural circumstance of a rat in the world or is it simply enough to provide a rat with toys and a running wheel?

6.4 PROPOSED FRAMEWORK

I conceive of the proposal that I provide for increasing validity in neurobiological experiments within the context of a framework of descriptive and physical models (*Figure 13*). I want to go through the steps of this model with respect to the case that I have been considering throughout the dissertation, namely extending an interpretive claim from a learning situation in the lab to a learning situation in the world with respect to the rat.

In brief, in *Figure 13*, a working descriptive model ("Learning Situation in the World") is developed to capture those general features of learning situations in the world. I have attempted to lay the foundation for one such model in this chapter, by considering several features that differentiate experimental systems in the laboratory from learning situations in the world. I conceive of this as a working model that ultimately will become more detailed in light of taking into account findings about rat behavior from a wide range of areas of science, including other areas of neuroscience,

neuroethology, cognitive ethology, cognitive psychology, behavioral ethology, and critical work done in the history of psychology on learning. Ultimately, neurobiologists cannot go out into the field, they can simply partake in the lessons obtained from those who have.

The "Learning Situation in the World" working model is supposed to inform decisions made about what features of the world should be included in an experimental design and protocol. The basic idea is that an investigator will identify a feature of the learning situation in the world that could be included in the laboratory to increase validity. Then, the investigator will translate that feature into the context of an experimental design and protocol. The experimental design and protocol will then be physically instantiated in an "Experimental Model". The experimenter will then go through the data production process, and, if it is reliable, she will arrive at a claim about the role of the molecule of interest in the effect under study in the laboratory. She can juxtapose this interpretive claim with that of an interpretive claim arrived at using the experimental model before the feature was added. She can then integrate the findings from these two results to say something broader about the role of the molecule of interest in learning situations in the world.[12]

To give an example, one option would be to alter the housing in which the rat is raised in a way to make it more closely resemble a wild rat's natural environment prior to the learning situation. Then in these rats, if this is taken to be a partially "enriched environment", the activation profile of a particular molecule could be investigated and compared against what is already known about that molecule in lab rats raised in unenriched environments. Another option may be for a neurobiologist to begin with a fear-conditioning paradigm but to alter the noxious stimulus in such a way as to make it more like a noxious stimulus that a rat would encounter in the natural world. Then, findings with respect to molecular activity in the two cases could be compared. A third experiment could

[12]David Danks (2005) provides an account of strategies that might be employed to integrate or "stitch together" results across experiments.

involve training the rat in an area less restricted than a Skinner box, which will permit it to exhibit different types of natural responses–e.g., a fleeing response as opposed to a freezing response–and to then investigate changes with respect to a molecule of interest. These simple parametric analyses are in essence "incremental experiments"–incremental primarily because what is achieved with respect to interpretive claims is an incremental increase in validity. This is of course necessary in order to achieve reliability. Parametric analyses are commonplace in neurobiology, but they have not necessarily been used to consider importing features of the world into the lab to determine whether specific molecules are operative in learning under more natural conditions or what their activation profiles look like, or whether there are differences in where they act in a cell.

There may also be more *radical* types of experiments, for example, in which an entirely new learning paradigm is developed that more closely resembles a learning situation that the rat encounters in the natural world. I have no examples in mind of this kind of change for the future. Yet, we find examples in the history of neuroscience of the development of new learning paradigms. For example, in 1973, Richard Morris developed the water maze model for studying learning in rats, which opened up a new field of research of "spatial learning" in rodents. Yet, the learning situation is not one that resembles a natural learning situation that a rat might encounter in the world. Still, this experimental paradigm opened up an interesting new area of research (Please see Appendix for a description of this paradigm)[13]. Yet, if new kinds of learning or stimulation paradigms are developed that more closely approximate natural learning situations in the rat, then incremental changes of the kind that I have described above could be used to make the entire experimental model more like that of the world.

There is a second approach that I have in mind that achieves the goal of increasing validity

[13] As an aside, Morris (2003) claims: "I carried out my last behavioural experiment in an operant chamber in 1972 and have never been tempted back into the world of response rates and schedules of reward and punishment."

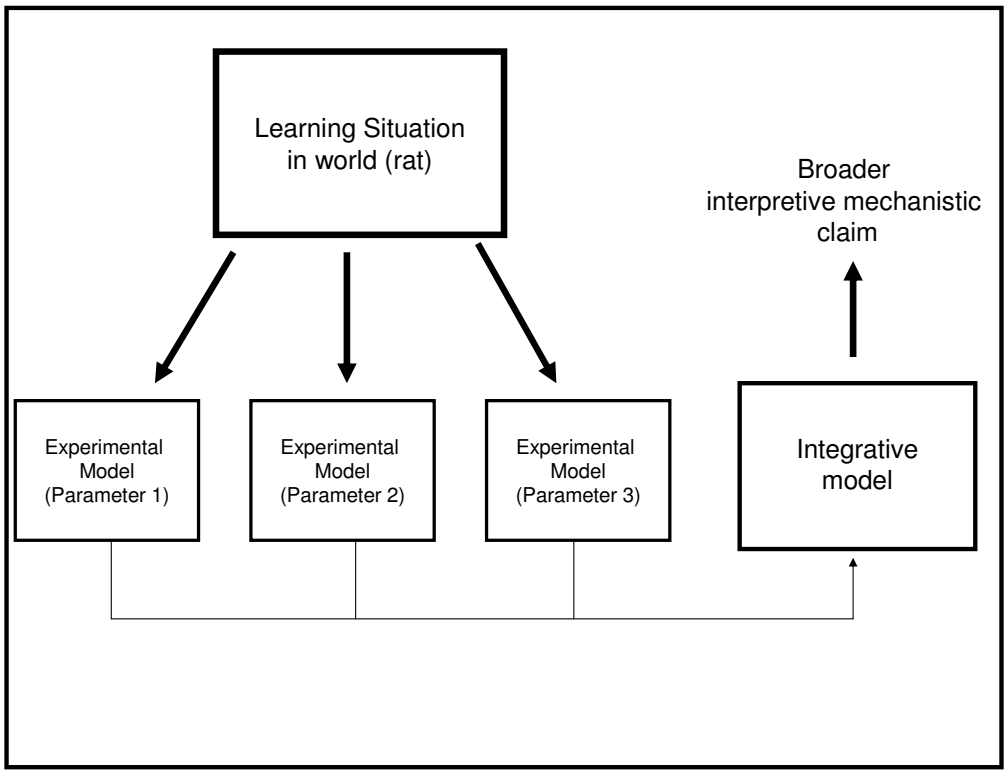


Figure 13: Framework for Increasing Validity: Integrative Model

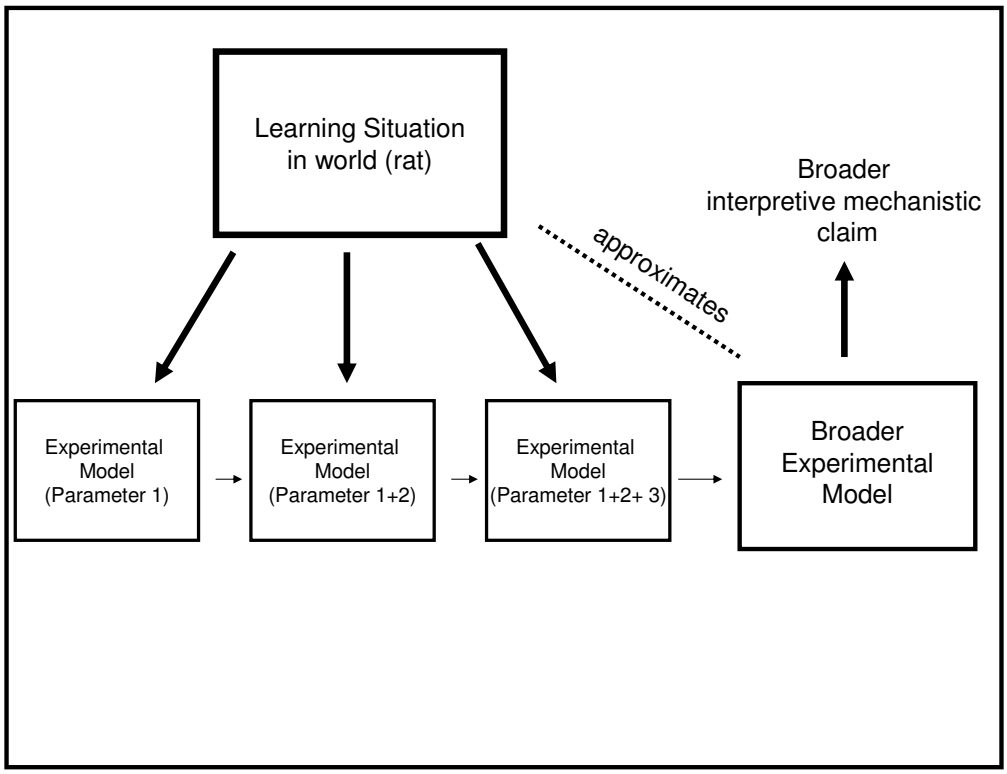


Figure 14: Framework for Increasing Validity: Additive Model

while striking a balance between reliability and validity *Figure 14*). This second approach involves the gradual building of new features into an experimental model across experiments, so that the learning situations contained in the models gradually come to resemble those real-world learning situations of the rat. In this framework, the reliability of the data production process of each individual experiment will be maintained by virtue of the relationship that it bears to all those experiments that have preceded it. The idea is that reliability can be in a sense "monitored" across data production processes, while validity can gradually be built in. If reliability appears at some point to be minimal or completely lost, the experimenter might have to make a change in how a feature of the learning situation in the world has been physically represented in the laboratory, in order to keep reliability fairly stable. Although this may result in a partial loss of validity, the investigator will be further towards achieving her interpretive and explanatory goals by making such moves.

I take this basic strategy that I have outlined to be applicable to plasticity experiments as well. I suggested in section 6.3 above, how changes in the features of the world of the lab rat may result in changes in the brain that will make the plastic features of synapses more similar to those of a rat in the world. Clearly what is problematic in plasticity experiments is the artificiality of the stimulation paradigms being used. Developing new paradigms that better mimic how neurons are activated during learning events by "natural" stimuli would be ideal. But, barring that, there are other avenues that can be explored.

But this brings us back to the question of extrapolation from the rat to the human case. Ultimately I am claiming that by making the lab look more like the world, neurobiology will more closely approximate the mechanisms of learning and memory as they are in the world, rather than as they are in the laboratory. Investigations in model systems like the rat will only bring us closer

to our goals if they look more like us.[14] The point is that they can look more like us than current neurobiology suggests, but of course they will never be perfect models. There are most likely other options for making them look more like us than those that I have identified here. Another possibility for example, would be try to develop learning paradigms that are better correlates of human learning. I cannot say what the features of such paradigms will have to look like, but most likely they will be more complicated than current learning paradigms used in neurobiology. And most likely such experiments will be more interesting in a laboratory context that better approximates the world.

6.5 CONCLUSION

Descriptive models of humans or rodents, human and rodent learning, and the environment as humans and rodents encounter it in *the natural world* are not the starting points for neurobiological experiments. The starting point is rather a set of descriptively simple models of the laboratory organism, the environment, and learning (that might be taken to form a "theoretical core" of neurobiology) that either do not correspond at all to the world (i.e., they are not, what Cartwright refers to as "representative models" (Cartwright 1999, 180)) or correspond to it only very roughly. Such models leave a vast number of details of environments, organisms, and learning as we encounter them in the world out of the picture.

Yet, this starting point is not exclusive to neurobiology–it puts it on a par with other areas of science that begin their experiments with very idealized models that bear little to no resemblance

[14] What I mean by "look like us" is more than mere "face validity". The idea is that there will be complex internal changes in these organisms as a result of altering certain features in their environments.

with the natural world that is the target of interest (including physics, economics, and evolutionary biology). Second, given the complexity of the subject matter, namely, learning in organisms in environments, we can anticipate that some degree of idealization will always be necessary in neurobiology, with respect both to models of experimental and natural systems. What is implicit in criticisms of areas of science that make *global claims* about phenomena is that these areas of science could better approximate a more accurate representation of reality, i.e., achieve their global aims, if they took more of the world on board in experimentation. In this chapter, I have sketched one proposal for neurobiologists to move in this general direction.

The solutions that I have proposed in this chapter for increasing the validity of experiments in neurobiology have obvious drawbacks. First, the piecemeal process of adding in parameters in experiments will be extremely time consuming and require far more resources (money, organisms, techniques and methods, apparati) than current neurobiological experiments on learning do, which makes what I am suggesting a very unattractive option. Furthermore, in order to achieve success with respect to this new overarching goal of experimentation, to cut-time and cut-costs, neurobiologists most likely will be required to engage in collaborations across the community–something that will take an element of individuality away from science that is very important. In my opinion, however, resolving what I have identified as the current crisis in neurobiology should encourage neurobiologists to join in a concerted effort to increase the validity of their interpretive claims.

7.0 CONCLUDING DISCUSSION

The generating circumstance of this dissertation was to understand the nature of what I took to be the primary "crisis" in the contemporary neurobiology of learning and memory–namely, that the cellular and molecular mechanisms currently on offer for explaining organism-level learning were insufficient given the simplicity of the learning paradigms and concepts of learning used in experimental neurobiology in contrast to the complexity of the learning phenomena of interest. Essentially, the aim of the dissertation was to get a handle on the foundations of experimentation in neurobiology, in order to provide solutions to the current crisis.

My way of approaching this problem was to characterize the experimental process and identify its specific features and then to assess at every step of the way in neurobiological experiments, what constrained it. What I found was that in neurobiology the desire for the production of data to be reliable far outweighed the desire to be able to legitimately extend causal claims beyond the context of the laboratory. Two options were available for addressing this problem: (1) allow neurobiology to persist as is or (2) characterize the problem in a way that afforded the development of solutions.

To attack the problem head-on is to call into question the very assumptions that drive research in the neurobiology of learning and memory. In this dissertation, I took a less direct approach and what I conceived to be a more constructive one, by pointing out that given the relationship between reliability and validity, while achieving the goal of reliability in scientific experiments is a rational

goal, when it becomes the exclusive goal, it threatens the possibility of neurobiology arriving at interpretive mechanistic claims that will be applicable beyond laboratory circumstances.

The solution as I outlined it, is essentially as follows. To achieve the goal of validity across experiments on learning, neurobiologists must engage in parametric analysis across experiments for example, on the role of a specific molecule or signaling cascade in learning, by varying individual parameters across experiments and then combining the experimental results in an integrated model. This model will better capture the world than any model that accompanies the individual experiments in which one parameter is varied. Reliability is essentially maintained with respect to each individual experiment, and the validity of interpretive claims is achieved gradually across experiments.

7.1 PRIMARY CONTRIBUTIONS

I see this dissertation as making two primary contributions: one to the philosophy of experiment and the other to the philosophy of neuroscience. First, the issue of how reliability and validity are related in the context of experimentation has not been explicitly addressed in detail in the literature and I offer a conceptual framework for understanding the experimental process and the relationship between reliability and validity that can be readily applied to experiments in other areas of science.

With respect to the philosophy of neuroscience, to date, no one has considered in detail the epistemological problems that arise with respect to experimentation in neurobiology. I fill in this gap in the literature. My primary hope is that this dissertation may serve as a starting point for engaging in discourse with experimentalists in neurobiology about the epistemological problems that arise in the context of experimentation and how those problems may be solved.

There are a lot of aspects of this dissertation that require further development. First, my analysis of validity, its types and parameters is superficial at best. It served the purpose here of illuminating some of the differences that we encounter when we compare a model of the world and its contents to the features of the world that are actually captured in the physical context of the laboratory. But the context of the laboratory can itself be represented in a model, and clearly when we integrate experimental results, we do not do this in any physical sense, we do it in the context of a broader model, which requires us to think of each individual experiment in which a new parameter is introduced as a local causal model. But clearly the relationship between models, experiments and the world requires an analysis more detailed that I have given it in this dissertation. What I have done is laid the groundwork for the development of such a project in the future.

8.0 APPENDIX

On Squire and Kandel's ([1999] 2003) definition, *non-declarative* memory is taken to be "revealed through unconscious performance", "inaccessible to conscious awareness", and "inflexible". It is divided into four different subtypes, which I will consider in the following order: (1) *nonassociative learning*, (2) *simple classical conditioning*, (2) *priming and perceptual learning*, (3) *procedural* (skills and habits). If these categories are taken to be both mutually exclusive and exhaustive with respect to what is known currently, presumably all forms of *nondeclarative* learning not classified under *nonassociative* learning are taken to be *associative* forms of learning.

In *non-associative* learning, an organism learns about the properties of a single type of stimulus (in contrast to associative learning, in which a relationship is taken to be established between at least two stimuli). At least three types of nonassociative learning have been identified. The first type, *habituation* is defined as "learning to recognize and ignore as familiar, unimportant stimuli that are monotonously repetitive"(Squire and Kandel [1999] 2003, 26) and as "a type of learning that operates to eliminate inappropriate or exaggerated responses". *Sensitization*, on the other hand is defined as an altered response to a (one) stimulus as a consequence of exposure to some other, usually noxious stimulus. *Dishabituation* is not discussed by Squire and Kandel ([1999] 2003), so I will leave out a detailed discussion of it here. It essentially involves the termination of a habituated

state as the result of a sensitizing stimulus.[1]

Habituation and sensitization have been studied most extensively in simple reflex pathways in invertebrate model systems. The most celebrated cases are habituation and sensitization studies of the gill-siphon withdrawal reflex in *Aplysia Californicum*, the sea mollusk.(See Schaffner (1993) for a philosophical analysis of this case study.). Typical habituation studies involve repeated application of a tactile stimulus to the organism's tail, which results in retraction of the gill and siphon into the mantle shelf, underneath the organism. The defensive reflex begins strong to the first application of the tactile stimulus, but the response decrements after repeated presentations to the point at which the measured extent of the actual reflex is negligible. In sensitization studies, baseline response to the tactile stimulus is first established and a tail shock is later introduced. After this sensitizing stimulus, the animal shows an increased reflexive response to the tactile stimulus that lasts over multiple trials, and gradually decrements towards baseline response after repeated trials in which the tactile stimulus is applied.

How might these types of non-associative learning be extended? Often times terms like "habituate" and "desensitize" are employed rather loosely. For example, we might say that one can habituate to the sound of a lawn mower running across the front lawn while one is reading a book in the sunroom. However, the similarity between what the term captures in the controlled environment of the laboratory and what it captures with respect to the everyday experiences of more

[1] One thing that is of interesting note with respect to these two learning types is that while habituation seems to fit the profile of nonassociative learning, sensitization, in contrast, involves learning some kind of temporal relationship between two discrete stimuli as opposed to one discrete stimulus. So, it is not obvious why it should be considered as involving learning about a single stimulus when two different stimuli are relevant for the type of alteration in response that is said to constitute this type of learning. The fact that it is not classified as an instance of *associative learning* raises the interesting question of what kinds of conditions a given scientist believes must be satisfied in order for a type of learning to count as associative, or as involving the learning of a relationship between two stimuli. And an interesting question is, even if these conditions are met, does the type of learning in the two cases amount to basically the same type of thing?

complex organisms appear to be radically different things.[2]

The second type of nondeclarative learning that Squire considers is "simple" classical conditioning. In this paradigm animals are assumed to learn a predictive relationship between two stimuli or the predictive value of one stimulus for another; or, under another description, animals learn to associate two stimuli. This type of learning can be studied in any animal just so long as they elicit a simple measurable reflex that may be subject to conditioning.

Squire most likely places classical conditioning in the category of non-declarative learning because the original reflex that is conditioned in a classical conditioning experiment is not under conscious control (e.g., salivation, startle or reflexive fear response: reflex, eye-blink reflex)–and since this reflex is not under conscious control, then the assumption is that if it is conditioned to another stimulus, its elicitation in response to that stimulus cannot be brought under conscious control–so this type of learning is thought to be unconscious and the elicitation of what is learned is thought to be unconscious as well (or not under conscious control).

Squire subdivides this classical conditioning category into emotional responses and skeletal responses, which correspond to two subsets of experiments undertaken in rodents (fear-conditioning paradigm), rabbits (eye-blink/nictitating membrane reflex) and humans (eye-blink). In both types of experiments variations on the traditional Pavlovian model of classical conditioning are employed.

What is simple classical conditioning? According to the traditional model, classical conditioning paradigms begin with a simple innate reflex (an unconditioned response or UCR) that is elicited

[2]For one, it does not seem like we are dealing only with a simple reflex in this case. Two, one cannot exert the same kind of control in the world outside of the laboratory, and so, extension of these terms to other than simple reflex pathways will be more complicated, and will most likely involve more than just simple mechanisms–simple sensory-motor circuits or reflex pathways. [If this criticism is true of the most simple forms of learning, one can imagine the kinds of difficulties that we will encounter when we try to approach more complex forms of learning. [In turn, it will most likely be difficult to isolate as pure a form of non-associative learning in more complex organisms.]

by an animal in response to a stimulus (an unconditioned stimulus (UCS)). The stimulus that will elicit the reflex is then paired concurrently and repeatedly with another stimulus (conditioned stimulus CS) that would not elicit the reflex under normal conditions (e.g, salivation does not usually happen when a bell rings). In a successful classical conditioning experiment, after concurrent and repeated pairings of the UCS and CS, the CS, when presented alone, is found to elicit the reflex (conditioned response CR).

So, on the traditional model, there will be several things at issue. First, a simple measurable reflex must be selected for experimental study. Second a simple stimulus that elicits it must be identified and a stimulus that does not elicit it under normal conditions must be identified. Third, the induction paradigm has to satisfy the conditions of (1) contiguity and (2) repetition. Fourth, the reflex response measured after training must be equal than or greater to that measured in response to the stimulus that elicited it in the first place.

If we consider the experimental paradigms that are all thought to qualify as examples of classical conditioning, only some of these paradigms will fit this model. Examples that do not fit the model will fail to fit it with respect to one or more of these four conditions. Although the last two conditions are seemingly unproblematic, the first two are clearly less so, both taken independently and taken together. It is not clear what it means for something to be a "simple stimulus", since most stimuli are complex in so far as they may possess properties that act on multiple sensory receptors for different modalities simultaneously. Similarly, they may be complex even with respect to only one sensory modality, acting on different receptors each associated with that modality but which each carry different sensory information. It is extremely difficult to design controlled experiments that contain stimuli that only operate with respect to one sensory modality or one sensory-motor pathway. So, it will be particularly difficult for certain experiments to fit the traditional model of

classical conditioning with respect to the choice of unconditioned and conditioned stimuli. Second, some experiments will fail to fit the model due to the fact that the type of response under study does not appear to be a simple innate reflex. We run into problems here, precisely because it is not clear how we should specify what it means for something to be a reflex.[3]

The next category that Squire places under nondeclarative memory is a combination of two learning types: *priming* and *perceptual* learning. These two forms of learning appear to be grouped together primarily because they are thought to take place in the neocortex during the early stages of perceptual processing, immediately downstream of sensory receptors in early cortical areas prior to "higher-order" processing. Otherwise these two forms of learning have very little in common.

Priming is essentially the ability to recover recently studied visual information, without awareness of previous or recent exposure to such information. So, one is not aware of having learned anything, nor can one consciously recall what has been learned. Rather, what has been learned can be shown to be elicited, but unconsciously. In a classic example of a priming experiment, subjects are presented with a set of words (e.g., Motel), and are later exposed to parts of those same words (e.g. "mot") and are asked to name the first word that comes to mind. Typically subjects will elicit those words to which they have been recently exposed. In another task, a picture/word-naming task, a subject is shown a picture of an object (e.g., an airplane) and asked to name the object in

[3]Pavlov himself proceeded with some difficulty with respect to the notion of *reflex*, at least theoretically. In his theoretical work he claimed that *all* behavior (behavioral phenomena) is comprised exclusively of reflexes, some simple and some more complex. He attempted to provide a taxonomy of different reflexes in much the same way that contemporary neuroscientists classify different types of learning. If it is true that all behaviors are reflexes, then it would make sense that all reflexes could be subject in principle to classical conditioning. However, in practice Pavlov thought that only "pure" and measurable reflexes could be subject to experimentation, and of these he identified relatively few, in fact, his more celebrated experimental work concentrated on only one: the alimentary reflex in dogs. He took it to be a simple, innate, measurable reflex capable of being elicited in response to definite application of food to the taste receptors in the mouth. A little less celebrated was his work on the defensive reflex in dogs, which is, like fear-conditioning paradigms, still employed today.

the picture. The delay in answering is measured and compared to the delay in naming the object in the picture each time the subject is presented with the same picture later during the same task.

Although it may be wrong to call priming a form of learning (it seems more like an immediate and perhaps momentary retainment of information in the visual system), it would seem as if the mapping between the experimental paradigms used to study priming, and the definition and position in the taxonomy all appear to be on target. In other words, if priming has any place in Squire's taxonomy, if it is indeed an unconscious phenomenon, then it is rightly placed under the nondeclarative category.

Perceptual learning, on the other hand, is defined as an "improvement in the ability to discriminate simple perceptual attributes, such as tones or line orientations, simply as the result of performing the discrimination repeatedly." (Squire and Kandel 2003 [1999], 164) It is considered to be a form of nondeclarative learning because it involves changes that "accrue gradually over time in the machinery of perception" that are not accessible to conscious awareness (Squire and Kandel 2003 [1999], 164). Like priming, such changes are thought to accrue in the early stages of perceptual processing in cortical structures that first receive information about the external world (Squire and Kandel 2003 [1999], 164).

A typical paradigm for testing perceptual learning is the "circles illusion", in which subjects are presented with two different arrangements of objects side-by-side. One side contains an arrangement with a flat disk surrounded by a circle of smaller flat disks. The other contains a central disk of the same size that is surrounded by a circle of larger disks. Which side a given arrangement is on varies from trial to trial. Although both center disks are the same size, when surrounded by smaller disks, the center disk looks larger than when it is surrounded by larger disks. Subjects are asked to pick up the central disk from one side when the two disks appear to be the same size,

and to pick up the center disks from the opposite side when the two circles appear to differ in size. Subjects are required to use their thumb and forefinger to pick up the central disk and the distance between the forefinger and thumb (the grip aperture) was recorded and measured. Results obtained from this paradigm suggest that subjects succumb to the illusion in so far as their grip aperture corresponds directly with whether they take the center disk to be larger (when the surrounding disks are smaller) or smaller (in the reverse case).

There are a variety of aspects of this and other perceptual tasks that involve discrimination that would seem to make it presumably difficult to categorize this task as a pure perceptual task that involves only cortical areas involved in the earliest stages of visual processing. The task engages both the perceptual and motor systems, it requires subjects to make a finger-eye coordinated judgment about disk sizes. Even though subjects appear unable to master the illusion because of the perceptual information coming in and updated feedback from the motor system, it does not provide grounds upon which to ascribe learning in this case exclusively to the perceptual system. This task raises the issue of the "credit assignment problem", that is, to which brain area are we entitled to ascribe the learning when a task seems to involve aspects that tax *at least* two different brain systems simultaneously. It would seem that the lack of rigor that can be achieved with respect to such tasks that are purely *perceptual* in nature makes it difficult to isolate this as a pure learning category. Furthermore, it is not clear that this task does not involve any degree of conscious awareness of the properties of the stimuli that are the objects of the tasks. Although I do not want to go into further detail about this here, it may at least be said that this task clearly differs from the priming task where we may be more tempted to say that subjects are not "aware" of what has been primed. The interesting question, then, is: if perceptual learning is truly different than priming, should it be in its own proper category under nondeclarative learning, should it be in another available category

or should it not be classified as a form of nondeclarative learning at all? (For example, maybe learning how to discriminate is an example of procedural rather than perceptual learning; maybe there is no purely perceptual learning.)

It is interesting that Squire only identifies priming and perceptual learning as part of an independent category in the taxonomy when in his written work (e.g. Kandel and Squire [1999] 2003) he mentions at least one other type of learning that is not obviously a case of either priming or perceptual learning and yet, based on the description of the type of learning that is taking place, the kind of structural changes in the brain that are supposed to accompany it are identical to those thought to accompany priming and perceptual learning. This type of learning is *emotional learning*, but clearly, the conditioning paradigms that are used to induce emotional learning have a different structure than those that are used to prime or induce perceptual learning.

The final type of learning that Squire places in the non-declarative learning category is *procedural learning*. It was initially the case that declarative learning was directly contrasted with procedural learning as the difference between *knowing that* and *knowing how*. Yet, as more learning types emerged in the literature that were considered to fall under the category of "non-declarative" memory, procedural learning became just one special instance of nondeclarative memory.

Procedural learning is perhaps best construed as learning how to perform a procedure. The resulting "know-how" is thought to be capable of unconscious expression, through doing, and is oftentimes characterized as unavailable for declaration. As a category of learning it is supposed to include any type of skill that might be learned primarily through repetition or repeated performance of an action to the point at which it achieves the status of a skill. Yet, Squire also lumps "habits" into this category, as if he takes learning how to engage in repetitive activities to be an instance of learning how to perform a procedure. So, the procedural category is taken to include: (1) motor

skills, (2) perceptuo-motor skills, (3) skills (general) and (4) habits. There are specific experimental learning paradigms that correspond to each category. In what follows, I want to consider each of these categories and the tasks with which they are associated.

Motor skill learning is defined by Squire and Kandel ([1999] 2003) as learning "to acquire a procedure for operating in the world" or "learning to perform coordinated movements". One can think of commonplace examples of the kinds of things we have in mind when we talk about motor skills: things like playing tennis, flying a kite or riding a bike. A common experimental paradigm used to study motor skills is the reverse mirror drawing task in which subjects must trace two outlines of a star while viewing their hand in the mirror as they draw the star. The task is thought to induce learning by means of repetition; if subjects, over repeated trials improve in accuracy in tracing the star and the amount of time taken to trace the star decreases, then they are said to have learned a new motor skill.

It is rare that tasks designed to teach subjects a specific type of motor skill involve exclusively the motor cortices that comprise the motor system in the mammalian brain. Even a simple task such as learning how to trace a star in a mirror involves both the sensory and motor cortices. So it is not clear that it is right to call a skill exclusively "motor" when all skills seem to involve both sensation or perception and motor action, and it would seem that the learning will be happening within a sensory-motor circuit rather than just in the motor system. [In fact, it does not seem like there could be such "purely" motor tasks.]

The definition of perceptuo-motor skill learning is identical to that of motor skill learning, however there are specific tasks that are designated as "perceptuo-motor tasks" rather than just simply "motor" tasks. As I mentioned above, it seems more likely that most of the tasks used to test skills will be accompanied by learning that involves the integrated activities of the perceptual

and motor systems. So it just might be the case that there is a certain degree of sloppiness and inconsistency in how these different terms are applied and maybe that term that better captures the nature of the learning under study ought to be employed and the other terms discarded.

One example of a perceptuo-motor skill paradigm is the sequence learning task. In this task, subjects are presented with a visual signal: an asterisk on a computer screen. The screen contains 4 blank dashes at the bottom that correspond to 4 possible locations in which the asterisk may appear. Subjects must press lettered keys that correspond to each of the four locations (A,B,C,D). Subjects are trained over up to 400 trials, which means that they repeat the same sequence of 10 key presses 40 times. The speed with which the subjects press the keys is compared across training trials. Improvement in the speed with which the keys are pressed accurately in response to the cues is taken as indicative of perceptuo-motor skill learning.

Perhaps there is nothing problematic about claiming that this task induces perceptuo-motor skills in subjects. Yet, the description of this learning paradigm does not differ substantially from at least one of those used to study what Squire identifies as "habit learning", causing one to wonder why he attempts to draw a terminological distinction between the two forms of learning. I will consider this issue in the context of looking at habit learning.

On a final note about skill learning, Squire and Kandel identify a subset of other skills which Squire wants to classify as neither motor or perceptuo-motor skills. These types of skills are more cognitive tasks, such as learning a language or learning how to read text on a page. He also claims that these kinds of skills, perhaps in contrast to other kinds of skills, "involve the development of a general sense or intuition about how to proceed on a task".

Whereas the above types of learning are all classified as "skills", there is another type of learning that Squire and Kandel identify as "habit learning", which they claim is "different from

remembering the experience in which a particular type of skill was learned". Furthermore, they claim that "it does not require obvious effort and does not require us to notice that the learning is taking place", which may differentiate it from some of the above types of skills, which presumably require some concentration for learning to occur.

There are experimental paradigms that are used in humans, monkeys and rodents to induce habit learning. I here want to concentrate on the task that was used by Squire and colleagues (Knowlton, et al. 1996) to dissociate the declarative memory system from the procedural memory system. The task is called the "Weather-prediction game", and as it is fairly involved, certainly more involved than the other tasks that fall into the procedural memory category. I want to describe it in detail here. In it, an experimenter attempts to demonstrate that subjects with brain damage restricted to a given brain area fail on one aspect of a task while they are successful on another aspect, whereas subjects with brain damage to a different brain area exhibit the opposite type of performance. One task that has been used to affect a double dissociation between the declarative/explicit memory and procedural/implicit memory systems is this "weather-prediction task" (Knowlton, et al. 1996). In this computer-based learning task, subjects are presented with four tarot cards. Each card contains a unique pattern of a particular geometric shape. On each trial, subjects are presented with a stimulus that consists of 1, 2 or 3 of the 4 cards presented in a specific spatial order (14 different card patterns in total), and subjects must press one of two buttons designating (1) sun or (2) rain, to signal their answer that the cues are predictive of either sun or rain. Correct responses result in feedback in the form of a high-pitched tone that is accompanied by a bar on the left-hand side of the screen that increases by one unit. Incorrect responses are followed by a low-pitched tone and a decrease in the bar of one unit. Each tarot card is probabilistically associated with the two weather patterns and these probability relations are worked out over 50 trials. For example, the

probability relationships used in the Knowlton study and in the current study were: squares (sun: 77%; rain: 23%), triangles (sun 58%; rain: 42%), circles (sun 42%; rain: 58%) and diamonds (sun: 23%; rain 77%). Subjects begin naive to the probability relationships of the task and they begin by making guesses. The task is thought to "defeat the normal tendency to try to memorize a solution, and individuals can learn without being aware of the information they have acquired" (Knowlton, et al. 1996). In turn, the task is taken to be an implicit learning task. In support of this claim is the fact that normal subjects cannot achieve a performance of more than 20% above chance levels on the task. One might reason that with memorization, performance levels in normal subjects would be higher, and since they are not, memorization must not be involved. Subject performance is evaluated over 50 trials, and, after the task, an eight-item multiple-choice questionnaire, taken to be indicative of declarative memory, is distributed to subjects. The test contains questions about the layout of the computer display, the cues, and the testing procedure. If, over time, subjects become better able to predict the weather using the visual cues by performing significantly above chance in so far as they make fewer errors, then they are said to have learned implicitly the probabilistic relationships between the cues and the weather–they are said to have successful habit/implicit learning. If subjects do not improve, then they have not learned how to perform the task and most likely their ability to learn implicitly is impaired. If subjects when queried about features of the training episode can answer those questions correctly, then they have intact episodic memory, if they cannot, then they are said not to have intact episodic memory.

It seems likely that Squire and Kandel ([1999] 2003) lump skill learning and habit learning together under the blanket of procedural memory because these types of learning are all thought to involve the striatum. Yet, processing in the striatum is a stage in the processing between perception and action and so, it seems we once again run up against the credit assignment problem and in part

primarily because of findings from subjects who have damage to the striatum who fail to be able to successfully perform these tasks.

The contrast of non-declarative memory is declarative memory. Squire and Kandel (2003) define *declarative memory* as "memory for events, facts, words, faces, music–all various pieces of knowledge that we have acquired during a lifetime of experience and learning; knowledge that can potentially be declared that is, brought to mind as a verbal proposition or as a mental image" (Squire and Kandel 1999] 2003, 70-71). They emphasize the fact that this memory can be consciously recalled, but no emphasis is placed on whether so-called declarative learning (i.e., encoding as compared to recall) is conscious or not. Declarative memory is typically divided into two subtypes: (1) memory for facts, sp., *semantic memory* and (2) memory for events, sp., *episodic memory*. Both forms of learning are declarative, because the information learned can be made conscious and somehow declared (i.e. verbally or in written form). However, they differ in so far as in the case of episodic memory, a subject is required to remember particular past events. Squire and Kandel describe both types of declarative memory as essentially associative, although the types of associations forged in each case differ. In some ways these types of learning are considered to be more conceptual in terms of the fact that what is being associated are comprised of "stimulus-stimulus" associations as opposed to the kind of associations that appear to occur with respect to those experimental paradigms that are used to study non-declarative memory. It appears that when the motor system is involved in a learning event, it will most likely be a non-declarative memory, but when it comes to the formation of concepts or thoughts, the associations are thought to be purely between stimuli.

The common paradigms used to study semantic memory in monkeys are the "delayed non-matching to sample" and the "delayed matching to sample" paradigms. The aim of these paradigms

is to induce specific kinds of associations between stimuli in monkeys, so that when prompted, they can exhibit whether or not they have formed these associations. Both are typically thought of as working memory tasks, although they are often used in monkeys with hippocampal lesions. They require that a monkey subject hold in ”working memory” a memory of a stimulus during a delay period between stimulus presentations and to differentiate it from (“non-match”) or match it to (“match”) another stimulus in order to receive a food reward. Since the animal clearly has to learn what is requisite of it on the task during the training session and prior to the testing session, it is difficult to say exactly what is being learned.

In one example of a non-matching task, a monkey is presented with an object. Then a screen is lowered, and when it is raised, two objects are visible: the one previously presented and a new one. The monkey has to choose which one is novel in order to receive a food/juice reward. In one example of a matching task, a monkey is presented with a color sample (flash), then a brief delay occurs and the monkey has to pick which of 2 sample colors is the same as the one shown previously. If it selects correctly, it receives a food reward. If the monkey guesses correctly in either case, then it is said to have learned an association between the reward and the novel object or to have learned the prediction that picking the novel object will result in a reward.

A variety of other tasks are also taken to test declarative semantic memory. The first I want to consider is the Morris water maze. In this *spatial learning* task, mice or rats are placed into a circular shallow pool of opaque water and are required to swim in order to learn to locate an escape platform hidden just below the surface of the water. The pool is typically placed in a room with distal spatial cues (e.g., pictures on the surface of the walls of the room) or local spatial cues (such as pictures of objects on the wall of the pool). Since rats prefer to be on a stable platform out of the water, they are motivated to find a way out of the water (i.e. to find the platform). A

typical training protocol generally consists of 2 blocks of 4 training trials a day with an inter-block interval of approximately 1 hour. Subjects are released into the pool from 1 of 4 starting positions, and the location of the platform remains constant throughout training. The training is given for about 6 consecutive days. Time taken to find the escape platform is measured. To determine if the rat is actively using the available spatial cues (which is said to amount to "using a spatial learning strategy"), probe trials are later undertaken in which the platform is removed and the rat is placed in the pool for 60 sec to search for it; the amount of time spent in each of the four quadrants of the pool is calculated and crossing the exact position where the platform was during training is also quantified. Amount of time spent in each quadrant is compared. It is supposed that more time spent in the quadrant in which the platform was during training have learned the location of the platform (by the use of the spatial cues). In addition, the swimming behavior of a rat or mouse during the probe trials can also be assessed using devices that allow the measurement of total path length and swim velocity for individual mouse or rat.

Mice and rats that display significant improvement in their performance in locating the hidden platform over several blocks of training trials, as assessed by the animals' escape latencies, i.e. the time taken to locate the escape platform, are identified as having learned the location of the platform. However, on probe trials, correct search for a longer period of time in the quadrant in which the platform was during training is taken as the best indication overall of the use of spatial cues to learn the location of the platform. Richard Morris himself has demonstrated that the task overall is cognitively quite complex and can be experimentally dissociated into at least two components. One component is learning the task, i.e. that there is a platform, that spatial cues are relevant, etc. A second component is learning the specific location of the escape platform. Some types of lesions can lead to a loss in an animal's ability to learn the task, while not affecting the

ability of the animal to learn a specific platform location, for example.

Another spatial learning task, the Barnes maze, consists of a thick well-lit round circular table with small holes punched out all around the periphery. Rodents are placed in the open well-lit space, which they find aversive since they prefer dark spaces, which forces them to run around on the platform looking for a way out of the light. All of the holes lead to the floor except 1 that leads to a darkened chamber into which the rats/mice will escape from the light. There are visual spatial cues on the walls surrounding the maze, and the location of the escape hole is constant with respect to these spatial cues. An animal is said to have learned the location of the escape hole the fewer errors it makes before finally reaching the hole, or the more rapidly it navigates to the hole.

Typically the Barnes maze is used to establish hippocampal dependent spatial learning and is used in conjunction with or as an alternative to the Morris water maze in adult rodents that are not fit enough to swim or to rule out that mobility impairments are the reason for a lack of spatial learning.

Two other kinds of spatial learning connected with reward are the 4- and 8-arm radial mazes with distal cues. In either of these 2 types of mazes, rats or mice have to learn to locate food rewards in various limbs of the maze that is set up on a platform. Each maze has a set of 4- or 8-arms and is placed in a room with 1 unique distal visual cue on each wall. A food reward is placed at the end of one or more of the arms of the maze. The animal is meant to learn to use the spatial cues to navigate to food rewards. Errors that the animal makes in locating the food rewards are counted over trials. Also, delays between training trials are used to differentiate short-term from long-term memory. When an animal makes few errors in locating food rewards, or shows a reduced latency in locating food rewards, then the rat is described as having learned to use spatial

cues to navigate the maze.[4] Findings from these paradigms are used essentially to substantiate claims about the role of the hippocampus in spatial learning; claims about the type of information that hippocampal cells encode.

Another maze, the 4-arm radial Maze, contains local (proximal) cues and ix used to train animals in order to determine what kind of information hippocampal ”place cells” encode or what type of information is processed by the hippocampus. This paradigm consists of a 4-arm radial maze with distinct (1) visual, (2) tactile or (3) olfactory cues placed in each of the 4 arms, in addition to distal cues placed on the wall. Additionally, rotation of the maze experiments are used to determine how this will affect the firing of cells in the hippocampus. In rotation experiment, animals are trained in a multi-cue maze that contains both local and distal visual cues first, and then the maze is rotated such that the visual cues are then located 90 degrees counter-clockwise and the local cues 90 degrees clockwise to the original position in order determine effects of the firing of cells in the hippocampus.[5]

Another task that falls into the declarative learning category is the odor-association task: variously referred to as ”contingency learning” or the ”learning of transitivity over paired associates”. The task is thought to build on associative learning and still be a form of declarative learning. The task consists of small cups of sand. During preliminary training, rats learn to dig into the sand

[4]If hippocampal place cells fire dependent upon the animal’s location in space relative to the distal visual cues, then that means that they are coding information about spatial location relative to these cues. In addition, if they fire selectively when the animal is moving either outward into an arm or inward back toward the center of the maze, they are encoding the animal’s position in space relative to the ends of the arms of the maze, which is again associative.

[5]If the cell fires when an animal is in a select position with respect to any of the 4 different kinds of cues, that neuron is said to code for that type of cue. If a cell fires when the animal passes a tactile surface, then it is a texture encoding cell. If it responds with respect to a distal cue or a local visual cue, then perhaps it is a cell that fires to distal cues. Some fire only when all modalities of information are present: or an aggregate representation combining all of the cues. This data has been interpreted as supporting the idea that hippocampal pyramidal neurons are place cells, but they also can be texture cells and olfactory cells, and they also can be place + texture + olfactory cells–they are multi-modal association cells that are involved in encoding a wide variety of contingencies and relationships.

cups in order to obtain food rewards. Later, rats are exposed to sand cups with up to 9 different odors mixed into the sand (like thyme, cumin, cinnamon). On each trial the rat is presented with a sand cup with a different odor (odors are presented in sequential order). It will only encounter a sand cup that contains a food reward if that sand cup has an odor that differs from the odor of the sand cup presented on the previous trial. So it is supposed to learn to dig only when the smell is different. Later, it is presented with one sand cup (the sample) that has an odor, and on the subsequent trial, with two sand cups with different odors and it has to learn to dig in that sand cup that has an odor that it is supposed to learn to associate with the sample. So, for example, it may learn that if the sample has the smell of cinnamon, then, when it is presented with the odor of cumin and thyme, it should always dig in the sand cup that smells like thyme in order to obtain a food reward (presumably this requires trial and error learning). They are trained on 3 problems of this type (e.g., if A then B, if B then C, if X then Y, if Y, then Z). Once learning is established, they are then tested for flexibility and they are presented with an odor such as cinnamon (A) and then asked to select between C or Z (e.g., if A then C or Z?; or if X then C or Z?). In this task, if the rat learns to dig in the appropriate cups, then it is said to have learned the requisite associations, and if it exhibits transitive learning by digging in the appropriate cup, then it has learned how to flexibly use associations it has made.

In contrast to semantic memory, episodic memory is difficult to study in non-human animals because they lack language and there is no easy way to test whether they have memories for where and when information was acquired, which is typically referred to as “source memory”. In turn the best paradigms that we have to test episodic memory involve human subjects. A typical case of such a paradigm is reflected in the ”Weather Prediction Game” (mentioned above; the Knowlton et al. 1996 study) in which subjects are asked specific questions about the testing episode during

which they learned how to perform the task. Most episodic learning tasks contain questions asking subjects about their memory for a particular episode or event.

Episodic memory, the other form of declarative memory identified in the taxonomy is characteristically taken to be a type of memory unique to humans that plays a role in identity formation and in the maintenance of a coherent notion of the self. It is memory. Because it is supposed to be uniquely human, and even if not unique, characteristically difficult to study in non-human animals, few experimental paradigms exist for its study.

Of the aforementioned paradigms, only a subset of them are used in modern neurobiology of learning and memory.

BIBLIOGRAPHY

Achinstein, P. (2001) *The Book of Evidence*. New York: Oxford University Press.

Achinstein, P. (2000) "Why Philosophical Theories of Evidence Are (and Ought to Be) Ignored by Scientists", *Philosophy of Science* 67:3 (*Supplement*): S180-S192.

Achinstein, P. (1983) *The Concept of Evidence.* New York: Oxford University Press.

Ankeny, R (2001) "Model Organisms as Models: Understanding the 'Lingua Franca' of the Human Genome Project", *Philosophy of Science* 68: 3 (*Supplement*): S251-S261.

Ankeny, R. (2000) "Fashioning Descriptive Models in Biology: Of Worms and Wiring Diagrams", *Philosophy of Science* 67: S260-S272.

Aronson, J., Harre, R. and E. Cornell Way. (1995) *Realism Rescued*. Chicago: Open Court.

Atkins, C.M., Selcher, J.C., Petraitis, J. J., Trzaskos, J.M., Sweatt, J.D. (1998) "The MAPK cascade is required for mammalian associate learning." *Nature Neuroscience* 1: 602-609.

Azarashvili, A.A. (1999) "The Complex Perception of Stimuli during the Acquisition of Conditioned Reflexes", *Neuroscience and Behavioral Physiology* 32 (3): 247-254.

Baerends GP, Kruijt, J.P. (1973) *Stimulus Selection*. London: Academic Press.

Barco A, Bailey CH, Kandel ER. (2006) "Common molecular mechanisms in explicit and implicit memory." *Journal of Neurochemistry* 97(6): 1520-33.

Barnett, S. A. (1975) *The Rat: A Study in Behavior*. Chicago: University of Chicago Press.

Bechtel, W. (2005) *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2000) "From imaging to believing: Epistemic issue in generating biological data" in R. Creath and J. Maienschein (Eds.), *Epistemology and Biology*, 138-163. Cambridge: Cambridge University Press.

Bechtel, W. (1994) "Deciding on the data. Epistemological Problems Surrounding Instruments and Research Techniques in Cell Biology", *Philosophy of Science* (*Supplement*) v.2.

Bechtel, W. and Richardson, R. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press.

Bechtel W. & Stufflebeam, R. S. (2001) "Epistemic Issues in Procuring Evidence about the Brain: The Importance of Research Instruments and Techniques", in Bechtel, P. Mandik, Jennifer Mundale and R. S. Stufflebeam (eds.), *Philosophy and the Neurosciences: A Reader*. Oxford: Basil Blackwell.

Belelovsky K., Maroun M., Rosenblum K. (2007) "MAPK activation in the hippocampus *in vivo* is correlated with experimental setting", *Neurobiology of Learning and Memory* 88 (1):58-64.

Bennett, M.R. and P.M.S. Hacker. (2003) *Philosophical Foundations of Neuroscience.* Malden, MA: Blackwell Publishing.

Berger, T.W., Rinaldi, P.C., Weisz, D.J., and Thompson, R.F. (1983) "Single-unit analysis of different hippocampal cell types during classical conditioning of rabbit nictitating membrane response", *Journal of Neurophysiology* 50: 1197-1219.

Berger, T.W., Laham, R.I., and Thompson, R.F. (1980) "Hippocampal unit-behavior correlations during classical conditioning", *Brain Research* 193: 229-48.

Berger, T.W. and Thompson R.F. (1978) "Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictitating membrane response. I. The hippocampus", *Brain Research* 145: 323-46.

Blacker, D. and Endicott, J. (2000) "Psychometric Properties: Concepts of Reliability and Validity", in *Handbook of Psychiatric Measures*, pp. 7-14.

Bliss, T.V., Lomo, T. (1973) "Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path", *Journal of Physiology* 232:331-356.

Bliss TV, Collingridge GL, Morris RG. (2003) "Long-term potentiation: Enhancing Neuroscience for 30 Years", *Philosophical Transaction of the Royal Society London B Biological Sciences* 358(1432): 607-11.

Bogen, J. (2002a) "Epistemological Custard Pies from Functional Brain Imaging." *Philosophy-of-Science* 69(3) (*Supplement*): S59-S71.

Bogen, J. (2002b) "Experiment and Observation" in *The Blackwell Guide to the Philosophy of Science*, P. Machamer (ed.), Cambridge: Blackwell, 128-148.

Bogen, J. (2001a) "'Two As Good As a Hundred': Poorly Replicated Evidence in Some Nineteenth Century Neuroscientific Research", *Studies in the History and Philosophy of Modern Physics* 32C(3): 491-533.

Bogen, J. (2001b) "Functional Imaging Evidence: Some Epistemic Hot Spots" in *Theory and Method in the Neurosciences*, P. Machamer, R. Grush, and R. McLaughlin (eds.). Pittsburgh: University of Pittsburgh Press.

Bogen, J. and Woodward, J. (1992) "Observations, Theories and the Evolution of the Human Spirit", *Philosophy of Science* 59: 590-611.

Bogen, J. and Woodward, J. (1988) "Saving the Phenomena", *The Philosophical Review* 97:303-352.

Breland, K. and Breland, M. (1961) "The misbehavior of organisms", *American Psychologist* 16: 681-684.

Brewer, W.F. (1974) "There is no Convincing Evidence for Operant or Classical Conditioning in Adult Humans" in W.B. Weimer and D.S. Palermo (eds.), *Cognition and the symbolic processes*. Hillsdale, NJ: Erlbaum, 1-42.

Bronfenbrenner, U. (1979) *The ecology of human development: Experiments by nature and design*. Cambridge: Harvard University Press.

Brunswick, E. (1947) *Systematic and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswick, E. (1943) "Organismic Achievement and Environmental Probability", *The Psychological Review* 50: 255-272.

Calhoun, J. B. (1963) "The ecology and sociology of the Norway rat", *U.S. Public Health Service Publication* No. 1008. Washington, D.C.: U.S. Government Printing Office.

Campbell, D.T. (1957) "Factors Relevant to the Validity of Experiments in Social Settings", *Psychological Bulletin* 54: 297-312.

Campbell, D.T. and Stanley, J.C. (1963) *Experimental and Quasi-experimental Designs for Research.* Chicago: Rand-McNally.

Campbell, D.T. and D. Fiske (1959) "Convergent and Discriminant Validation by Multi-trait-Multi-method matrix", *Psychological Bulletin* 56: 81-105.

Cartwright, N. (1999) *The Dappled World.* Cambridge: Cambridge University Press.

Cartwright, N. (1983) *How The Laws of Physics Lie*. Oxford University Press.

Cook, T.D. and Campbell, D.D. (1979) *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand-McNally.

Craver, C. (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.

Cronbach L. (1960) *Essentials of Psychological Testing*. New York: Harper.

Cronbach L. and P. Meehl (1955) "Construct Validity in Psychological Tests". *Psychological Bulletin* 52: 281-302.

Danzinger, K. (1990) *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge: Cambridge University Press.

Delehanty, M. (2005) *Empiricism and the Epistemic Status of Imaging Technologies*. Doctoral Dissertation, University of Pittsburgh.

Dennett, D. (1978) "Why the Law of Effect Will Not Go Away" in *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: MIT Press.

Dennett, D. (1978) "Skinner Skinned" in B*rainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: MIT Press.

Disterhoft, J.F., Kwan, H.H., and Lo, W.D. (1977) "Nictitating membrane conditioning to tone in the immobilized albino rabbit", *Brain Research* 137: 127-43.

Disterhoft, J.F., Quinn, K.J., Weiss, C. and Shipley, M.T. (1985) "Accessory abducens nucleus and conditioned eye retraction/nictitating membrane extension in rabbit", *Journal of Neuroscience* 5: 941-950.

Dudai, Y.(2002) *Memory from A to Z Keywords, Concepts, and Beyond*. Oxford: Oxford University Press.

Dudai, Y. ([1989] 1994) *The Neurobiology of Learning and Memory*. Oxford: Oxford University Press.

Duhem, P. (1954) *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

Eichenbaum, H. and Cohen, N.J. (2001) *From conditioning to conscious recollection: memory systems of the brain*. New York: Oxford University Press.

English J.D., Sweatt, J.D. (1997) "A requirement for the mitogen-activated protein kinase cascade in hippocampal long term potentiation". *Journal of Biology Chemistry* 272:19103-19106.

English J.D., Sweatt, J.D. (1996) "Activation of p42 mitogen-activated protein kinase in hippocampal long-term potentiation". *Journal of Biology Chemistry* 271: 24329-24332.

Feest, U. (Doctoral Dissertation 2003, unpublished) *Operationism, Experimentation, and Concept Formation*. University of Pittsburgh.

Fodor, J.(1968) *Psychological Explanation*. New York: Ramdom House.

Franklin, A. (2005) *No easy answers: science and the pursuit of knowledge*. Pittsburgh: University of Pittsburgh Press.

Franklin, A. (2002) *Selectivity and discord : two problems of experiment*. Pittsburgh: University of Pittsburgh Press.

Franklin, A. (1999) *Can that be right?: essays on experiment, evidence, and science*. Boston: Kluwer.

Franklin, A. (1986) *The neglect of experiment*. New York : Cambridge University Press.

Galef, B. (1985) "Social learning in wild Norway rats", in T. D. Johnston and A. T. Pietrewicz

(eds.). *Issues in the ecology of learning*, Hillsdale, NJ: Erlbaum, 143-165.

Galison, P. (1987) *How Experiments End*. Chicago: University of Chicago Press.

Gibson, J.J. (1960) "The concept of the stimulus in psychology." *American Psychologist* 15:694-703.

Gibson, J.J. (1941) "A critical review of the concept of set in contemporary experimental psychology". *Psychological Bulletin* 38: 781-817.

Gioia M., Galbiati S., Rigamonti L., Moscheni C. and Gagliano N. (2001) "Extracellular signal-regulated kinases 1 and 2 phosphorylated neurons in the tele- and diencephalon of rat after visceral pain stimulation: an immunocytochemical study", *Neuroscience Letters* 308: 177-180.

Glymour, C. (1980) *Theory and Evidence*. Princeton: Princeton University Press.

Goldman, A. (2004) *Pathways to Knowledge: Public and Private*. Oxford: Oxford University Press.

Goldman, A. (1986) *Epistemology and Cognition*. Harvard: Harvard University Press.

Goldman, A. (1979) "What is Justified Belief?" in *Justification and Knowledge*, George Pappas, (ed.), Dordrecht: D. Reidel.

Goodenough, F.L. (1950) *Mental Testing*. New York: Rinehart.

Granado N., Ortiz O., Surez. LM, Martn E., Cea V., Sols J., Moratalla R. (2007) "D1 but not D5 Dopamine Receptors Are Critical for LTP, Spatial Learning, and LTP-Induced arc and zif268 Expression in the Hippocampus", *Cerebral Cortex*: EPUB.

Guala, F. (2003) "Experimental Localism and External Validity", *Philosophy of Science Supplement* (70): 1195-1205.

Guilford, J.P. (1946) "New Standards for Test Evaluation", *Educational Psychological Measurement*6: 427-439.

Gulliksen, H. "Intrinsic validity" *American Psychologist* 5: 511-517.

Guthrie, E. (1935) *Psychology of Learning*. New York: Peter Smith Pub Inc.

Hacking, I. (1992) "The Self-Vindication of the Laboratory Sciences" in A. Pickering (ed.), *Science as Practice and Culture*. Chicago: University of Chicago Press.

Hacking, I. (1983) *Representing and Intervening*. Cambridge: Cambridge University Press.

Hammond, E. and D. Horn ([1958] 1984) "Smoking and death rates–report on forty-four months of follow-up of 187,783 men", JAMA 251(21):2840-53.

Hárre, R. (1998) "Recovering the Experiment", *Philosophy*, Vol. 73, No. 285, pp. 353-377.

Hárre, R., Aronson, J.L. and E.C. Way. (2000) "Apparatus as Models of Nature" in *Metaphor and Analogy in the Sciences*, F. Hallyn (ed.), pp. 1-16.

Hebb D.O. (1949) *The organization of behavior; a neuropsychological theory*. New York:Wiley.

Hebb D.O. (1932) "Conditioned and Unconditioned Reflexes and Inhibition", Unpublished MA Thesis, McGill University, Montreal, Quebec, April 1932 (excerpts supplied by Richard E. Brown, Dalhousie University)

Hempel, C.G. (1954)"A Logical Appraisal of Operationism", *Scientific Monthly* 79, pp. 215-20.

Herrnstein, R.J. (1970) "On the Law of Effect", *Journal of the Experimental Analysis of Behavior*13: 243-266.

Hesse, M. (1962) 'Models and Matter' in *Quanta and Reality*, S. E. Toulmin (ed.), pp. 49-57.

Hinde, R.A. (1973) *Constraints on Learning–An Introduction to the Problems*. London: Academic Press.

Ho O., Delgado J., O'Dell T. (2004) "Phosphorylation of proteins involved in activity-dependent forms of synaptic plasticity is altered in hippocampal slices maintained in vitro", *Journal of Neurochemistry* 91(6):1344-57.

Jenkins, J.G. (1946) "Validity for What?" *Journal of Consultational Psychology* 10: 93-96.

Kandel, E.R., Spencer, W. A. (1968) "Cellular Neurophysiological Approaches in the Study of Learning", *Journal of Physiology* 48:65-134.

Kandel, E.R., Jessell, T.M. and Schwartz (2000) *Principles of Neural Science*, *Fourth Edition*. New York: McGraw-Hill.

Kelly, K. (1996) *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.

Klein, R.M. (1999) "The Hebb Legacy", *Canadian Journal of Experimental Psychology* 53: 1-3.

Knowlton, B.J., Mangles, J.A., Squire, L. R. (1996) "A Neostriatal Habit Learning System in Humans", *Science* 273:199-1402.

Krauth, J. (2000) *Experimental Design: A Handbook and Dictionary for Medical and Behavioral Research*. Amsterdam: Elsevier.

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

LaBar, K.S. and Disterhoft, J.F. (1998) "Conditioning, Awareness, and the Hippocampus", *Hippocampus* 8: 620-626

Lamprecht, R. and LeDoux, J. (2004) "Structural Plasticity and Memory", *Nature Reviews Neuroscience* 5: 45-54.

Latour, Bruno (1988) *The Pateurization of France*. Cambridge: Harvard University Press.

Ledoux, J.E., Iwata, J., Cicchetti, P., and Reis, D.J. (1988) "Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear", *Journal of Neuroscience* 8: 2517-2529.

Ledoux, J.E., Sakaguchi, A., Iwata, J. and D.J. Reis. (1986) "Interruption of projections for the

medial geniculate body to an archi-neostriatal field disrupts the classical conditioning of emotional responses to acoustic stimuli in the rat", *Neuroscience* 17: 612-627.

Ledoux, J.E., Sakaguchi, A. and D. J. Reis. (1984) "Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned by acoustic stimuli", *Journal of Neuroscience* 4(3): 683-698.

LeFrancois, G. (2000) *Theories of Human Learning: What the Old Man Said* (Fourth Edition). Belmont: Wadsworth.

Lewin, K. (1943) "Defining the Field at a Given Time", *The Psychological Review* 40: 292-310.

Liu L, Zhang S, Zhu Y, Fu Q, Zhu Y, Gong Y, Ohtsu H, Luo J, Wei E, Chen Z. (2007) "Improved learning and memory of contextual fear conditioning and hippocampal CA1 long-term potentiation in histidine decarboxylase knock-out mice", *Hippocampus*: EPUB.

Lomo, T. (2003) "The discovery of long-term potentiation". *Philosophical Transactions of the Royal Society of London B Biological Sciences* 358:617-620.

Lorente de No (1938) "Synaptic stimulation of motoneurons as a local process", *Journal of Neurophysiology* 1: 195-206.

Lorenz, K. (1965) *Evolution and Modification of Behavior*. Chicago: University of Chicago Press.

Machamer, P., Darden, L. and Craver, C. (2000) "Thinking About Mechanisms", *Philosophy of Science* 67(1): 1-25.

Machamer, P., R. Grush, and P. McLaughlin. (2001) *Theory and Method in the Neurosciences*. Pittsburgh: University of Pittsburgh Press.

Mapel, B.M. (1977) "Philosophical Criticism of Behaviorism: An Analysis" in *Behaviorism* 5:17-32.

Martin, S., Grimwood, P. and Morris, R. (2000) "Synaptic Plasticity and Memory: An Evaluation of a Hypothesis", *Annual Review of Neuroscience* 23: 649-711.

Mauk, M.D., and Thompson, R.F. (1987) "Retention of classically conditioned eyelid responses following acute decerebration", *Brain Research* 403: 89-95.

Mayo, D. (2000) "Experimental Practice and an Error Statistical Account of Evidence", *Philosophy of Science* 67(3): S193-S207.

Mayo, D. (1996) *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Mayo, D. (1991) "Novel Evidence and Severe Tests", *Philosophy of Science* 58:523-52.

Messick, S. (1989) "Validity" in *Educational Measurement 3rd Ed.*, R.L. Linn (Ed.). New York: Macmillan, 13-103.

Milner, B. Squire, L.R., Kandel, E.R. (1998) "Cognitive Neuroscience and the Study of Memory",

Neuron 20: 445-468.

Mitchell, S. (2003) *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Monfils M., Teskey G. (2004) "Skilled-learning-induced potentiation in rat sensorimotor cortex: a transient form of behavioural long-term potentiation", *Neuroscience* 125(2):329-36.

Monfort P, Felipo V. (2007) "Hippocampal long-term potentiation is reduced in mature compared to young male rats but not in female rats." *Neuroscience* 11; 146(2):504-8.

Morgan, M. and Morrison, M.(1999)"Models as Mediating Instruments", in *Models as Mediators: Perspectives on Natural and Social Science*, Morgan and Morrison (eds.), Cambridge: Cambridge University Press.

Morris, R.G. (2003) "LTP and Memory" in Long-Term Potentiation: enhancing neuroscience for 30 years, *Philosophical Transactions of the Royal Society of London, Biological Sciences*: 643-648.

Morris, R.G., Frey, U. (1997)"Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience?", *Philosophical Transactions of the Royal Society of London Biological Science* 352(1360): 1489-1503.

Nozick, R. (1981) *Philosophical Explanations*. Cambridge: Harvard University Press.

Pantev, C., B. Ross, T. Fujioka, L.J. Trainor, M. Schulte and M. Schulz. "Music and Learning-Induced Cortical Plasticity" in *The Neurosciences and Music*. New York: *Annals of the New York Academy of Sciences*, 438-450.

Pare, D. (2002) "Mechanisms of Pavlovian fear conditioning: has the engram been located?" *Trends in Neuroscience* 25:436-438.

Pavlov, I.P.P. (1927) *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford: Oxford University Press.

Pavlov, I.P.P. (1906) "The scientific investigation of the physical faculties or processes in the higher animals". *Science* 24: 613-619.

Quaranta A, Siniscalchi M, Vallortigara G.(2007) "Asymmetric tail-wagging responses by dogs to different emotive stimuli", *Current Biology*, Vol. 17(6): R199-R201.

Rescorla, R.A. (1980) *Pavlovian second-order conditioning: studies in associative learning*. New York: Halsted Press.

Rescorla, R.A.(1976) "Stimulus generalization: some predictions from a model of Pavlovian conditioning". *Journal of Experimental Psychology and Animal Behavior Process*, 88-96.

(1967) Rescorla, R.A. "Pavlovian conditioning and its proper control procedures". *Psychological Review* 74(1): 71-80.

Rescorla, R.A. and Solomon, R.L. (1967)"Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning", *Psychological Review* 74(3): 151-182.

Belelovsky K, Maroun M, Rosenblum K. (2007) "MAPK activation in the hippocampus in vivo is correlated with experimental setting", *Neurobiology of Learning and Memory* 88(1):58-64.

Roush, S. (2005) *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Clarendon Press.

Salmon. W. (1967[1966]) *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Schaffner, K.(2001) "Extrapolation from Animal Models: Social Life, Sex, and Super Models" in *Theory and Method in the Neurosciences*, P. Machamer, R. Grush, and P. MacLaughlin (eds.), 200-230. Schaffner, K. (1998) "Model Organisms and Behavioral Genetics: A Rejoinder", *Philosophy of Science* 65: 276-288.

Schaffner, K.(1994) "Interactions Among Theory, Experiment, and Technology in Molecular Biology", *Proceedings of the Biennial Meetings of the Philosophy of Science Association* 2: 192-205.

Schors T. and Matzel, L. (1997) "Long-term Potentiation: What's Learning Got to Do with It?", *Behavioral and Brain Sciences* 20: 597-655.

Schmuckler, M. (2001) "What is Ecological Validity? A Dimensional Analysis", *Infancy* v. 2, No. 4: 419-436.

Seligman, M.E.P. (1970) "On the generality of the laws of learning", *Psychological Review* 77: 406-418.

Shadish, W., Cook T. and Campbell, D. (2002) *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.

Sharp, P., McNaughton, B., and Barnes, C. (1985) "Enhancement of hippocampal field potentials in rats exposed to a novel, complex environment", *Brain Research*, Vol. 339, Issue 2: 361-365.

Shen C., Tsimberg Y., Salvadore C., Meller E. (2004) "Activation of Erk and JNK MAPK pathways by acute swim stress in rat brain regions", *BMC Neuroscience* 5(1):36.

Sherrington, C.S. (1941) *Man on His Nature*. New York: MacMillan.

Sherrington, Charles. (1941) *Man on His Nature*. New York: The Macmillan Company.

Shettleworth, S.J. (1972) "Constraints on learning" in *Advances in the study of Behavior*, v. 4, (D.S. Lehrman, R.A. Hinde and E. Shaw, eds.). New York: Academic Press.

Skinner, B.F. (1969) *Contingencies of reinforcement; a theoretical analysis*. New York: Appleton-Century-Crofts.

Skinner, B.F. (1938) *The behavior of organisms; an experimental analysis*. New York: D. Appleton Century Company Inc.

Spence, K. “The Postulates and Methods of Behaviorism”. *Psychological Review* LV: 67-78.

Squire, L.R. (2004)“Memory systems of the brain: a brief history and current perspective”, *Neurobiology of Learning and Memory* 82(3):171-7.

Squire, L.R. (1996) “Mechanisms of Memory”. *Science* 22:1612-1619.

Squire, L.R. and Kandel E.R. ([1999] 2003) *Memory: From Mind to Molecules*. New York: W.H. Freeman.

Squire, L.R. (1987) *Memory and the Brain*. Oxford: Oxford University Press.

Stevenson-Hinde, J. (1973) *Constraints on Reinforcement*. London: Academic Press.

Suppes, P. (1975) “From Behaviorism to Neobehaviorism”. *Theory and Decision* 6: 269-285.

Sweatt, J.D. (2003) *Mechanisms of Memory.* San Diego: Elsevier.

Sweatt, J.D. (1999) “Toward a molecular explanation for long-term potentiation”, *Learning and Memory* 6: 399-416.

Taylor, C. (1970) *The Explanation of Behavior*. London: Routledge and Kegan.

Thompson, R.F., Barchas, J.D., Clark, G.A., Donegan, N., Kettner, R.E., Lavond, D.G., Madden, J., Mauk, M.D., and McCormick, D.A. (1984) “Neuronal substrates of associative learning in the mammalian brain” in *Primary Neural Substrates of Learning and Behavioural Change*, D.L. Alkon and J. Farley, eds., pp. 71-99.

Thompson, R.F. (1976) “The search for the engram” *American Psychology* 31: 209-217.

Thorndike, E.L. (1911) *Animal Intelligence; Experimental Studies*. New York: The Macmillan company.

Thorndike, E.L.(1898) “Animal intelligence: An experimental study of the associative processes in animals”, *Psychological Review* 5, No. 5: 551-553.

Tinbergen, N. (1951) *The Study of Instinct*. Oxford: Claredon Press.

Waelti, P., Dickinson, A. & Schulz, W. (2001) “Dopamine responses comply with basic assumptions of formal learning theory”, *Nature* 412: 43-48.

Weber, M. (2005) *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

Weber, M. (2001) “Under the Lamppost: Commentary on Schaffner” in *Theory and Method in the Neurosciences*, P. Machamer, R. Grush, and P. McLaughlin (eds.), pp. 231-249.

Whitlock, J.R., Heynan, A.J., Shuler, M., Bear, M.F. (2006) “Learning Induces Long-Term Potentiation in the Hippocampus”, *Science* 25, v. 313: 1093-1097.

Wimsatt, W. (1981)“Robustness, Reliability, and Overdetermination” in *Scientific Inquiry and the Social Sciences*, M.B. Brewer and B.E. Collins (eds.). San Francisco: Jossey-Bass Inc.

Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Woodward, J. (2000) “Data, Phenomena and Reliability”, *Philosophy of Science* 67(3): S163-S179.

Woodward, J. (1989) ”Data and Phenomena”, *Synthese* 79: 393-472.

Yang J, Hou C, Ma N, Liu J, Zhang Y, Zhou J, Xu L, Li L. M. (2006) “Enriched environment treatment restores impaired hippocampal synaptic plasticity and cognitive deficits induced by prenatal chronic stress”, *Neurobiology of Learning and Memory* 87(2):257-63.