The Practice of Evidence in Evidence-based Practice: A Conceptual History of Evidence in Psychological Clinical Practice

by

Nathalie Lovasz

M.A. (Clinical Psychology), Simon Fraser University, 2009
B.A. (Hons.), Wilfrid Laurier University, 2007

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Psychology
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY

Summer 2013

Approval

Name: **Nathalie Lovasz**

Degree: **Doctor of Philosophy (Psychology)**

Title of Thesis: ***The Practice of Evidence in Evidence-based Practice: A Conceptual History of Evidence in Psychological Clinical Practice***

Examining Committee: **Chair:** Dr. Robert McMahon
Professor

Dr. Jack Martin
Senior Supervisor
Professor

Dr. Kathleen Slaney
Senior Supervisor
Associate Professor

Dr. Roger Frie
Supervisor
Professor

Dr. Richard Young
Internal Examiner
Professor
Department of Educational and
Counselling Psychology, and
Special Education
University of British Columbia

Dr. Alexandra Rutherford
External Examiner
Associate Professor
Department of Psychology
York University

Date Defended/Approved: August 26, 2013

Partial Copyright Licence

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

The concept of 'evidence' plays an important role in the epistemology of science, a role that has been amplified recently within psychology with the advent of the evidence-based practice (EBP) movement. However, psychologists have devoted little attention to exploring the meaning of the concept of 'evidence.' The purpose of this project was to examine the concept of 'evidence' in social contexts throughout history, in psychology, and finally in the EBP movement in clinical psychology, with the aim of elucidating: (1) definitions and contexts of employment of 'evidence' throughout the history of Western thought, (2) definitions and the role of evidence in psychology and EBP, (3) philosophical and conceptual issues related to various conceptualizations of 'evidence' and the implications of these for psychological practice and research.

Historical, archival, and qualitative methods employed towards this end included: (1) a review of historical literature and sources on 'evidence,' (2) interviews with members of the APA Task Force on EBP, (3) a review of relevant archival records of Task Force deliberations, (4) a qualitative analysis of published articles that relate to the EBP movement, and (5) a review of philosophical treatments of evidence and conceptual issues that arise in the EBP literature.

The historical review revealed that conceptualizations of 'evidence' and its role in the generation of knowledge have shifted throughout human history and across social contexts. Shifts in conceptualizations of evidence within the narrower context of science have been accompanied by changes in scientific practices and conventions. A review of the evidence-based practice literature in psychology revealed a multitude of conceptualizations of 'evidence' and its role within the EBP literature. The review of philosophical treatments of 'evidence' helped to highlight different implications of conceptualizations of 'evidence' for scientific practice. For instance, the assumptions inherent in conceptualizing evidence in various ways are at times conceptually, logically, and practically incompatible. The definitional plurality of evidence in the EBP discourse is problematic for coherent scientific and clinical practice in psychology. By drawing attention to such matters, this thesis encourages clinical psychologists to attend more carefully to EBP initiatives and the implications these carry for clinical practice.

Keywords: Evidence-based practice; evidence; history of psychology; clinical practice, theoretical psychology

To Alex, who sacrificed almost as many evenings and weekends to this project as I did, and Andrew who was there all the way through.

Acknowledgements

I owe a great debt of gratitude to many people who contributed to this project in various direct and indirect ways.

I consider myself very fortunate to have been able to work with my supervisors, Dr. Jack Martin and Dr. Kathleen Slaney. They modeled integrity in scholarship and supervision and provided consistent mentorship, support, and encouragement in a manner that far exceeded what I possibly could have wished for.

I am also indebted to Dr. Roger Frie, who generously provided his input and advice on my project, and to my examiners, Dr. Richard Young and Dr. Alexandra Rutherford, who volunteered their time to review and discuss my work, despite its length.

I am grateful to my research participants, the members of the APA Task Force on Evidence-based practice, who took time out of their very busy schedules to speak to me, and make this project a reality. Dr. Bruce Wampold deserves special mention for his invaluable advice and guidance in the early days of this project. Dr. Lynn Bufka helped me locate the documents of the Task Force proceedings, without which the history section would be much sparser.

My lab mate, Snjezana Huerta played an invaluable role in bringing the project to its completion. She stepped in with an incredibly generous offer of her time, proofread hundreds of pages of my work, and gave me encouragement and the final push to the finish line when I most needed it.

Funding from the Michael Smith Foundation for Health Research and the Social Sciences and Humanities Council of Canada gave me the freedom to devote my full attention to this project.

Throughout the years of my doctoral program, the Research Assistants of the History and Theory Lab provided a comfortable home base for my studies, supported many of my research projects, and allowed me to teach them and learn from them.

Dr. Alexander Chapman mentored my development as a clinician and shaped my career development and professional identity. My colleagues at the Mindfulness Clinic believed in me before I was able to finish this project, and encouraged me throughout.

My dear friends, Sophie Yeung, Katherine Dixon-Gordon, Kris Walters, Theone Patterson, Lisa Dreger, and Jill Satin, encouraged me, worked alongside me, balanced me, and shared many meals that were a product of my procrastination of this project. My dearest friend, Andrew Spice, was truly there for all of it, and no words could adequately describe his contributions.

My family motivated me towards this goal from the start, and shaped who I am today.

Finally, my partner, Alex, patiently postponed numerous evening and weekend plans, got on many airplanes, relocated twice, provided computer support, brightened my days of work with his company, and held me through the many meltdowns that inevitably accompanied the long days, nights, and weekends of work that this project required.

Table of Contents

List of Acronyms

APA	American Psychological Association
EBP	Evidence-based Practice
EST	Empirically-supported Treatment
HMO	Health Management Organization

Chapter 1:
An Introduction

To say that 'evidence'[1] occupies a privileged position in the epistemology of science is hardly an overstatement. In fact, many have argued that reliance on evidence is the defining feature that distinguishes science from other knowledge-seeking endeavours, pseudoscience, and the humanities. It is thus not surprising that 'evidence' also plays an important role in psychology, a discipline that in the past century has striven to establish itself as a legitimate scientific discipline in its own right, on par with natural sciences such as physics, chemistry, and biology. The importance of 'evidence' for psychology has been further amplified in recent decades, with the advent of the Evidence-Based Practice (EBP) movement. 'Evidence,' it seems, has come to be conceptualized as a central commodity or even currency in psychological interchange; conceptualized at times as the product of psychological research, 'evidence' is talked about as something to be consumed and applied by psychological practitioners, and something which also provides a measure of the value of psychological research and psychological practice. The prolific researcher, it is argued, produces evidence with wide-ranging applications. The responsible clinician likewise uses interventions that are supported by a large body of evidence. Funding agencies, HMO's, and insurance providers use 'evidence' to gauge the value of psychological research and practice.

[1] For increased clarity, throughout this work, single quotation marks will be used to distinguish discussions of the concept of 'evidence' from references to actual evidence or general phenomena designated by the concept.

Some Relevant Conceptual Definitions and Clarifications: A Glossary of Terms

For clarity of communication, some definitions of the concepts and distinctions that play a central role in the following discussion will be necessary. Many of these terms have been used frequently in various contexts with varied connotations and denotations, and thus their intended meanings in the present context must be clarified.

Psychotherapy is herein defined as any psychological intervention that is undertaken in a structured setting between two individuals, a trained psychotherapist and a patient (or patients) who seek help. Psychotherapy aims to ameliorate emotional suffering, enhance functioning, and increase understanding (including assessment and intervention) through behavioural, cognitive, and interpersonal (rather than physiological) means and is administered by someone with expertise in such interventions. Psychotherapy is provided by members of various mental health disciplines including psychology, counselling, psychiatry, psychiatric nursing, and clinical social work.[2] **Psychological practice** is herein defined as the practice of interventions and assessments with the aims of psychotherapy, as described in the previous definition. **Clinical research** is defined as research that could be relevant to clinical practice, including research on specific interventions, methods, measures, or treatments.

It is also important to distinguish between empirically validated treatments, empirically supported treatments, and evidence-based-practice, as confounding of these concepts abounds in the literature of psychology. **Empirically Supported Treatments (EST's)** are treatments that meet a set of criteria for efficacy proposed and established by Division 12 (Clinical Psychology) of the American Psychological Association (see Chambless & Ollendick, 2001). Lists of such treatments and categorization schemes for empirical support (e.g., probably efficacious, possibly efficacious) exist and judgments of

[2] Psychotherapy, as defined in the present work is not limited to any specific form of treatment or any particular theoretical orientation. Rather, it is acknowledged that treatments based on various theoretical frameworks (e.g., cognitive, behavioural, psychodynamic) are of interest in discussions of EBP, despite the fact that the EBP literature tends to be biased towards more short-term interventions (i.e., cognitive-behavioural treatments).

efficacy within this movement are based on the availability of randomized-controlled-trials. Support for the efficacy of a treatment in this context is conceptualized in a categorical manner (i.e., a treatment is either considered to be shown to be efficacious or not) rather than a dimensional manner (i.e., demonstrations of efficacy are not considered to be a matter of degree; there is little room for the efficacy of some treatments to be considered more clearly demonstrated than that of other treatments). The term **Empirically Validated Treatments (EVT's)** is the historical predecessor of the term EST's and was used in original proposals advocating the establishment of such criteria (e.g., Chambless et al., 1996). This term was discarded following criticisms of its implications of finalized determination of efficacy, a criterion that is thought to be contrary to the provisional nature of scientific evidence. **Evidence-based practice (EBP)**, based on the statement of the APA Task Force on Evidence-based practice, is "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA Presidential Task Force on Evidence-based Practice, 2006, p. 3). EBP is not based on lists of treatments or specific criteria for efficacy and assumes that evidence is multi-faceted, contextualized, and exists on a continuum of certainty. It is the most recent attempt to establish quality criteria for psychological treatments.

In the present project, a distinction between empirical and conceptual issues will be drawn based on works by Baker and Hacker (e.g., 1980). Drawing on Wittgenstein's post-*Tractatus* writings, Baker and Hacker argue that language is essentially normative and that the meaning of a concept is given by the rules of its employment according to the conventions of the language(s) and language contexts in which it has a use. Subsequently, they draw a distinction between conceptual and empirical issues. **Conceptual issues** have to do with the meaning of a concept. Conceptual issues are adjudicated by normative standards and concern the correct application of a term. In other words, a conceptual issue concerns the meaning of a term or the boundaries of its correct application. **Empirical issues** on the other hand are issues concerning the properties (etiology, quantity, ontology, etc.) of a concepts' referent and are adjudicated by empirical means. The clarity, veracity, and utility of empirical investigation are contingent on conceptual clarification. For example, the criteria for the classification of an animal such as a bird into a given class (i.e. BIRD) is a conceptual issue, while the

number of birds that live in a certain area is an empirical issue. Only after specifying what constitutes a bird can one begin to count them, or else one would not know what to include and exclude from one's count. Conceptual issues precede empirical investigation and are in-principle non-empirical.

The Unexamined Question: What is the Evidence in Evidence-based Practice?

One question appears to be central to EBP, and yet remains unaddressed both by the Task Force and the various responses to the Task Force statement on EBP. Specifically, the Task Force's general abstinence from theoretical and philosophical discussions and lack of specification of the boundaries of the concept 'evidence' leave the psychological community with a lack of conceptual clarity about the very concept that lies at the heart of EBP. Psychologists who follow the recommendations of the Task Force ought to base their practice on evidence, it is argued, and yet what exactly constitutes such evidence is far from obvious. How 'evidence' is conceptualized has practical and theoretical implications for EBP and therefore this question is central to many of the debates regarding EBP.

Prior to the publication of the Task Force statement, various debates regarding the ability of particular research designs to generate acceptable evidence were waged. Some (e.g., Edwards, Datillio, & Bromley, 2004; Messer, 2004) stressed the importance of evidence from individual case studies. Others defended the value of correlational findings as a source of evidence in instances where experimental evidence may not be available (e.g., Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005). A few also advocated for the incorporation of evidence derived from qualitative investigations (e.g., Bohart, 2007; Reynolds, 2000). A number of psychologists (e.g., Chwalisz, 2003) argued for a broad conception of 'evidence,' one that incorporated various types of evidence, including experimental evidence, qualitative evidence, and evidence from case studies. As these debates demonstrate, the question of what constitutes evidence clearly is of great practical importance to psychologists in the context of the EBP movement.

As EBP entered the psychological landscape, a few psychologists began to raise this very question. Years before the formation of the APA Task Force on EBP, in response to early discussions on EBP, Chwalisz (2003) argued that "The definition of 'evidence' is critical to a scientific discipline and, thus, critical to the activities of professional psychologists" (p. 499). She points out the extent to which adaptation of a positivistic epistemological framework and a narrow conception of 'evidence' may unduly constrain fruitful exploration of the complexity of psychological phenomena. In a response to the APA Task Force's statement on EBP, Davison (2006), then president of Division 12 (Clinical Psychology) of APA, urged psychologists to consider theoretical issues that underlie evidence-based practice, such as what constitutes evidence. Stuart and Lilienfeld (2007) described the lack of operationalization of the concept of 'evidence' as one of the major shortcomings of the APA Task Force report. They pointed out that the report lists a number of data sources available to clinicians, and implicitly differentiates between treatments that are and are not supported by evidence, but provides no specific guidance to clinicians on how to determine and recognize evidence-based treatments. It remains unclear what constitutes the best available evidence in a number of circumstances (e.g., where a popular treatment has not been studied, or where individual clients differ from those studied in important ways) and this lack of clarity makes EBP an almost insurmountable challenge for the clinician.

In the context of earlier debates regarding the role of science in clinical psychological practice, which anticipate much of the current EBP debates, Rychlak (1959) provided a relatively thorough theoretical examination of the nature of 'evidence.' More recently, some consideration of theoretical and conceptual issues that relate to matters of 'evidence' has appeared in the context of evidence-based medicine (e.g., Worall, 2010). However, for the most part, years after the publication of the Task Force's report, the psychological literature remains remarkably silent with respect to the nature of 'evidence' assumed and/or required in EBP proposals. EBP has been widely accepted and celebrated, and the extent to which a treatment is deemed to be evidence-based has come to be viewed as the ultimate measure of value of any psychological intervention. Nonetheless, substantial ambiguity in the use of the term 'evidence' remains. To date, there simply is no clear standard of application for the "evidence-based" label, as used to describe psychological practices.

Consequently, it would be prudent for the psychological community to devote some of its resources to an exploration of this question, in order to ensure that any commitment to EBP occurs in a context of requisite knowledge about the nature of such a commitment. In an examination of facts and values in psychological research, Sturdee (2001) very aptly pointed out that "scientific evidence cannot provide proof, it can only affirm our commitment to the conceptual structures and theoretical constructs provided by the paradigm within which what counts as 'evidence' has already been defined" (p. 61). If so, it is readily apparent that attempts to define the nature of 'evidence' assumed in proposals for evidence-based practice must proceed on conceptual, historical, and theoretical grounds. Such conceptual, theoretical, and historical explorations of other concepts of importance to psychology, and to science in general (e.g., Danziger, 1990, 2008; Daston & Gallison, 2007; Lovasz & Slaney, 2012; Martin, 2010), have proven to be fruitful in recent decades. Given the increasing disciplinary and socio-political importance of EBP in contexts of managed care in the United States, and similar trends towards EBP in Canada, such lines of inquiry seem especially timely (Hunsley, 2000; Tanenbaum, 2006; Wampold & Bhati, 2004).

Research Aims and Questions

In broadest terms, the aim of the present project is to conduct a conceptual history of the concept of 'evidence' in psychology in the context of the EBP paradigm of the past decade. More specifically, the present inquiry aims to explicate (1) the various definitions and contexts of employment of the concept of 'evidence' in the relevant history of the term, in general parlance, and in clinical psychological research; (2) the manner in which the concept of 'evidence' is employed in the discourse of EBP; and (3) some of the main epistemological and ontological commitments evident in EBP and the implications of these commitments for psychology research and practice.

The following open-ended questions and sub-questions guided this inquiry:

1. What is the history of the concept of 'evidence'?
    1.1 What are some of the social contexts in which the concept of 'evidence' played a role and how was it employed in these contexts?

1.2 What are some of the roles that 'evidence' has played in science across human history?

2. What is the history of 'evidence' in psychology?
   2.1 What roles has 'evidence' played in psychology during the past century?
   2.2 What are some of the social/cultural/historical/disciplinary/political factors that contributed to the emergence of the EBP paradigm?
3. What conceptions of 'evidence' are implicitly and explicitly endorsed in discussions of EBP in psychology?
   3.1 How do Task Force members, applied researchers, and practitioners in psychology use the term 'evidence,' especially in the psychological literature that pertains to EBP, and can these uses be classified; if so, based on what criteria should they be classified?
   3.2 What assumptions about the nature of 'evidence' are inherent in the ways in which researchers and practitioners in psychology use the term 'evidence' in the EBP literature?
4. What are some of the conceptual, ontological, and epistemological commitments associated with commonly-held conceptions of 'evidence' in psychology and what implications do these commitments have for research and practice vis-à-vis the EBP paradigm?
   4.1 What are some major treatments of 'evidence' in the relevant philosophical literature?
   4.2 How do major philosophical treatments of 'evidence' map onto conceptions of 'evidence' in the EBP paradigm.
   4.3 What are some of the implications of assumptions about 'evidence' in EBP for psychological research and practice?

Method

The present study employed what could be classified overall as a critical conceptual approach consistent with methods commonly employed in conceptual history. Various sources were consulted in order to trace shifts in the employment of the term 'evidence' across history and various social contexts. As is common in conceptual histories (e.g., Danziger, 1997: Hacking, 1995), conceptualizations of 'evidence' were used as means to understand values and practices related to the employment of evidence across time and social contexts. The main methods employed were (a) a review of literature on evidence in epistemology, philosophy of science, and psychology through iterative tracing and "sourcing" of the references provided in the Task Force report on EBP and relevant literature searches of academic databases and historical and philosophical works; (b) a review of documents from the proceedings of the APA Task Force, in order to establish a contextual historical framework for the present investigation; (c) in-person/phone/email interviews with members of the APA Task Force

on EBP, in order to explore some of the issues pertaining to definitions of 'evidence' that the Task Force dealt with during their proceedings, the decisions reached with respect to these issues, and how these decisions came about; (d) a qualitative (modified grounded theory) conceptual and philosophical analysis of the research literature on EBP, in order to examine common uses of the concept of 'evidence' by psychological practitioners and researchers, and to consider such uses against conceptions of 'evidence' available in relevant literature in epistemology and philosophy of science.

Question 1:
The History of 'Evidence'

I addressed the first major research question using an examination of relevant historical literature in order to provide a brief overview of the history of the concept of 'evidence' in Western societies. Practices related to evidence in language, law, and science were chosen as the main foci of this investigation. I examined a variety of etymological dictionaries (e.g., The Oxford English Dictionary) to provide information concerning the various social contexts in which the notion of 'evidence' emerged, the role that evidence played in these contexts, and how these practices and contexts transformed the concept of 'evidence' over time. Based on this initial investigation, it appeared that practices related to evidence in legal and scientific contexts were particularly relevant. I therefore consulted various sources on the history of law and the history and philosophy of science in order to trace the transformation of 'evidence' in legal contexts and in scientific practices.

Question 2:
'Evidence' in Psychology

A variety of works in the history of psychology were consulted as sources on the role that 'evidence' has played in psychology and psychotherapy research across time. To more closely examine the history of the evidence-based practice movement, I reviewed documents from the proceedings of the APA Task Force on EBP. I also conducted interviews with members of the APA Task Force on EBP in order to supplement available information about the proceedings.

Archival method: Proceedings of the Task Force. In light of the relatively sparse extant research on the history of EBP and the general lack of information about particular decisions and discussions with respect to the nature of 'evidence' that occurred during the deliberations of the APA Task Force on EBP, additional archival research was necessary. I obtained available working documents, reports, and minutes available from the proceedings directly from the APA.

Qualitative method: Interviews. I further contextualized and interpreted information gathered during the archival review with information gained during interviews with members of the APA Task Force. I conducted interviews with five of the members of the Task Force and analyzed interview responses using an open-ended, interpretive analysis. I incorporated perspectives and information obtained into my history of 'evidence' in EBP. Participants were members of the APA Presidential Task Force on EBP (for a list of members see Appendix A). I used internet search engines to find contact information for all members and contacted Task Force members by email to invite them to participate in this research project. I conducted open-ended interviews with five Task Force Members who agreed to participate between February and August of 2011. Participants were given a choice between phone interviews and in-person interviews at the APA annual convention in Washington, DC in August 2011. Four participants chose to complete phone interviews and one participant agreed to an in-person interview at the convention. Participants were emailed a copy of the interview guide. The interview guide consisted of a series of open-ended questions to elicit participants' conceptions of 'evidence' and their experiences in considering the nature of evidence during the deliberations of the APA Task Force (see Appendix B). Interview guides were emailed at least one week prior to the interview to allow participants to review and consider interview questions. Participants were given a chance to alter or remove questions from the interview guide as they saw fit; no participants chose to exercise this right. Interviews were semi-structured: I asked similar questions of all participants but altered the sequence of interview questions and added probing questions to enhance the flow of conversation. Interviews were audio-recorded. After interviewing participants, interviews were transcribed verbatim for content (non-verbal responses were included only in so far as they enhanced the understanding of content

and pauses were not timed or transcribed as they were irrelevant for the purposes of this project).

Transcripts were analyzed using an interpretive approach loosely based on the methods of Grounded Theory (Glaser & Strauss, 1967). All transcripts were read once to provide an overview of their content in their entirety. Parts of transcripts that were relevant to the research questions were then identified and categorized in a first round of open coding; descriptive labels were assigned to sections of interviews based on their content. Once all relevant parts of the transcripts had been coded, broad themes in Task Force members' interview responses were identified and labeled sections were sorted and organized based on these themes. Constant comparisons of coded sections of transcripts were made in order to identify differences and similarities between participants' views and experiences. Such differences and similarities became the focus of the written overview of findings from interview transcripts provided in Chapter 5.

Question 3:
Conceptions of 'Evidence' in EBP in Psychotherapy Research
– Archival Content Analysis

A qualitative content analysis of articles on evidence-based practices in clinical psychology journals was conducted to examine conceptions of 'evidence' espoused in the relevant literature. A PsycInfo search was conducted, using the title search term "evidence-based."[3] The results were reviewed systematically and filtered to include only those results that pertained to EBP in clinical psychology. Results pertaining to EBP in other areas of psychology (e.g., educational psychology, industrial/organizational psychology) were excluded from the analysis, since the focus of the present work was clinical practice rather than applied psychology more broadly. The academic journal articles sampled were analyzed and coded using a qualitative data analysis software program (MaxQDA). A more detailed lexical search of all articles sampled was conducted using the search term "evidence." All references to 'evidence' in these articles and the particular paragraphs within which the reference word occurred were

[3] A search using a non-hyphenated spelling of "evidence based" was also conducted but yielded identical results given how PsycInfo handles keywords in its search algorithms.

coded. References to 'evidence' were examined systematically and further categorized according to general themes. Relevant examples that illustrated distinctions between various conceptualizations of 'evidence' in the literature were selected and used in the write-up of the results of this work.

Question 4:

Philosophical Issues in Conceptualizations of EBP – Philosophical Investigations

Philosophical conceptions of evidence were summarized and an analysis and interpretation of their implications for the EBP discourse were conducted.

A Preview

To call my project a conceptual history may misleadingly suggest a primarily diachronic focus[4]. Instead, my investigation of the concept of 'evidence' proceeded from a diachronic approach to a synchronic approach. In fact, this distinction defines the two major sections of my work as reported herein. In the first major section, I attempt to provide a diachronic understanding of the concept of 'evidence,' with a focus on its first appearance and historical evolution in language (Chapter 2) and various social contexts, including the practices of law (Chapter 2), science (Chapter 3), and finally clinical practice (Chapter 4). I have chosen this structure in an attempt to proceed from the most general social contexts of 'evidence,' to the ones that are more specifically the focus of the present project. In this section, I aim to elucidate current understandings of the concept of 'evidence' by tracing the ways in which such conceptions have evolved historically.

The aim of the second part of my work is to provide a more synchronic conceptualization of 'evidence' in the context of practices in clinical psychology. Here, I provide an overview of current conceptualizations of 'evidence' in the context of the EBP

[4] I apply here somewhat loosely a conceptual distinction borrowed from linguistics, between the synchronic approach, which focuses on a phenomenon at a fixed point in time, and the diachronic approach that focuses on the evolution of a phenomenon over time.

movement in clinical psychology, and consider current perspectives on 'evidence' in philosophy to draw implications relevant to EBP. An overview of the history of EBP (Chapter 5) will be followed by a presentation of results from my empirical qualitative review of the psychology literature on EBP and interviews with Task Force Members, which will provide a sketch of some of the conceptions of 'evidence' that are currently found in the EBP literature (Chapter 6). A few philosophical accounts of 'evidence' will also be summarized and their implications for the EBP movement discussed (Chapter 7). My aim for this section is to describe what the current practices are with respect to 'evidence' in clinical psychology. The first section thus provides an overview of the historical route towards current practices related to evidence in psychology and clinical psychology, while the second section attempts to provide a sketch of what these present practices are.

It is a daunting task to attempt a definition of 'evidence' that does not assume a particular conception of 'evidence,' and thus, one that I do not seek to accomplish. Instead, I provide some observations about commonalities in accounts of 'evidence' that may serve as a conceptual organizational scheme. At the risk of pre-empting myself, I reveal that the following sections of this conceptual history highlight great diversity in conceptualizations of 'evidence' across time and social contexts. This is also apparent from the definitions of 'evidence' discussed below. One feature, that seems to be common to all or at least the majority of accounts of 'evidence,' irrespective of their context and philosophical assumptions, is that 'evidence' could be classified as what I will call a "relational concept." What I mean is that there is no meaningful way to define 'evidence' without referring to its relationship with other concepts. More concretely, evidence is always evidence for something (see Kim, 1998) and thus is defined by its relationship to that for which it is evidence. For example, Hacking (1975) in his work on probability, highlights that "many philosophers claim that probability is a relation between a hypothesis and the evidence for it" (p. 31) but also points out that evidence has not always been conceived in relation to a hypothesis and the relation of evidence to its objects has not always been defined probabilistically. Thus, accounts of 'evidence' that will be discussed in the following sections vary along three main dimensions: (a) the nature of evidence: the limits of what can and cannot constitute 'evidence'; (b) the object of evidence: that which evidence is taken to be evidence for (e.g., beliefs, propositions,

theories, hypotheses, divine revelations, etc.); and (c) the nature of the relationship between evidence and its objects (e.g., causal, probabilistic, logical/deductive, etc.). All accounts of 'evidence' I have come across posit objects and a relationship between evidence and those objects. It is difficult to conceive of an account of 'evidence' that does not contain these components. This framework will be used as a way to organize discussions of various conceptualizations of 'evidence' in the present project.

Chapter 2:
Evidentiary Practices

Wittgenstein (1953), in his *Philosophical Investigations* writes:

> Man kann für eine *große* Klasse von Fällen der Benützung des Wortes »Bedeutung« - wenn auch nicht für *alle* Fälle seiner Benützung - dieses Wort so erklären: Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache [For a great many cases of the use of the word 'meaning', although not for all cases of its use, the word can be explained thus: the meaning of a word is its use in language.] (p.40)

Consistent with this Wittgensteinian conception of language, which informs the current conceptual investigation, a study of the meaning of the concept of 'evidence' by necessity must entail an analysis of its uses in language and its employment in the various language contexts in which it plays a central role. The first four chapters of this work are devoted to just such an approach to understanding the meaning of the concept of 'evidence.' In the present chapter, I provide an overview of definitions and uses of the word 'evidence' in the English language and then in the context of legal proceedings and theory. Chapters 3 and 4 are focused on a more detailed look at the use of the concept of 'evidence' in science and psychology/psychotherapy practices, respectively.

The Concept of 'Evidence' in Language

At a most general level, a clarification of the concept of 'evidence' and its meaning can proceed first at the level of common usage, by examining how the uses of the concept of 'evidence' in ordinary language have emerged and evolved over time.

The Emergence of 'Evidence' in Language

An examination of the origin and uses of the concept of 'evidence' can elucidate the diverse social contexts in which this concept has emerged and similarities and differences across various conceptualizations in these contexts. The Oxford English Dictionary (1989) traces the word 'evidence' to its Latin root, where the word 'evidentem' is derived from the prefix 'e' (out, a variant of 'ex') and the verb 'videre' (to see). It is stressed that this form of the Latin word uses the active form of 'videre' in contrast to the passive 'evideri' (from videri – to appear). Cresswell (2011) in the Oxford Dictionary of Word Origins defines its meaning as "obvious to the eye or mind" (n.p.). In later discussions of philosophical treatments of 'evidence' in Chapter 7 of this work, it becomes clear that even the early roots of the concept emphasize its relationship to sensory data (i.e., evidence is not conceptualized as a product of reason) and the active role such data play in the concept's meaning. Both these aspects of the roots of the concept of 'evidence' are key feature of later definitions and conceptualizations.

The definitions of 'evidence' offered in the Oxford English Dictionary (1989) also highlight the varied conceptions of 'evidence' in different social contexts over time. In the first set of definitions, 'evidence' is defined as "the quality and condition of being evident, clearness, evidentness," (Section I, 1a) "actually present; prominent, conspicuous" (Section I, 1b), and a "manifestation; display." (Section I, 2) Thus, 'evidence' is conceptualized as something that exists in the world and is visible, directly observable or clear, and indubitable.

In contrast, 'evidence' is also defined as "that which manifests or makes evident" (Section II), including "an appearance from which inferences may be drawn; an indication, mark, sign, token, trace" (Section II, 3a). In other words, according to these definitions, 'evidence' is a sign or indicator of something that may not be directly visible or accessible.

Further, 'evidence' is defined as "ground for belief, testimony or facts tending to prove or disprove any conclusion" (Section II, 5a) or "something serving as proof" (Section II, 5b). Here 'evidence' is defined as that which justifies a belief or a basis for a

belief. In this definition, its epistemological relation to knowledge claims, rather than its empirical accessibility, defines 'evidence.'

The final set of definitions of 'evidence' focuses on its role in legal contexts and includes two broad categories. First, it is defined as "information, whether in the form of personal testimony, the language of documents, or the production of material objects, that is given in a legal investigation to establish the fact or point in question" (Section III, 6a); "The testimony which in a particular case has been received by the court or entered on its record. Similarly to be or produce in evidence" (Section III, 6b); and "statements or proofs admissible as testimony in a court of law" or "one who furnishes testimony or proof" (Section III, 6c). These definitions primarily emphasize the active process whereby evidence is produced or given in a legal context in the form of testimony. In contrast, a second definition of 'evidence' in legal contexts is given as "a document by means of which a fact is established" (Section III, 8). Thus, 'evidence' is simultaneously defined as that which is established through an active process of testimony and an objective material object that establishes fact.

Based on these definitions as a whole, 'evidence' has been defined as tangible or observable sensory data, signs, or indicators that form the basis of inference, the basis or justification of belief, and testimony or documents supporting certain facts in legal contexts. Notably, these definitions mirror many of the theoretical and philosophical accounts of 'evidence' that have been advanced, as becomes clear in later sections of this work.

Although these definitions may at first appear to be overlapping and complimentary, the differences among them become apparent when consideration is given to the various examples provided of the use of the term for each of the definitions. For instance, an example of usage based on the first set of definitions, where 'evidence' is essentially defined as observable sensory data, includes "so evident that we require no grounds at all for believing them save the ground of their own very evidence" (Mivart, 1885 as cited in "evidence", Oxford English Dictionary, 1989, Section I, 1a). In other words, the evidence is immediately and directly accessible, or in philosophical terms "the given." In contrast, 'evidence' is treated as a *sign or marker* of something else in phrases such as "an opportunity was offered them of giving evidence of their loyalty"

(Froude, 1858 as cited in "evidence," Oxford English Dictionary, 1989, Section II, 3a), or as that which justifies a claim: "evidence is not that which the mind does or must yield to, but that which it ought to yield to" (Mill, 1846 as cited in "evidence," Oxford English Dictionary, 1989, Section II, 5a). Here, evidence forms the basis of inference or justifies beliefs or inferences. In legal contexts, evidence can be material as in the case of a document -- "the document is not in evidence" (N.E.D, 1984, Section III, 6b); a person who testifies -- "two infamous and perjured evidence made oath to the prisoners having expressed themselves interested in the great confederacy of the Catholics" (Scott, 1823 as cited in "evidence," Oxford English Dictionary, 1989, Section III, 7a); or the testimony provided by a person -- "what a witness states on hearsay is not evidence" (N.E.D., 1894 as cited in "evidence," Oxford English Dictionary, 1989, Section III, 6c). It is notable that instances of usage consistent with each of the various conceptualizations span the 14th to the 19th centuries. Thus, these conceptualizations co-existed historically, as they continue to co-exist at present (see later section on treatments of 'evidence').

The Concept of 'Evidence' in the Law

One of the most obvious social contexts in which 'evidence' plays a central role is in law. Most people have participated in the production of evidence for legal purposes (as, for example, in the creation of a legal document through the signing of a contract) and evidence appears to play such a central and essential role in contemporary legal procedures that it is virtually impossible to imagine a legal system that does not rely on the presentation, examination, and consideration of evidence. This very brief survey of legal practices across history will show that legal systems have relied on some form of evidence at all times. The formal study of 'evidence' and rules of evidence in jurisprudence however, has a surprisingly short history, dating back only to about the 19th century (e.g., Ho, 2004; Twining, 2006). Both the theory and history of the law of evidence are extensive and I will eschew any attempt to provide an adequate treatment

or exhaustive overview of these topics.[5] Instead, I will focus my discussion much more narrowly, using the framework laid out above, and provide brief descriptions of kinds of legitimate evidence, the purposes of evidence (the objects of evidence), and the manner in which various kinds of evidence have been weighted and combined in legal proceedings from Antiquity to the Modern Age.

At times, it has been argued by legal historians that since Antiquity, society has moved from irrational methods of arbitration and fact finding to a rational system of law based on facts established through evidence (e.g., Thayer, 1898 cited in Twining, 2006). Similarly, with reference to evidence, many legal historians attribute the short history of evidence law to the fact that evidence did not play a major role in the irrational systems of arbitration of the courts during Antiquity and the Middle Ages (e.g., Ho, 2004). Instead of taking such a view of legal history as a history of progress however, I seek a contextualized understanding of the history of practices related to evidence in legal contexts across time. I argue that evidence, defined broadly, as that which has been used as the basis of legal decision-making, has in fact played an important role in all legal settings, but that the nature of evidence and its role in the legal process has depended on the purposes of legal proceedings. The purposes of legal proceedings have changed substantially throughout Western history. An understanding of the purpose of legal proceedings is thus essential to any understanding of what constituted evidence in these proceedings and the manner in which evidence was considered and evaluated in them. As the following sections illustrate, however, a more contextualized understanding of legal proceedings and their purposes reveals how procedures that might seem foreign to the contemporary reader could well serve the purpose of evidence in their time, and how the process of decision making based on such evidence could indeed be highly rational from a pragmatic perspective.

[5] Many good sources on these topics exist. A particularly useful treatment of the history of jurisprudence and legal process can be found in the multivolume A Treatise of Legal Philosophy and General Jurisprudence (Eds. Pattaro, Rottleuthner, Shiner, Peczenik & Sartor, 2005-2012). A thorough discussion of evidence law is to be found in Twining's (2006) Rethinking Evidence: Exploratory Essays.

Evidence in the Courtrooms of Antiquity

An understanding of the legal proceedings of Antiquity requires that one take a step back from the assumptions inherent in contemporary legal practices. Most specifically, if one assumes that, as Langbein (1996) argues, "the main work of a legal system is deciding matters of past fact" (p. 1168), the procedures and evidence used in the courtrooms of Antiquity and the Middle Ages seem wholly irrational. Some scholars have indeed arrived at such an evaluation as a result of their explorations of these legal systems. A more charitable premise, however, may be that differences in evidentiary practices of current times and Antiquity suggest a very different objective of the ancient legal process, in light of which particular legal practices become much easier to comprehend. In line with such an interpretation, it has often been suggested that the objective of ancient legal systems was the settling of disputes (and thus the maintenance of peace in society) rather than the establishment of past facts (e.g., Gagarin & Woodruff, 1995). This interpretation, in turn, raises the question of what, if any, role evidence can play in a legal process, which does not concern itself with the establishment of fact. The following exploration of 'evidence,' its object, and the relationship between the two in Antiquity will provide some answers to this question.[6]

Evidence. Evidence and the practices used in its establishment in Antiquity were closely tied to the practices of rhetoric (Ferguson, 1960; Gagarin & Woodruff, 1995). This statement has implications for all aspects of evidentiary practice, but in terms of the nature of evidence, it means that evidence in Antiquity was primarily taken to be of an oral or spoken, rather than material nature. A number of notable changes in practices surrounding evidence occurred between the various time periods of Antiquity: evidence in ancient Greek society was exclusively oral; however documentary and material evidence began to play a minor role in the legal proceedings of Roman society and began to be favored over oral evidence by Justinian times (Ferguson, 1960). Nonetheless, testimony was the primary form of evidence throughout Antiquity. Not surprisingly then, rules with respect to the admissibility of evidence focused primarily on the types of witnesses who were excluded from testifying. Women, minors, slaves,

[6] Of course, the practices surrounding 'evidence' differed greatly over time in Antiquity and only a very general treatment of this time period is offered here.

those who had come into disrepute in society, and, in the case of Greek society more than Roman society, parties with an interest in the trial outcome (e.g., relatives, the accused, and defendants themselves) were excluded from testimony (Campbell Black, 1892). Many of the people who were most privy to the facts of a trial were excluded from testifying in the trial. These exclusions did not present any major problem in the trials of Antiquity since the goal of testimony differed greatly from that in the contemporary courtroom; consistent with the objective of the legal process, the purpose of testimony was not the delineation of fact but rather often involved pleas for a fair outcome put forth by various parties. Testimony in one's support was often a form of character reference (Ferguson, 1960), where the assumption was that the ability to rally a number of credible witnesses in one's case provided proof of the legitimacy of one's case. It would thus be a mistake to confuse the testimony of Antiquity with that of the modern trial.

The relative lack of importance placed upon the delineation of facts can be most clearly illustrated by describing the oath, a common oral form of evidence in ancient legal processes. In both Greek and Roman judicial procedure, the swearing of an oath to tell the truth by witnesses and persecution of perjury were used to ensure the veracity of testimony in a way that is consistent with current legal practice (e.g., Ferguson, 1960; Gagarin, 1986). Methods of torture were also used on occasion as a means to elicit true testimony (Ferguson, 1960). In addition, decisory oaths were used as a source of evidence that does not exist in a similar form at present. In the context of a trial, participants could be asked to swear an oath to the truthfulness of their claims. A refusal of the oath was taken to be an automatic admission of guilt, resulting in a loss of one's case (Ferguson, 1960; Gagarin, 1986). Decisory oaths were thought to be enforced through divine retribution. A person who swears a false oath it was thought, would invite divine retribution from the Gods, and the fear of such retribution was to serve as a deterrent against false oaths. Decisory oaths were taken to be a form of ascertaining or establishing the truth that did not rely on the delineation of past facts. Nonetheless, a decisory oath was quite literally taken to be a way to establish the facts with certainty (Ferguson, 1960). Divine retribution was thought to be the mechanism whereby any false oath would be identifiable and fear of retribution also may well have served as a powerful deterrent against swearing a false oath. The decisory oath derived much of its

value in establishing facts from the social consensus and beliefs that existed about its power. Therefore, it is important to draw a distinction between the delineation of facts that occurs in laying out and supporting the details of a past occurrence with the end goal of determining what occurred in some narrative factual way, and the establishment of fact through other procedures, such as torture and decisory oaths. The former did not play a major role in the courtrooms of Antiquity (though some evidence exists that it was preferred during later times of the Roman empire, see Ferguson, 1960) but the latter were commonly used. Facts were not irrelevant in the ancient trial. Rather, various procedures other than the retracing of the details of past events were thought to be legitimate means of establishing facts.

The objects of evidence. To understand evidentiary practices in the legal contexts of Antiquity, it helps to look at what evidence was taken to be evidence for, hereon referred to as the object of evidence. Since, as stated above, the purpose of a trial in Antiquity was not necessarily to retrace past facts, evidence likewise was not necessarily taken to be supportive or refutative of certain accounts of the past. Rather, in line with the aims of settling disputes and re-establishing peace and order, evidence was presented with the aim of establishing a just outcome and convincing others of the justice of one's case. In simple terms then, evidence in Antiquity, especially in Greek society, sought to establish the justice of a certain outcome (Gargarin & Woodruff, 2007). Evidence was presented in order to establish to the judge, jury, and society at large that one's claim was just or that a proposed settlement would be just and fair. A variety of mechanisms were built into Greek legal procedures to provide what was at the time perceived to be a public and fair hearing with significant amounts of community participation and mediation. Gargarin and Woodruff (2007) attribute the importance of justice and just procedure to the lack of formal and established mechanisms of enforcement in Greek society. Social sanctions provided the only means to enforce legal judgments. As a result, a judgment required the support of the community if it was to be upheld through social sanctions. In some sense then, evidence was intended to serve as a foundation for the judgment of the community. Evidence had to be presented in order to convince the community of the fairness of a verdict. In light of this very different role of evidence in Antiquity, the kinds of things that were considered to be relevant evidence become comprehensible. Oaths regarding one's character, one's

behavior during a trial, along with moral arguments, were all considered very relevant to deciding a just outcome of a trial.

Somewhat similarly, in Roman society, evidence in a trial was to convince the court of one's case (Ferguson, 1960). The presentation of evidence was viewed as a rhetorical process aimed at persuading the court rather than retracing and establishing facts. Thus, evidence came in the form of arguments presented in a convincing manner and members of the Roman court were expected to make a decision reflecting their inner conviction (a judgment in the literal sense of the word) rather than a decision as to the facts of a case.

The relationship between evidence and its object. The legal procedures of Antiquity did not rely on an inductive approach to the consideration of evidence. Greek procedures relied on the principle of "straight justice," which included the hearing of pleas from both sides. However, this was done in the interest of arriving at justice rather than hearing all the facts in order to establish events of the past (Gagarin & Woodruff, 1995). The task of the trial was not to arrive at an account that incorporates all evidence and pleas presented, but rather to arrive at a decision based on subjective evaluation of pleas along with considerations of social order and benefits. A trial was to help ensure continued moral order in society by acting as a forum for deliberation and expression, not by issuing punitive, deterrent verdicts. A just society was a moral and harmonious society in Greek thought, one that follows just and fair processes. Similarly, in ancient Rome, moral considerations along with subjective deliberation by the court played major roles in legal procedure. Compared to much contemporary legal procedure, which limits admissibility of evidence but asks the court to consider all available evidence, Roman courts admitted all evidence but allowed the court considerable discretion in their subjective decision of what evidence was convincing. The relationship between evidence and its object in ancient jurisprudence was thus much more subjective and judgment-based than inductive or cumulative.

Legal Evidence in the Middle Ages

The legal proceedings of the Middle Ages were similar to those of Antiquity in the sense that the delineation of past facts was not the primary goal of a trial. There was

also some overlap in the types of evidence that were presented in the courtrooms of Antiquity and the Middle Ages. Nonetheless, the objective of trials in the Middle Ages differed from that of Antiquity in that the former was strongly influenced by the ascendance of Christianity as a major social force. The legal proceedings of the Middle Ages were viewed as instruments of the divine and the purpose of the trial was to obtain divine justice from God (Ho, 2004). With this purpose in mind, the evidentiary practices of the Middle Ages and their relationship to the object of evidence, divine justice, become easier to comprehend.

Evidence. There are three main modes of proof that are unique to the Middle Ages: ordeal, the wager by law/compurgation, and battle (Ho, 2004). Specific practices pertaining to these modes of proof varied across time and region (Ho, 2004). Nonetheless, a general description of each of these modes or practices of evidence can be provided. The ordeal is likely the most well-known of the trial procedures of the Middle Ages, made famous by the witch trials of the time. In a trial by ordeal, a defendant underwent a physical challenge such as being submerged in cold water or exposed to extreme heat through boiling water or being forced to carry hot metal. The accused's ability to withstand or recover from the challenge provided an indication of guilt or innocence. For example, in the water method, guilt was demonstrated if the suspect floated, while in the fire version, healing of one's hands after a few days was taken to be a sign of innocence. Many other versions of ordeal existed.

The wager by law mirrored the oath of Antiquity. A defendant had to summon a particular number of compurgators who would support their swearing of an oath to their innocence. It is important to note that compurgators were not required to be privy to any knowledge of the events surrounding an alleged offense (e.g., as in the role of eyewitnesses) and swore to the truth of the oath rather than the facts of the case (Carpenter, 1958). In later years, rules were established according to which compurgators had to be chosen from a pre-selected group of candidates determined by the sheriff, with the assumption that trustworthy individuals would be able to get almost any member of the community to act as a compurgator while known scoundrels would find this task more difficult (Carpenter, 1958).

Finally, in trial by battle, combat between the parties of a trial was used to decide the outcome of the trial such that the case was won by the party that won the battle. This presented some challenges to older and less physically able members of society. In later decades of the Middle Ages, parties were able to choose representatives in battle who would fight the combat for them, a practice especially common among those of higher social standing.

The object of evidence. Ho (2003) critically analyzes a variety of theories that have been put forth to explain how various types of evidence during the Middle Ages may have provided evidence for the facts of a case. For example, in ordeal by fire anxiety and guilt may have slowed down the healing of defendants who were guilty. Such theories, Ho argues, miss the point that establishment of fact was not the objective of trials of the Middle Ages. Rather, Ho contends that truth and justice were viewed as being intricately linked and were taken to be epistemically accessible to God. The purpose of these methods of evidence was therefore to gain direct access to divine judgment so as to determine a just outcome. The factual component of a judgment and its moral and just components could not be separated in the epistemology of the time, which conceived of God as "the supreme knower" who could grant humans limited access to knowledge through divine acts. God's will was necessarily taken to be just (based on God's omniscience) independently of whether it constituted a direct establishment of fact relevant to a particular case. Presumably, a defendant could pass the ordeal despite having committed the offense he was accused of, if God deemed a judgment of innocence to be a just outcome based on the defendant's character as a whole.

The relationship between evidence and its object. Ho (2003) points out that the evidentiary practices of the Middle Ages, despite their ritualistic nature, still provided some avenues for human involvement. Interpretation of the outcomes of ordeal or battle left some room for judgment. In addition, judgment of previous fact based on testimony was used to determine whether to use particular methods of adjudication. Nevertheless, as has been previously stated, overall the objective of these methods was to gain access to divine judgment. Divine intervention was thought to ensure the validity of outcomes of relevant evidentiary procedures: God, it was thought, lent strength in battle to the party with the strongest claim and aided the healing of the innocent. In wager by

law, fear of divine retribution was to deter compurgators from false oath. The legitimacy of these procedures was entirely contingent on presumed divine involvement. The procedures of the trial were thought to be a direct route of access to, or mode of receiving communication from God.

The relationship between evidence and the divine is perhaps best explained by Ho (2003) as follows:

> The outcome of a trial by ordeal, compurgation or judicial battle was not a proof outcome, if "by proof' we mean the proof of facts. When these methods of adjudication are described as "modes of proof," the term "proof' is used in the sense of "vindication." We understand the word in this way when, for example, a victim of sexual harassment asserts that the conviction of her tormentor "proves" that she was telling the truth all along, or when a person claims that an acquittal "proves" his innocence. The result of the medieval modes of proof is, rather, the adjudication outcome: it marks the termination of the dispute and is not the finding of fact to which rules of law have to be applied to reach the verdict. The justice sought through use of the medieval modes of proof was not grounded in substantive norms operating on the facts of the case; it was based, rather, on submission to and faith in a spiritual power. (p. 259)

Such a mode of adjudication was particularly appropriate in the social structures of the time, as relatively small and tightly-knit communities required final and definite outcomes that re-established peace through justice and vindication (Ho, 2004). The dismissal of guilt on the basis of insufficient grounds for conviction, which is common in contemporary society, would be ill-suited to a resumption of peaceful co-existence in Medieval society. Legal practices in the Middle Ages therefore involved retribution and ritual intended to re-establish justice much more than a forensic analysis of the past.

Practices of Evidence in Renaissance Law

The Renaissance was a time of great diversity in legal practices across Europe. While much of continental Europe saw a continuation of the legal practices of the Middle Ages, most famously within the context of the witch trials, the English legal system began a transition towards a system of jury trials (Carpenter, 1958). No new types of evidence in legal proceedings emerged during this time period. Nonetheless, evidence became a focus of attention in legal scholarship. With the establishment of universities,

first in Italy and gradually throughout the rest of Europe, law emerged as a major discipline of scholarly study. Attention came to be focused on the relationship between evidence and its object rather than the nature of evidence itself.

Evidence. As already stated, no new forms of evidence emerged during the Renaissance. Ordeals were still practiced widely across continental Europe, although participation of the clergy in such practices was outlawed by the church in the Lateran Council of 1215 (Carpenter, 1958) and replaced by the wager by law. At the same time, torture was commonly used as a way to elicit testimony or force compliance and in fact became more widespread in British feudal society with the emergence of the jury trial system. Those accused of a crime could enter a plea and choose to stand trial with a jury. If they refused a jury trial or refused to enter a plea however, torture could be used in an attempt to change their decision (Carpenter, 1958). Finally, testimony re-emerged as a major form of evidence, especially in the English legal system, but the role and admissibility of testimony began to receive much greater focus.

The object of evidence: The judgment of the jury. The use of a jury was introduced to England from Normandy during the conquest and originally used as a means to divide land. Prominent and respected men of the community would be asked to decide land claims between the church and state. The use of juries was extended to legal trials in the 13th century (Carpenter, 1958). Jurors were originally considered to be witnesses to the testimony presented in court. They were chosen from knowledgeable and respected members of the community and oftentimes even had an interest in the trial. They were encouraged to use their knowledge and judgment of the community and case to decide on pertinent facts (Carpenter, 1958). The emergence of the jury trial brought with it, for the first time, a division between the facts of a case and the laws of a case. The role of the jury was to decide the facts of a case based on the testimony presented to them. The judge was then to make a decision with respect to how the law applies to the facts of the case. The purpose of evidence in the Renaissance thus became to persuade the jury.

The relationship between evidence and its object. For the first time then, during the Renaissance the manner in which evidence was to be evaluated became a focus of scholarly attention (Maclean, 2000). In the early Renaissance, jurors could be

convicted of perjury if they entered what was taken to be a false verdict in a case. Their role in evaluating testimony and evidence was taken extremely seriously. Consequently, the credibility of witnesses became a major focus of the law. The manner in which evidence was to be evaluated when various witnesses contradicted each other became a central question in jurisprudence. Various formulas and rules were developed. Credibility was often equated with social standing. In an example cited by Maclean (2000), fifteen rules were put forth in a tract by Farinacci published in Germany in the 17th century, which included preference for a male witness over a female witness, a clergy witness over a lay witness, the testimony of a person with wealth and status over that of a lower status person, that of a person with good eyesight over the testimony of a person with poor eyesight, and so forth. Rules with respect to consistency, plausibility, and completeness of a testimonial account were also put forth. Although these rules may appear arbitrary, they were grounded in relevant social contextual factors. It was thought that the testimony of a person of lower status would be more easily influenced or purchased, thus rendering such individuals potentially less credible.

Nonetheless, it is also easy to see how in the "gentlemen culture" of English society at this time, such rules could be problematic, especially in the context of a trial where jury members were not disinterested or removed from the case but actively participating in it as involved members of the community. More elaborate schemes for the evaluation of evidence, including rules about the relative weight of pieces of evidence (evidence which could be a half-proof for example, as in the case of a less credible witnesses) were also devised (Carpenter, 1958). Towards the end of the Renaissance, attempts were made to provide procedurally fair means of adjudication. Juries had shifted from interested parties to disinterested, impartial jurors. Jurors were no longer legally responsible and subject to prosecution for perjury based on their judgment. Jurors also were increasingly encouraged to consider only testimony presented in court and to steer clear of including personal judgment in their verdicts. The road was thus paved, for what Twining (2006) called "the rationalist tradition" in jurisprudence.

Evidence in the Enlightenment

To outline the various social shifts that occurring during the Enlightenment that contributed to shifts in legal theory would be an extensive project on its own, and one that has at any rate already been undertaken by others (e.g., Hernandez Marcos, 2009). Suffice it to say that values of the Enlightenment (including the rights of the individual, the equality of persons, and principles of rationality) significantly shaped the legal proceedings of the time. The Enlightenment brought with it a legal tradition based on the value of rational legal procedures that could be applied to all individuals. It is in this time period that legal evidence came to play a role that will be more familiar to a contemporary audience. To summarize briefly, the establishment of truth and fact came to be the object of the Enlightenment trial. It was thought that this was best attained through rational, standardized procedures that could be equally and fairly applied to individual cases.

Evidence. The main evidence in the courtroom continued to be testimony. A novel element of the courtroom of the Enlightenment was the use of cross-examination of witnesses as a means of ascertaining the truth of their statements. In addition, the oath came to serve a new role in the ascertainment of truth; witnesses were asked to swear an oath to the truth of their testimony, a practice that continues to the present day (Hernandez Marcos, 2009). In addition, empirical evidence, or what Hacking (1975) calls "internal evidence," for the first time came to play a role in the legal process. Consistent with the empiricist epistemology of the day, the senses were taken to be a vital route to the ascertainment of fact. Thus, the presentation of physical objects that could be used to establish the facts of a case came to be particularly valued. Of course, some such evidence had been part of trials dating to Antiquity, but only in the Enlightenment did material evidence come to be valued above that of testimony. This shift can be directly related to the new value of objectivity that arose during the Enlightenment (Daston & Gallison, 2007). Increasingly, aspects of the individual person (including their perception and memory) were viewed as potential obstacles to the apprehension of truth. The physical object provided an objective proof of fact that could be apprehended by all who were present in the courtroom, and was not subject to the distortions of individual interests and perceptions.

Evidence itself, also for the first time, became a formal subject of law, with the emergence of various treatises on evidence (e.g., Gilbert, 1750; Wigmore 1913). The purpose of these treatises was to establish standardized procedures and rules regarding the admissibility of various kinds of evidence. For the first time, hearsay evidence was excluded from trials, as was the evidence of parties with a direct interest in the case (Oldham, 1994), and distinctions between the competence and credibility of witnesses were drawn. Prior to this time, evidence was presented in courtrooms and the judge and jury were allowed considerable discretion in their decisions of what evidence to consider in which manner. With the increasing value of objective, rational, and equal process, individual discretion was no longer desirable. The judge and jury now were expected to consider all the evidence presented. Subsequently, it became much more important to exclude evidence that was not to be considered (Twining, 2006).

The object of evidence. The goal of the trial in the Enlightenment was to determine the facts or truth of a case (Hernandez Marcos, 2009). This may appear to be a trite statement, unless the significance of the shift that occurred towards this objective is further elucidated. During the Middle Ages, and also in the Renaissance, God was taken to be the ultimate source and arbitrator of the law. Courts and human legal systems were a means to enforce laws but were always dependent on the divine for the ultimate determination of law. Relative to God, humans were afforded only limited knowledge of laws, facts, and principles of justice. In the Enlightenment, this conception shifted, and laws came to be viewed as man-made social contracts between the individual members of a society (Hernandez Marcos, 2009). Humans also came to be conceptualized as rational individuals, who could apprehend knowledge and reality through their reason. For the first time, it became theoretically possible for the jurors and judge in a courtroom to attempt to ascertain the facts of a case, through rational and logical analysis of the evidence.

In addition, however, principles of justice came to be equated with principles of fairness. Ideas such as the notion that the severity of a punishment should be proportionate to the severity of a crime arose as central tenets in a fair conception of justice. For the first time, the facts of a case became relevant and important to the judgment of the case. Judges no longer had only to decide whether individuals were guilty of the crime they were accused of, or in which they had a legitimate claim, but

were also required to retrace all the details of a claim or case, in order to reach a procedurally fair and just ruling. As a result, the facts and truths of a case came to play a much more important role in Enlightenment legal systems, first because they became relevant to the outcome of a case, and secondly because they were thought to be apprehensible in the first place. As Twining (2006) describes it, according to thinkers at the time "the direct end of adjective law is rectitude of decision, that is the correct application of valid laws (presumed to be consonant with utility) to true facts" (p. 41).

The relationship between evidence and its object. The emergence of evidence law as an area of jurisprudence brought with it a wide variety of rational procedures that were thought to be of use in objectively adjudicating evidence in a consistent manner. Principles of mathematics, including the newly created concept of mathematical probability came to play a central role in these procedures. It is not without accident that many of the early probability theorists (e.g., Pascal, Fermat, Leibniz) were formally trained in the law. In the oldest treatise on evidence by Gilbert (1750), what came to be known as the 'best evidence rule' was established. Gilbert devised a way to assign mathematical probabilities to various types of evidence and established a hierarchy of evidence based on its probability, with public records considered the most reliable source of evidence (If this sounds familiar to the reader of contemporary EBP literature, the parallel will be more clearly spelled out in later parts of this work). Gilbert's approach was not without its critics, including Bentham (1802), who in his antinomian thesis argued strongly that questions of admissibility and weight of evidence could not and should not be rule-governed. Rather, Bentham advocated for an inductive approach to the law, one that values completeness of evidence above all. Later theorists, including Wigmore and Chamberlayne, tended to criticize purely quantitative approaches to the law and attempted to separate admissibility from proof in consideration of the manner in which truth could be established in the courtroom.

By the time of the Enlightenment, the relationship between evidence and truth in the legal context was taken to be a logical one, if not a mathematical one, that could be studied in a scientific manner. Evidence was to be considered in its entirety and any evidence that might aid in a more complete apprehension of truth was deemed relevant. Objective evidence was valued over and above the subjective evidence of testimony, and legal procedures came to be viewed as rational processes. In fact, the presentation

of evidence in trials came to serve as a model for the demonstrative science of the modern age, as is explained further in the next chapter.

Evidentiary Practices in Law: A Summary

In Antiquity, testimony served as a main form of evidence in legal proceedings that aimed to re-establish and maintain order in society. Ordeals, wagers of law, and battle were used as means to obtain divine justice during the Middle Ages. Testimony and torture provided ways to establish legal judgments by juries during the Renaissance. Finally, testimony and objects provided a way to apprehend the facts of a case in modern trials, starting in the era of the Enlightenment. If the foregoing, cursory summary can highlight one thing, it is that practices of evidence in the law have been far from stagnant throughout human history. In order to understand the types of evidence that were considered valid in the legal contexts of various times, one must understand the objects of evidence or the objectives of the legal proceeding, as well as the manner in which the evidence was thought to be related to these objectives. In isolation from these legal practices, one could be left with a fragmented and decontextualized understanding, perhaps giving an impression that evidentiary practices had been deeply flawed and without rationality of any kind until the present time.

Chapter 3:
The Concept of Evidence in Science

Attempts to ascertain when the concept of 'evidence' began to play a role in science depend on one's focus, as well as the manner in which one defines 'evidence.' For example, those like Allen (2001), and to some extent also Franklin (2001), who seek to trace intellectual developments related to probability and evidence, define 'evidence' most broadly, and include all concepts whose role was akin to the present day concept of 'evidence' in scientific inference. Consequently, they describe conceptual predecessors of 'evidence' dating back all the way to science in Antiquity. Hacking (1975), on the other hand, who aims to explore the emergence of probability as a concept in Modernity defines 'evidence' more conservatively as inductive evidence connected to propositions through probabilistic relationships. He then views 'evidence' as a concept that emerged in the 17th century, during the time of the scientific revolution. As Hacking correctly asserts, it is difficult if not impossible to study the concept of 'evidence' in isolation from the concept of 'probability.' Thus, the present investigation draws heavily on some very valuable studies of the history of probability (e.g., Allen, 2001; Franklin, 2001; Hacking, 1975). The method used will employ case examples selected to illustrate broader ideas, concepts, and movements that arose during various time periods. Other examples could well have been selected, and no claim is made that the examples chosen are particularly important, relevant, or exhaustive, only that they provide samples of ideas that existed during certain time periods.

For the purpose of this study, practices related to current conceptions of 'evidence' in any of the domains related to current disciplines of science are of interest, and consequently very broad definitions of 'evidence' and 'science' will be adopted. 'Science' will be defined as any knowledge-seeking endeavour in the context of domains now considered to be part of science. 'Evidence' will be defined as anything that is used as the basis for making inferences in science. Practices related to all such inferences in

science and predecessors of science will be the focus of this study. The purpose will be to examine the manner in which scientific inference and practices related to such inference have shifted repeatedly and substantially over the course of human intellectual history. The word 'evidence' itself, I argue, is a concept of Modernity, and only quite recently entered the vocabulary of science during the scientific revolution in approximately the 17th century as part of the emergence of modern science. It would be a mistake to assume that conceptions of 'evidence' employed in Antiquity evolved into contemporary conceptions in a continuous fashion or were analogous to the current concept. Rather, the practice of defining and delineating what constitutes proper grounds for inference in knowledge-seeking ventures goes back as far as Antiquity. For present purposes, what is of interest, much more then when people came to speak of 'evidence' as they do now, is how scientific investigation across time relied on very different forms of inference and evidence than contemporary science, and how conceptions of evidence in science were contingent on the social context in which they were employed. My premise is that the current model of science along with the role evidence plays in this model emerged only in recent centuries. Practices related to utilizing evidence in reasoning and inference, however, date back to Antiquity. A study of these practices shows that even though science and knowledge without evidence, in a contemporary sense, seem unthinkable, science in fact has relied on very different kinds of evidence and conceptions of 'evidence' for much of documented human history. In order to support this argument, a look back at practices starting in Antiquity is necessary.

Evidence in Antiquity: Demonstration, Tokens, Signs, and Rhetoric

I begin my study of evidence in Antiquity with the rather bold assertion (albeit one that others, e.g., Hacking, 1975, have made before me) that there was no role for evidence, in the contemporary sense of the term, in the science of Antiquity. Rather, what could be considered evidence in Antiquity may seem rather foreign to the contemporary reader. The science of Antiquity was driven by metaphysics and rational demonstration. The purpose of science, according to the ancient thinkers, was to discover demonstratively certain knowledge through deduction from universal essences

and first principles. A demonstrative truth, in Antiquity, was defined as a truth that is certain (i.e., entailed), in that it could not be any other way. As such, no evidence is needed to establish demonstrative truths (I will elaborate this claim below). At the same time, I also argue that one would be remiss to assume that none of the ancient debates related to evidence mirrored contemporary debates on evidence. Thinkers in Antiquity contemplated at least some of the issues considered in discussions of evidence today. Thus, the focus of my discussion of Antiquity will be twofold. First, I will attempt to explain the practices of science in Antiquity and the role, or lack of role, of 'evidence' in its contemporary sense in the science of Antiquity. I then will highlight some of the intellectual contributions from the science of Antiquity to the concepts of 'evidence' of the modern period, while acknowledging that these contributions occurred primarily outside the realm of what was considered to be science and do not map directly onto current conceptions of 'evidence' and its role in science.

Aristotle's (384 BCE- 322 BCE) Demonstrative Science: Evidence as Knowledge[7]

The consensus story of science in Antiquity in most introductory philosophy of science or history of science textbooks goes somewhat as follows. Socrates and Plato did not concern themselves much with the physical world, which they took to be imperfect, compared to the world of ideas. Aristotle was the first natural philosopher who, despite his concession that the physical world may be imperfect and our senses fallible, devoted himself to observing and classifying the natural world (e.g., Ede & Cormack, 2004). Despite his focus on the natural world, however, Aristotle did not endorse an empirical, inductive approach to scientific investigation.

Aristotle viewed the universe as governed by first principles, essential truths that cannot be deduced from anything else and universal essences that determine current and future states of all aspects of the universe. The goal of science was to gain an

[7] The views articulated here are mere sketches of ancient thought. For more detailed analyses of these topics, I refer the reader to the excellent studies provided by Allen (2001) and Franklin (2001).

understanding of these essences and first principles and the truths that can be deduced from them. This is illustrated by the first paragraph in Aristotle's Physics.

> When the objects of an inquiry, in any department, have principles, conditions, or elements, it is through acquaintance with these that knowledge, that is to say scientific knowledge, is attained. For we do not think that we know a thing until we are acquainted with its primary conditions or first principles, and have carried our analysis as far as its simplest elements. Plainly therefore in the science of Nature, as in other branches of study, our first task will be to try to determine what relates to its principles. (Aristotle, Physics, trans. Hardie & Gaye, 1194, 184a, 10-15)

The process whereby knowledge can be gained through deduction from certain first principles has been referred to (in translation from Greek) as demonstration (Franklin, 2001). A demonstration does not establish what will occur but rather what caused it to occur and why it had to occur as it did. The goal of science, for Aristotle, was to establish the necessary causes of things (their essences), and science was to concern itself only with logically necessary, deductively established truths. Aristotle's notion of causality, however, was much broader than that of the present day.[8] Some of what Aristotle considered to be a logically necessary cause (e.g., that of which something is made – its material cause; the form something has - its formal cause, and the purpose something serves – its final cause) would be characterized by most contemporaries as a naturally or conceptually necessary cause. Aristotle took knowledge gained from demonstration to be necessary and certain. As Franklin (2001) puts it:

> Aristotle's ideal of science requires that the universal propositions used in demonstrations should be established by a process of understanding or insight, as happens in establishing the axioms of geometry. Ideally, one will know that all A's are B's with certainty because one will understand why any A must be a B – even in very empirical sciences like medicine. (p.112)

As such, it seems that Aristotle, despite his attention to the physical world, positioned science in the realm of logic. Although he was concerned with classifying nature (Franklin, 2001), his scientific explanations are divorced from empirical observation.

[8] Aristotle's notion of causality famously included material, formal, efficient, and final causes.

Hence, empirical evidence played a very minor if not insignificant role in Aristotle's science. Nonetheless, if one broadens one's conception of 'evidence' to include any basis of inference or knowledge, one could argue that first principles played the role of evidence in Aristotle's demonstrative science. In Aristotle's model, through deduction from first principles one could understand the necessary causes of nature.[9] To apply the framework of evidence provided above then, Aristotle's evidence was known, certain first principles, the objects of his evidence were the causes of nature, and the relationship between the two was logical, deductive necessity. To abstract even further, in much more general terms, it could be argued that Aristotle admitted as evidence into his science all logically necessary certain knowledge. He thereby made science an essentially rational venture that was fundamentally based on logic and favored absolute certainty over the ability to draw conjectural inferences. Aristotle's evidence may seem foreign to the contemporary reader who may associate evidence with degrees of uncertainty. It is important to remember, however, that uncertainty as Hacking and many others have argued, did not come to play a role in science until a much later time.

Uncertainty in Antiquity: Rhetoric, signs and tokens. Despite differences between contemporary models of evidence and signs and models of science in Antiquity, some of the issues related to evidence in modern science have their roots in ancient thought and can be traced to ancient accounts of signs, as argued by Allen (2001) and Franklin (2001). Ancient contributions to modern thought on evidence occur primarily in two contexts, theories advanced in rhetoric and practices in what could be described as the applied sciences of Antiquity. Both of these will be surveyed briefly below.

Allen (2001), to some extent echoed by Franklin (2001), argues that Aristotle's contributions to theories of evidence occur outside of his philosophy of science, in his rhetorical works, most specifically in his *Prior Analytics*. Recognizing that many arguments, which may not be deductive in nature and thus may not be logically valid

[9] Once again, it is important here to keep in mind that, although a discussion of Aristotle's model of causality would be beyond the scope of this paper, Aristotle had a very broad conception of causes that included material, formal, efficient, and final causes.

syllogisms can nonetheless be convincing, Aristotle provided what may well be one of the first structured accounts of non-demonstrative, and thus non-deductive, inference or enthymemes from tokens and signs (Allen, 2001). Allen views tokens and signs as predecessors of 'evidence.'

According to Aristotle, tokens and signs can provide evidence for a conclusion anytime they are more epistemologically accessible than the conclusion they support. For example, smoke that can be seen at a distance can provide evidence for a fire, which may not be visible at the same distance, or a person's pulse, which can be observed on their wrist and neck, can be evidence for their heartbeat, which is not easily observable. Aristotle distinguished between those things that furnish necessary evidence for a conclusion, which he called tokens (e.g., smoke is a token for fire), and those that only furnish "for-the-most-part" evidence for a conclusion, for which he used the more general label sign (e.g., wet ground is a sign of previous rainfall – alternatively it may be a sign of the fact that somebody spilled a jug of water). What distinguishes signs from tokens is their relationship with the conclusion for which they furnish evidence. A token necessitates the conclusion in the sense that if used as a premise that is true, its conclusion must also be true. In contrast, syllogisms based on signs merely establish the conclusion as a possibility. A premise based on signs can be true, while the opposite of its conclusion also can be true in particular instances. Thus, the differences between demonstration, understood as inference from token, and inference from sign were defined by the conclusion for which they serve as evidence. In demonstrations, premises provide a logically necessary explanation of a conclusion such that the conclusion necessarily follows from the premises and the premises also explain why the conclusion has to be the case. In inference from token, premises do not explain but merely establish a conclusion in a necessary way, such that if they are true the conclusion necessarily follows. Finally, inference from sign merely establishes the possibility of a conclusion, and the conclusion can be false, even if the premise is true. Nothing can be sign or token independently of the conclusion for which it is used in inference. The same thing (e.g., a person's accelerated pulse as felt on the wrist) could thus serve as a token of one conclusion (that their heart rate is accelerated) and a sign of another conclusion (that the person is anxious).

Overall, Aristotle created a first articulation of inductive reasoning and made room for a role for evidence by allowing for non-demonstrative inference, including inference from signs that only support conclusions "for the most part," and by classifying premises based on their relationship to that for which they serve as evidence. Nonetheless, it would be a mistake to interpret Aristotle's concessions towards non-demonstrative inference as support of a science based on non-demonstrative inference from observation. As Allen (2006) puts it, "signs furnished by observation play a negligible part in Aristotle's official method of science, with its focus on the grasp of first principles by intuition and explanation by demonstration from them" (p. 258).

Sextus Empiricus (160-210): Indicative and Commemorative Signs

Following Aristotle, a number of theories of signs were advanced in Antiquity that also bear mentioning due to their relevance to accounts of 'evidence.' Sextus Empiricus, in an attempt to oppose dogmatism in various areas of theory and practice,[10] clearly articulated the relationship between observed signs and the unobserved nature of that which they signify. He distinguished between signs based on the epistemological accessibility of that which they signify, by distinguishing between "commemorative" and "indicative" signs (Manetti, 1993).[11] The commemorative sign co-occurs with that which it signifies and is directly associated with that which it indicates. In the case of the commemorative sign, that which it signifies is at least in principle observable, although its observability may be temporarily obscured, necessitating the use of signs to infer its presence. For example, the presence of daylight may serve as a sign for the presence of the sun in the sky when the sun is temporarily obscured from vision by clouds, given that as a general rule daylight and the sun co-occur. Indicative signs on the other hand indicate something that is unobservable and thus are never co-observed with that which is signified. For example, the symptom of fever may signify the presence of a disease organism, although the disease organism itself may not be observable. Sextus Empiricus dismissed the legitimacy of indicative signs, arguing that such signs were

[10] Particularly medicine. Sextus Empiricus was well-aligned with the empirical school of medicine, see discussion below.

[11] It is unclear whether this distinction should be attributed to the Stoics or is falsely credited to them, see Manetti (1993) and Allen (2001).

based on dogmatic inference, but defended the use of commemorative signs as a means of valid inference.

While Aristotle argued that signs must be more epistemologically accessible than that which they signify, and classified signs based on the strength of the probabilistic relationship between signs and what they signified, Sextus Empiricus classified signs based on the epistemological accessibility of the signified, rejecting the validity of inference to signs that, in contemporary terms, would be called "in principle unobservable."

The Stoics: Signs as Antecedents

The Stoics defined the sign as a true antecedent proposition in a conditional with a true consequent (Manetti, 1993). Contrary to Aristotle who viewed signs as a part of rhetoric that, although permissible and convincing, is separate from and inferior to a valid logical syllogism, the Stoics situated the sign in the domain of logic. The Stoics distinguished between signs based on their relationship to that which they signify, using the terms 'common' and 'particular.' In a distinction that mirrors Aristotle's own between signs and tokens, the common sign can occur independently from that which it signifies, while the particular sign stands in a necessary relation to that which it signifies, and can co-occur only with that which it signifies. The Stoics dismissed the value of common signs, which do not signify with certainty, and limited the pursuit of knowledge to inference from signs that necessitate that which they signify. This is consistent with their conception of signs as antecedents in true conditionals; the Stoic sign must stand in a necessary conditional relationship to the consequent that it signifies. However, the necessity they require, as Allen (2001) points out, is not one of necessary, logical entailment as in Aristotle, but one of constant and consistent co-occurrence. The Stoic sign was the antecedent in a valid conditional argument, where that which is signified is the consequent and the relationship of the sign to the signified is a necessary relationship. The Stoics also defined signs based on their relationship to the signified but defined this relationship as constant co-occurrence, in a manner that was less restrictive than Aristotle requirement of logically necessary co-occurrence.

The Epicurean Conception of 'Evidence': Signs as Analogy

Contemporaries of the Stoics, the Epicureans endorsed a conception of signs that was much more inductive in nature than most previous accounts. According to Franklin (2001), they rejected the notion that inference from the observed to the unobserved, as in the indicative signs of Sextus Empiricus, was inherently problematic, and provided a logical framework for such inference. Arguments from observed instances to unobserved instances, they argued, can be grounded through similarity in the logical form of an analogy. Through analogy, we can gain knowledge about unobservable things as long as they are similar to observable things in a sufficient way. Allen (2001) argues that the Epicureans thus bridged empirical and rationalist accounts of signs by allowing for projection of empirical experience onto analogous unobserved instances through rational analogical inference or projection. This account may seem both familiar and foreign to the modern reader. It mirrors current conceptions of evidence in its inductive nature and yet bases induction on analogy in a way that seems far-reaching. At the same time, the Epicurean account highlights the analogical nature of inductive inference, or more simply put, the assumption that that to which inference is drawn is similar to that which is observed, an assumption that is common in contemporary scientific investigation.

Converging Signs: The Rhetoric of Cicero (106 – 43 BCE)

Roman thinkers were famously concerned with the oratory arts and much Roman thought on signs and their role in inference is to be found in rhetorical works. Most famously, Cicero in his *De Inventione* further legitimizes inference from signs and distinguishes such inference from more deductive reasoning. Recognizing that there are various kinds of probable[12] arguments, Cicero distinguished four kinds of argument: (1) proof, (2) credible, (3) already determined, or (4) compared with something else. He

[12] Consensus among scholars (e.g., Allen, 2001; Hacking, 1975) is that "probable" as it is used here and in writings up to the Renaissance may be understood to mean "convincing" or "believable" and not probable in a mathematical sense or even in the sense of being likely to occur.

then specified the manner in which each of them can be established. Cicero defined proof as follows:

> That is a proof which falls under some particular sense, and which indicates something which appears to have proceeded from it, which either existed previously, or was in the thing itself, or has ensued since, and, nevertheless, requires the evidence of testimony, and a more authoritative confirmation, -- as blood, flight, dust, paleness, and other tokens like these. (Cicero, De Inventione, XXX)

Here, Cicero to some extent seemed to draw on earlier accounts of signs, but also introduced several new aspects in his conception of proof. Things that are credible are those that will be accepted because they are consistent with the existing opinion of the audience. Arguments for things that are already determined involve premises that have been established through common agreement, by authoritative judgment, or through consensus judgment. Arguments for things that are compared with something else are most similar to the Epicurean analogy and involve demonstrating that things in nature resemble each other, that they are alike because they share certain features, or providing an example that either confirms or invalidates an argument about things that are similar to the example.

When reading this account, it quickly becomes clear that what distinguishes these propositions to some extent is the level of support they require for their acceptance. Credible arguments are accepted by virtue of being consistent with the opinion of those to whom they are presented, while those that have been established beforehand are deduced from previously agreed statements of fact or value. Neither requires testimony or establishment of fact. In arguments from comparison, mere description and enumeration of relevant similarities suffice to ground the argument. Proof seems to be the only form of argument that requires what could be thought of as supporting evidence, either through testimony or through physical evidence (blood, flight, dust). Here, Cicero also appears to specify that in the case of proof, multiple pieces of converging evidence are necessary, including evidence from testimony and confirmation through tokens (what could be considered physical evidence).

Cicero's views on the necessity of proof become more explicit in his discussion of parts of an argument (Cicero, XXXV-XL). He proposes that arguments generally consist

of five parts: a proposition (some statement of fact), proof of that proposition, an assumption (an inference from the statement of fact), proof of that assumption, and a summary. He supports his choice of arguments in five parts as follows:

> And from these considerations that also is evident, that there is a certain kind of argumentation in which neither proposition nor assumption stands in need of proof, of this sort, that we may adduce something undoubted and concise, for the sake of example. "If wisdom is above all things to be desired, then folly is above all things to be avoided; but wisdom is to be desired above all things, therefore folly is above all things to be avoided." Here both the assumption and the proposition are self-evident, on which account neither of them stands in need of proof. And from all these facts it is manifest that proof is at times added, and at times is not added. From which it is palpable that proof is not contained in a proposition, nor in an assumption, but that each being placed in its proper place, has its own peculiar force fixed and belonging to itself. (Cicero, DeInventione, XXXVII)

Cicero therefore seemed to draw two distinctions, first between premises and proof of these premises and secondly between premises that require proof and those that are self-evident and do not require proof. Contrary to Aristotle's model of science, which was based on inference from self-evident principles, Cicero's model of rhetoric dealt with the need for proof in the context of rhetoric.

Summary: Rhetoric and Signs in Antiquity

Aside from providing an overview of early articulations of issues related to 'evidence,' some of which will seem very familiar to the contemporary reader, the foregoing discussion of signs in the logic and rhetoric of Antiquity shows the extent to which, despite a lack of role for empirical evidence in Aristotle's science, a rich body of theoretical discussions related to what would now be considered evidence can be found in the realm of rhetoric and in accounts of signs. It is notable that some of the theories that may well have provided the basis for later thought on 'evidence' and its role in science, emerged in a context where concern was with convincing others of an account or views much more than with the establishment of facts about the world and that the knowledge established in such a manner was considered to be less certain and thus epistemologically inferior to that established by rational logical inference alone.

Evidence in the Applied Sciences in Antiquity: Astrology/Divination, Physiognomics, and Medicine

As previously mentioned, the role of signs and tokens in inference in Antiquity was limited and science was concerned with deduction from logical principles, the explanation of necessary causes, and mathematical proof. Beyond the theoretical realm of rhetoric, practices involving signs in Antiquity emerged in areas that were considered to be arts, rather than parts of science. These areas include astrology and divination, physiognomy, and medicine. Many (e.g., Allen, 2001; Daston; 1991; Franklin, 2001, Hacking, 1975) have argued that current practices related to evidence can be most clearly traced to their roots in these contexts.

Signs and Divination in Antiquity

According to Allen (2001), one of the main uses of signs and tokens outside of rhetoric occurs in practices of divination. Astrology and divination were widespread in Ancient societies, as is apparent from the use of horoscopes by the Babylonians (Franklin, 2001) or the Greek's reliance on oracles (Franklin, 2001). It is clear that signs and inferences played an important role in these practices. For present purposes, an exploration of the Stoic account of divination will serve to explicate this claim, as their practice is well documented. The Stoic conditional account of signs (explored briefly above), Allen (2001) argues, mostly emerged from a need to account for the various types of signs used in the context of divination. The divination of the Stoics was based on the belief that "the world is ordered by divine reason, by which it is providentially ordered for the benefit of mankind" (Allen, 2001, p. 165). As humans possess some of that divine reason, they are able to recognize the divine order from signs. Subsequently, Stoics espoused both a natural and artificial variety of divination. The former occurs through dreams and prophetic revelations, and can be interpreted by anyone, while the latter requires specific skill or ability in divination. Artificial divination relies on either lengthy observation of constant conjunctions (e.g., the constant co-occurrence of certain astronomical phenomena with personality characteristics of those born during the occurrence of those phenomena) or the conjectural assignment of meaning to a sign (e.g., a swarm of locust as a sign of the impending decline of society). The signs or

evidence in Stoic divination consisted of natural occurrences, unusual occurrences, dreams, and revelations. In order to establish inference from such signs, the Stoics had to allow for a non-deterministic conditional relationship between signs and their consequences, as of course none of them was consistently related to guaranteed outcomes. Stoic divination merged observations of nature and the use of inferences drawn from these observations to predict future occurrences. The Stoics, however, were not concerned with explaining the causes of natural occurrences or relationships, but rather with their utility in predicting future occurrences, which at the time separated divination from science. In divination, signs provided evidence for the future and the relationship between signs and the future events they signified was non-deterministic.

Physiognomics: The Signs of Human Appearance

Physiognomics, a practice of Antiquity, involved the inference of personal characteristics from external appearances. Franklin (2001) states that various types of inferences were used to interpret these signs, including drawing similarities to animals, using analogies, noting differences between genders, and suggesting causal explanations of the sources of physical differences. Siraisi (2004) stresses that physiognomy could be understood as the interpretation or reading of signs to draw inferences about the natural world. Visible signs (e.g., facial features) were taken to be cues that allow for knowledge about the invisible (e.g., character) and prediction of the future (e.g., behaviours of individuals). The signs in physiognomics were varied and the use of multiple signs was encouraged as sound practice. Franklin argues that it was in the context of such non-deductive practices that the need to consider multiple signs emerged. Physiognomy represents an early practice of drawing inferences from visible signs to invisible traits and using signs to predict future occurrences, as well as combining multiple converging signs to increase the certainty of inferences from these signs. Physiognomics became influential again in the early 19th century, and influenced popular culture and literature beyond that point, but was replaced by the mental testing of psychologists in the early 20th century (Collins, 1999).

Medicine: The Dogmatist and Empiricist Schools

Inference from sign also played a major role in the practices of Ancient medicine. In medicine, inferences about causes of disease and predictions about possible cures had to be made and it is not surprising that the interpretation of signs (or symptoms) was central to this practice. However, the manner in which signs were utilized differed greatly between the major schools of medicine in Antiquity: the dogmatist, the methodist, and the empiricist schools (Hankinson, 1987). For the purpose of the present discussion, the empiricist and dogmatist views will be contrasted as examples of two very opposite models of medicine (the methodist model is often characterized as a middle ground between these two positions).

The dogmatists, a group of medical practitioners and theorists, including Aristotle, essentially conceptualized reason as the connection between signs that can be known through perception and the signified that is imperceptible. They were concerned with understanding the nature and causes of diseases of the human body. They thought that one must grasp the invisible causes or nature of disease in order to know what might cure the disease (Allen, 2001). Because the causes and nature of disease are usually invisible, the dogmatists thought that they could be grasped only through reason. In other words, the dogmatists promoted a form of medical practice that relied on what Aristotle would have considered demonstration: they advocated not just for the establishment of a relationship between a disease and its symptoms or signs, but also a logical explanation of the reason for this relationship through mechanisms whereby the disease causes the symptoms or anatomical principles that explain the relationship. Allen (2001) draws a parallel between the dogmatist conception of signs and Sextus Empiricus' and the Stoics' distinction between commemorative and indicative signs. Allen argues that the dogmatist conception viewed symptoms in medicine as indicative signs, which pointed to an underlying imperceptible cause.

The empiricist school of medicine emerged in response to the dogmatist school (Franklin, 2001) and rejected the notion that it is possible to gain knowledge about causes or the nature of disease that cannot be observed through the senses. Empiricists espoused what Allen (2001) considers to be a commemorative conception of signs, arguing that medical knowledge can occur only through the observation of the co-

occurrence of signs. Promoting an inductive model, they rejected the notion of knowledge from single observations, which they thought was unreliable, and argued that a large number of observations of the co-occurrence of signs with disease is necessary in order to establish knowledge (Franklin, 2001). The empirical school relied on experience through the senses as the sole source of knowledge and conceptualized experience as "a collection of instances sufficient to ground a theorem" (Hankinson, 1987, p. 332). Its practitioners allowed for the drawing of experience from others as transmitted in writing or oral teaching, but also established standards to evaluate the credibility of such accounts of experience (Franklin, 2001). To borrow the distinction made by the Stoics (many members of the empiricist school were Stoics), the empiricists rejected the validity of indicative signs and espoused only the commemorative sign (Allen, 2001).

Overall then, the empiricists and dogmatists espoused radically different conceptions of signs. The dogmatists viewed signs used in conjunction with reason as means to gain knowledge of invisible causes and nature. The empiricists on the other hand were concerned only with signs, which co-occurred observably with the signified, and viewed observation of numerous signs and their correlates as the main avenue to knowledge.

Evidence in Antiquity: A Summary

A few themes emerge from this brief exploration of signs in Antiquity. First, as has been argued by many others (e.g., Allen, 2001; Franklin, 2001; Hacking, 1975), the concept of the sign in Antiquity most closely approximates what we now consider 'evidence.' Secondly, signs did not play a major role in the science of Antiquity, a science that sought to establish certain conclusions through deductive logic. Instead, signs played a role primarily in rhetoric and conjectural practices such as divination/astrology, physiognomy, and medicine. It would be a mistake to view the role of signs in Antiquity as directly equivalent to the role of evidence in modern days. Signs were assigned much less epistemological weight than deductive logic in the search for knowledge in Antiquity.

Nonetheless, ancient theories of signs do anticipate many of the later debates surrounding evidence that continue into the present day and their contributions to current accounts of evidence are apparent. Aristotle established the rhetorical legitimacy of inference from signs and tokens, and described the relationship of signs to what they signified as one that may not be certain or deterministic. Sextus Empiricus highlighted two conceptions of signs: those that co-occur with what they signify (commemorative signs) and those that point to something that cannot be observed (indicative signs). Both he and the Stoics moved signs into the realm of logic by conceptualizing them as antecedents in conditional arguments, ultimately rejecting the legitimacy of the indicative sign. The Epicureans endorsed the legitimacy of inference to invisible instances from signs through the use of analogy. Both astronomers of Antiquity and Roman rhetoricians examined the manner in which divergent evidence can be used as a source of unified inference. Cicero further established the notion that proof may be necessary for some types of argument, while other arguments may be established deductively. Finally, the practices of divination, physiognomy, and medicine used signs as a source of inference to invisible causes (e.g., in the dogmatist account of medicine) or a source of prediction of future instances.

As mentioned in the evidential framework provided in the introduction to this thesis, accounts of 'evidence' differ with respect to the strength of the relationship they espouse between evidence and its objects. Although some accounts require that the relationship between evidence and its objects is one of logical or observable necessity, other accounts permit a relationship of increased likelihood or simple similarity. Secondly, accounts also differ with respect to the objects of evidence. A major distinction that emerged in Antiquity is the distinction between accounts that permit objects of evidence to be invisible or in principle unobservable and accounts that require an observable correlation between evidence and its objects. In the former account, evidence provides clues to the invisible, while in the latter it merely allows for prediction of future instances. Lastly, accounts differ in terms of how evidence itself is viewed and what can count as 'evidence' (e.g., physical features in physiognomy, dreams and revelation in divination, symptoms of disease in medicine, previous knowledge in empiricist medicine). It will soon become apparent that this diversity continues to be reflected in later accounts of 'evidence.'

Evidence in the Middle Ages

The decline of the Roman Empire, it is often argued, marked the end of scientific and intellectual pursuits, as civilization was thrown into the chaotic "dark ages" and the spread of Christianity promoted a life governed by faith rather than the pursuit of knowledge. Nonetheless, within both Western and Islamic societies, developments in conceptualizations of 'evidence' continued to occur during the 'Middle Ages,' defined here as the time period between the end of the Roman Empire and the beginning of the Renaissance (approximately 400-1400 CE). In Western society, the Middle Ages witnessed metaphysical and epistemological shifts with respect to 'evidence.' In terms of epistemology, thinkers in the Middle Ages distinguished the role of knowledge derived from first principles and knowledge derived from opinion, mirroring Aristotle's distinction between rhetoric and demonstration. Metaphysically speaking, however, demonstrative knowledge and knowledge derived from opinion increasingly became conceptualized as partial and epistemologically inferior or incomplete relative to the more complete knowledge of the world afforded to God. Science increasingly came to be viewed as a way to apprehend God's knowledge through human means. These epistemological and metaphysical shifts brought with them changes with respect to 'evidence,' its object, and the relationship between the two. Conceptions of 'evidence' in the Middle Ages were broadened to include testimony and empirical data. Increasingly, the goal of science became the partial apprehension of divine, absolute knowledge (complete apprehension of divine absolute knowledge was thought to be reserved to God himself, and thus impossible to attain). Subsequently, the use of authority and, later in the Middle Ages, sense data, increasingly came to be viewed as providing evidence for absolute reality, apprehended through divine illumination. Further, relationships posited between testimony and sense data and absolute, divine reality were thought to be incomplete, uncertain, and necessarily limited.

Augustine (354-440): Evidence as Testimony

Augustine of Hippo in many ways represents the transition from the Roman Empire to Christian thought, as he himself converted to Christianity in the course of his life (King & Ballantyne, 2009). Upon his conversion to Christianity, he devised an

epistemology that supported the Christian doctrine and as such served as a bridge between the philosophers of Antiquity and the Christian thinkers of the Scholastic age. The basis of knowledge, according to Augustine, was neither logically necessary deduction, nor inference from signs. In fact, Augustine viewed the world as contingently dependent on God's power and will, and human knowledge as dependent on divine illumination in the same way that human vision is dependent on illumination through light. Consequently, he was doubtful about the utility of inferences from necessary essences inherent in nature. Instead, his epistemology was closely connected to testimony and faith (Beilby, 2002; King & Ballantyne, 2009). Knowledge of divine power and causes, according to Augustine, is based on perceptions and testimony. One can perceive external stimuli or the contents of one's mind. Testimony, however, is provided by others about their perceptions or the contents of their minds and is apprehended through faith. Augustine advocated that faith in testimony ought to be the default position unless one has reasons or evidence that justify a lack of faith in testimony. The role of additional evidence to corroborate testimony for Augustine was only to distinguish between two equally plausible testimonial accounts (Beilby, 2002). In such instances, ultimately belief was still grounded in testimony and non-testimonial evidence merely served to select a testimonial account. As such, testimony was epistemologically superior to all other types of evidence in Augustine's account. In addition, Augustine conceptualized signs as testimony to the signified. Signs apprehended through the senses could therefore be conceptualized as testimony to that which they signified. This shift in conceptions of signs is quite important as it laid the groundwork for later Western accounts of evidence that will be elaborated below (King & Ballantyne, 2009).

Evidence Produced through Experiment:
Avicenna (980-1037) and Middle Eastern Thought

As Europe went through the Dark Ages, scientific progress continued in other parts of the world, including the Middle East. The origin of the scientific experiment is often traced to this time period and location, in particular to the medical works of Avicenna, who later came to influence Western thinkers during the Middle Ages. Avicenna was concerned with finding a way to separate necessary or essential effects of substances from accidental effects that occur through mere coincidence (Franklin,

2001). He developed a set of procedures to test the effects of substances, which included ensuring the purity of the substance administered and, additionally, ascertaining that the effect occurs continuously and in temporal proximity to the administration of the substance. In other words, Avicenna conceptualized evidence for causal inference, as something produced through a carefully considered and executed process. For the first time, 'evidence' came to be seen as produced through a humanly constructed procedure rather than something that is simply observed by the human senses or revealed immediately to the human intellect. Avicenna's idea that some observed attributes of substances or effects can be accidental rather than necessary would influence later thinkers of the Middle Ages, including William of Ockham, who will be discussed in more detail below.

The Scholastics: Evidence of the Divine

The scholastics, consistent with Augustine, questioned the possibility of the attainment of certain human knowledge through demonstrations. Nonetheless, they proposed a divinely ordered universe, which could be understood in a limited way through human faculties of sensing and intellect. In a perhaps surprising manner, this shift paved the way for greater attention to individual instances as perceived by the senses, in contrast to earlier attempts to understand universal essences.

Thomas Aquinas (1225-1274): Evidence for uncertain knowledge of first principles and opinion. Aquinas was a clear proponent of the incomplete nature of human knowledge relative to God's knowledge. Evidence, in the contemporary sense of the word, did not play a major role in the metaphysical views and epistemology of science of Aquinas (as argued by Hacking, 1975). Although he was heavily influenced by Aristotle's work, Aquinas' metaphysical views ultimately came to be consistent with a Christian worldview. As such, while Aristotle viewed demonstrative inference as a source of certain and complete knowledge, Aquinas ultimately viewed all human knowledge as inferior and incomplete relative to divine knowledge. In an ascending hierarchy from human to angelic to divine knowledge, Aquinas argued that reality is completely apprehended in its original form only by God. Human knowledge occurs indirectly through God and inferior material means. In his Summa Theologica, he wrote

"In this way the soul, in the present state of life, cannot see all things in the eternal types; but the blessed who see God, and all things in Him, thus know all things in the eternal types" (Summa Theologica, Part 1, Question 84, Article 5). Byrne (1968) emphasizes that to Aquinas the very fact that human knowledge has to be acquired by means of a process of inquiry bore testimony to the deficiency of the human intellect, as only God apprehends knowledge directly and automatically rather than through specific intellectual or sensory processes.

Within the imperfect realm of human knowledge, however, Aquinas identified two main sources of knowledge: demonstration and opinion. Byrne (1968) argues that, like Aristotle, Aquinas ultimately viewed demonstrative knowledge from necessary causes and first principles as the sole source of scientific knowledge. However, the first principles from which legitimate scientific knowledge could be inferred were thought to be logically necessary by Aquinas. They included the principles laid out in religious scripture rather than those deemed logically necessary by Aristotle's standards. In light of the imperfect nature of human knowledge, Aquinas also allowed for a second source of knowledge, knowledge from opinion. Opinion, according to Aquinas, was non-certain knowledge, which, if it is to be considered probable, must be arrived at through a process of dialectical reasoning (Byrne, 1968). This process is evident even in the structure of Aquinas' writing itself, which usually consists of a question followed by a review of common views on the question (difficulties), opposing views (to the contrary), Aquinas' view (reply), and a discussion of common and opposing views (answers to difficulties/answers to contrary difficulties).

What supports an opinion is a consideration of authoritative views through dialectical reasoning. Opinions are in some ways grounded in testimony or authoritative views and an opinion that is probable is one that is supported by authority and consistent with first principles (as argued by Byrne, 1968). Although ultimately Aquinas viewed opinion as outside the realm of science because it does not yield certain knowledge through demonstration, he still viewed it as a legitimate source of knowledge. Aquinas also viewed both sensory experiences and intellectual reasoning as essential means of acquiring knowledge and opinion: "We must not expect the entire truth from the senses. For the light of the active intellect is needed, through which we achieve the unchangeable truth of changeable things, and discern things themselves from their

likeness" (Summa Theologica, Part 1, Question 84, Article 6). At the same time, he viewed knowledge of particulars as directly achieved through sensory means and indirectly through the intellect, while knowledge of necessary principles occurs by necessity through the intellect.

> Therefore the contingent, considered as such, is known directly by sense and indirectly by the intellect; while the universal and necessary principles of contingent things are known only by the intellect. Hence if we consider the objects of science in their universal principles, then all science is of necessary things. But if we consider the things themselves, thus some sciences are of necessary things, some of contingent things. (Aquinas, 1473/1920, Part 1, Question 86, Article 3)

Nonetheless, all human knowledge is ultimately imperfect relative to God's knowledge. However, gradients of imperfection exist. Knowledge through demonstrations from first principles (derived from the use of sensory experiences in conjunction with the intellect) is less imperfect than knowledge derived from opinion (which again consists of sensory experiences in conjunction with intellectual reasoning in a dialectic manner).

Ultimately, according to Aquinas, all knowledge is apprehended by humans through inferior intellectual and material ways driven by the power of God. This view was articulated by Aquinas in the following: "passive intellect is reduced from potentiality to act by some being in act, that is, by the active intellect, which is a power of the soul, as we have said; and not by a separate intelligence, as proximate cause, although perchance as remote cause" (Part 1, Question 84, Article 4). The senses and intellect thus represented to Aquinas the inferior processes humans must rely on to apprehend knowledge. The two main sources of human knowledge for Aquinas, or in contemporary words, his evidence, were first principles derived from logic and religious doctrine and authoritative opinion examined dialectically in light of first principles.

Robert Grosseteste (1175-1253): Evidence of divine illumination. In line with Aquinas, Grosseteste also viewed human knowledge as uncertain relative to God's knowledge and viewed human senses and intellect as imperfect avenues to knowledge, reflecting the limits of human bodies (Laird, 1987). A departure from Aquinas however, is the manner in which Grosseteste believed that knowledge is apprehended. All knowledge, he thought, occurs through divine illumination. In a process akin to physical

illumination through light, Grosseteste thought that God creates the human apprehension of knowledge through the senses and intellect. Grosseteste also argued that, because this process is akin to vision, it follows similar mathematical laws, such as that governing the refraction of light (Serene, 1979). Thus, he saw mathematics as a useful tool in the apprehension of knowledge. He also viewed the universe as somewhat ordered according to laws that can be apprehended. He supported the use of tools such as mathematics and experiments, in the service of deriving premises of demonstrations for what he considered to be the highest level of knowledge, apprehension of the divine. Some (e.g., Crombie 1958) propose Grosseteste as the father of western experimental science. However, others (e.g., Serene, 1979) argue that the role of experiments in his epistemology has been exaggerated. Experiments were not a primary source of evidence for Grosseteste; rather they simply constituted one way to correctly identify divine illumination. Ultimately, for Grosseteste, evidence for divine illumination was apprehended through the senses and the use of reason (including mathematics) in a demonstrative process. Experiments merely constituted one way to generate sense data.

Roger Bacon (1214-1294): From demonstration to experience. In the work of Aquinas and Grosseteste, a shift away from demonstrative certainty is evident in their arguments about the fallibility of human knowledge. Nonetheless, both thinkers advanced various ways in which incomplete human knowledge can be attained through illumination by God in the context of demonstration guided by the intellect. Despite an increased recognition of the importance of sensory experience, these authors continued to privilege logical demonstration as a main source of human knowledge, consistent with Aristotelian epistemology.

Among the later philosophers of the Middle Ages, the shift away from logical demonstration towards a greater emphasis on experience as a source of knowledge became more pronounced. One thinker who will serve to exemplify this shift was Roger Bacon. Bacon, who attempted to formalize a method for the acquisition of knowledge in science, identified three main sources of knowledge: authority, reason, and experience. In doing so, he was consistent with previous thinkers. As a point of departure, however, Bacon dismissed both the value of authority and reason in acquiring certain knowledge about the natural world, relative to experience (Fisher & Unguru, 1971). Authority, he

argued provides belief without understanding, while reason does not allow one to distinguish between arguments that are true and those that only appear so. Thus, he argued, experience is the only means to apprehend truth completely. It is important, however, not to conceive of experience in its contemporary sense, as pure sensory experience. Rather, for Bacon, experience through the senses was limited, and more complete knowledge could be acquired through experiences of divine revelation, which he considered to be a superior form of experience (Bacon, Opus Majus, II, 169).

Furthermore, Bacon considered mathematics a source of certain experience, which should form the basis of all scientific knowledge. The truths of mathematics (geometry and algebra) can be directly experienced with certainty, he thought, while the remaining sciences were uncertain. Consequently, mathematics must serve as the foundation of all sciences, if they are to achieve certain knowledge. The evidence of science, for Bacon, was experience of truths, ideally formalized in mathematics. This shift, Fisher and Unguru (1971) argue, aligns with Bacon's conceptions of the purpose of experimental science. Although previous thinkers often viewed experimental science as a way to supply premises for the demonstrations of other sciences, Bacon viewed experimental science as a way to verify the conclusions of other sciences, to establish facts beyond the grasp of other sciences, and to establish practical knowledge (in the sense of it being knowledge that can serve the church). Thus, he attributed a more independent value to experimental science than did his predecessors.

William Ockham (1288-1347): The evidence of particulars. The shift towards an epistemology that gives primacy to experience is further apparent in the thought of William Ockham. Ockham, consistent with Christian sentiments of his time, posited that, given God's omnipotence, reality must be radically contingent (Goddu, 1984). He argued that God is unconstrained in the world and the reality he can create. This led him to his doctrine of occasionalism, the notion that God could have created the world differently in any way he wished. Using this argument, Ockham rejected the Aristotelian notion of universal essences. There are no universal essences to be known, but rather particular things that can be apprehended through a process he called intuitive cognition, the apprehension through the senses, which is the main source of knowledge. Universals are the result of a process he called abstractive cognition and are essentially a product of human imagination. Ockham advocated not only a science based on

experience, but a science based on experience of particulars. As sources of knowledge, he allowed for only experience and revelation (divine revelation). It is here that a major shift becomes evident; experience became a source of knowledge in its own right. Thus, experience of singular instances became primary, in effect reversing the epistemology of Aristotle. Laws that could be abstracted from particulars are of limited necessity in Ockham's thought. Consequently, evidence in scientific inference, for Ockham, was experience of individuals through the senses or divine revelation and that, which can be deduced from such experience. In a universe that is divinely ordered and ultimately dependent on divine will, human knowledge is always contingent and the propositions of science are possible and conditional -- they could be otherwise if God chose to make them so.

Evidence in the Middle Ages: A Summary

In very broad brushstrokes, it can be argued that the shift towards Christian metaphysics brought with it a number of epistemological shifts in the Middle Ages that seem to have elevated the role of experience – first, in the form of evidence by authority and later, in the form of direct sense experience, replacing the emphasis on reason that was prevalent in Antiquity. Augustine rejected the possibility of human knowledge without divine illumination and created a place for testimony based on experience and divine revelation as sources of evidence towards knowledge. In a separate development in the Middle East, evidence produced through experiments came to be viewed as a source of knowledge. Aquinas further emphasized the fallibility of human knowledge relative to divine knowledge, arguing that reliance on senses and intellect is a necessary consequence and limitation of human knowledge, and thus uncertainty entered into science. Grosseteste echoed the role of the divine in human knowledge, likening it to a process of vision through divine illumination. He opened the door for geometry to play a role in understanding the universe, while continuing to maintain his view of experience as a mere source of premises in logically necessary demonstrations. Bacon then moved away from the privileging of demonstrations and instead advocated the importance of experience in order to verify and create knowledge. He viewed mathematics as a certain and experiential science. Finally, Ockham began to advocate for the importance of experience of individual instances as a source of evidence. He dismissed the notion

of universal essences, which he thought was inconsistent with divine omnipotence. However, the shift towards what could be described as an increasingly experience-based (albeit not empirical in its current sense) epistemology was founded on a metaphysics that relied on divine omnipotence and illumination.

A Brief Summary of Intellectual Shifts in Antiquity and the Middle Ages

The abridged story of 'evidence' told thus far can be summarized as follows. 'Evidence' did not play a major role in the science of Antiquity. Ancient science was based on certain inference in the form of logically necessary demonstration and deductive inference from essential properties and first principles. The goal of science in Antiquity was to establish not only what the case was but also why it had to be the case. In other words, science sought to understand Aristotelian causes or ways of knowing (material, formal, efficient, and final cause). There was thus no need for evidence to ground inferences in demonstrative science. Explanatory principles or causes were inferred through a logical process based on reason and known with certainty. This is not to say, however, that evidence was not used in Antiquity to ground inferences. Rather, non-demonstrative inference was relegated to epistemologically inferior practices such as rhetoric and to the realm of soft sciences. 'Evidence' in these contexts was conceptualized as sign. Ancient works on inference from sign focused on clarifying the relationship between signs and the signified. Ancient thinkers conceptualized this relationship in degrees of certainty, conditional and analogical relationships, and whether the sign co-occurs with the signified or signifies something that is empirically inaccessible. In terms of practices involving inference from sign, such practices mostly occurred in the context of areas that at the time (or at present) would not have been considered part of the realm of science, such as divination, physiognomy, and medical practice. In these practices, the use of empirically accessible signs was established as a source of inference to conclusions about objects, properties, and processes that were empirically inaccessible. Nonetheless, it is important to remember that such practices were not thought to yield knowledge and were considered inferior to the certain knowledge of science and mathematics.

In the Middle Ages, opinion came to be viewed as a source of knowledge in addition to demonstration. Opinion it was thought, was knowledge based on authority and apprehended through faith in that authority. Much of the work of the Scholastics involved dialectical examination of the works of ancient philosophical authority (that of Aristotle) and divine authority (in scripture). The ultimate authority, the source of all knowledge in Medieval thought, was the omniscient God, who shared some knowledge with humans through divine illumination. Thus, the evidence of the Middle Ages can be conceptualized as authority, the object of which is divine knowledge. The relationship between evidence and its object was faith that privileged divine illumination.

The distinction between knowledge based on non-demonstrative inference from empirical observation, relative to the logically deduced knowledge from demonstration favored during Antiquity, lost much of its importance during this time. Instead, the difference between human knowledge relative to absolute divine knowledge became paramount. All human knowledge was taken to be necessarily incomplete and uncertain. Nonetheless, through the inferior human powers of reason/intellect and the senses, partial and imperfect access to divine knowledge through divine illumination was possible. Gradually, during the Middle Ages, the notion that knowledge derived from the senses and knowledge based on logic and reason could both aid in the apprehension of divine knowledge began to take hold. In addition, uncertainty (in light of humans' epistemological limitations), which had no place in the science of Antiquity, gradually came to be viewed as inevitable in the apprehension of knowledge in the Middle Ages, in light of the perceived necessary limitations of human knowledge.

Toward the end of the Middle Ages, the concept of a universe governed by necessary first principles was increasingly contested. In particular, Ockham's occasionalism highlighted the extent to which logically necessary first principles are incompatible with divine omnipotence. If knowledge in science is derived from the apprehension of divine knowledge, through will and reason, and God is radically omnipotent, such knowledge cannot be constrained by principles of logic or essential properties of things. An omnipotent God could create the world to be any way he wishes, and consequently, is not constrained by logical principles or reason. Even logic and reason were viewed as divine creations. In consequence, the individual instance

became a more legitimate object of scientific investigation, as the notion of knowledge of universal principles became untenable.

Evidence in the Renaissance: Hypotheses, Probability, and Laws of Nature

It is important to keep in mind the intellectual shifts laid out above. They contextualize the emergence of concepts and practices during the Renaissance that are more closely aligned with contemporary conceptions and practices of 'evidence.' In the present account, I eschew the notion of the Enlightenment as a time of scientific revolution that encompassed an entirely unprecedented change in conceptions of 'evidence' and constituted a radical break from previous practices. Such accounts of the history of science recently have been called into question by several historians and philosophers of science (e.g., Bowler & Morus, 2005; Shapin, 1996). Instead, I aim to trace the gradual development of concepts and ideas related to 'evidence.' I argue that widely held Renaissance accounts of the world as governed by divine will, and accessible only in an imperfect manner by humans through their inferior faculties of reason and the senses, provided the intellectual framework within which a number of the concepts that play a central role in contemporary conceptions of 'evidence' in science could emerge. These included the concept of internal evidence, evidence that consists of empirically observable "things" in nature, in contrast with the external evidence of testimony and authority. In line with the idea of internal evidence came practices associated with the concept of a hypothesis and the idea that empirical evidence may provide a means to adjudicate between competing hypotheses. In the metaphysical realm, conceptions of the world shifted with the emergence of notions of universal natural laws that govern natural phenomena. Relatedly, calculations of probability emerged as means to quantify uncertainty. It is to these developments that I now turn.

The Emergence of Empirical Evidence and the Study of Nature:
The Example of Paracelsus (1493-1541)

During the Renaissance, the study of nature through empirical means came to be viewed as a legitimate avenue to knowledge. Two preceding intellectual developments in the Middle Ages contributed to this shift. As previously stated, the supremacy of demonstrative, logically necessary inference was challenged by the absolute supremacy of divine knowledge over human knowledge. At the same time, authority as a source of knowledge came into question as the compatibility of necessary first principles with divine omnipotence became contested. In historical context, this was the time of the Protestant Reformation, which brought with it and epitomized a questioning of authority and doctrine leading to a democratization of knowledge (Trepp, in Daston & Stolleis, 2009). During this time period, the natural world came to be viewed as divinely ordered and governed, a reflection of divine knowledge. Nature, in addition to scripture, became a source of divine revelation, to be read by humans in an attempt to understand the divine (Findlen & Bianchi, 1990). The study of the natural world came to be viewed not only as a legitimate but also a sacred venture. These developments provided a fertile ground for the emergence of an increasingly empirical study of nature.

At the same time, the empiricism of the Renaissance mirrors contemporary science less than it diverges from it, appearing familiar, yet still very foreign to the contemporary reader. A good example is found in Paracelsus' medical studies, particularly in his observations of nature for the generation of knowledge. This example highlights both the empirical nature of many Renaissance knowledge practices and their more mythical underpinnings. Paracelsus viewed humans, nature, and the universe as analogical (Findlen & Bianchi, 1990). Similarities between elements of nature, he thought, were divinely ordained signs that yield cues to the underlying divine secrets of the universe. The task of the physician, according to Paracelsus, was to read the signs of nature in order to find cues to the cures of disease. His approach is epitomized in the following example:

> Behold the Satyrian root, is it not formed like the male privy parts? No one can deny this. Accordingly, magic discovered it and revealed that it can restore a man's virility and passion…Siegwurz root is wrapped in an

> envelope like armour, and this is a magic sign showing that like armour it gives protection against weapons. (Paracelsus, cited in Franklin, 2001)

To Paracelsus then, the study of nature quite literally involved the interpretation of physically observable resemblances, which he believed provided cues to the function of parts of nature. Despite the fact that his practice most certainly involved careful empirical attention to nature, as has often been pointed out (e.g., Franklin, 2001; Hacking, 1975), it does not resemble contemporary empiricism and, in fact, represents the very spiritual interpretive practice that most proponents of contemporary empiricism would be quick to reject. Several additional shifts in metaphysical and epistemological assumptions had to happen in order to provide a backdrop for the development of contemporary practices of evidence. Nonetheless, Paracelsus and his contemporaries most certainly legitimized the study of physical evidence in nature as evidence for the invisible workings of the universe (ordained by the divine), a practice that, without the involvement of the divine, resonates with certain current conceptions of the role of empirical evidence in science.

The Emergence of the Hypothesis:
Copernicus (1473-1543) and his Contemporaries

References to hypotheses date back to Aristotle's *Posterior Analytics*. Hypotheses continued to play a role in the science of the Middle Ages. Nonetheless, the definition of hypotheses and their role shifted profoundly during the Renaissance (Grant & Nelson, 1962). From Antiquity to the Middle Ages, hypotheses were assumed propositions not directly proven by the person who was assuming them (Grant & Nelson, 1962). Hypotheses were taken to be fictional assumptions invented to account for observed phenomena, with the goal of "saving the phenomena" [13] and so their truth or falsity was a non-issue (Grant & Nelson, 1962, p. 601). The two purposes of hypotheses were to explain observed phenomena and to fit or predict future observations. There

[13] This conception arose from astronomy in Antiquity where geometrical hypotheses of epicycles and eccentric circles were used to account for celestial phenomena. These hypotheses were used to account heuristically for such phenomena but their actual truth was considered irrelevant.

was no clear relationship between hypotheses used to account for observable phenomena and the metaphysical principles at the core of the philosophical demonstrations that were thought to constitute the certain knowledge of the scientia of Antiquity and the Middle Ages.

With the advent of Ockham's occasionalism at the end of the Middle Ages and the decreased dependency on authority as a source of knowledge associated with the Protestant Reformation, the beginning of the Renaissance can be characterized as a time of liberated scientific imagination and speculation (Grant & Nelson, 1962). No longer constrained by the literature and metaphysical writings of Antiquity, and resigned to epistemic uncertainty relative to divine knowledge, theories could be freely advanced with little regard for their veracity. Religious interpretations abounded (Harrison, 2006), both within the traditional Catholic and the new Protestant movement. In this context, various scholars (e.g., Daston, 1991; Grant & Nelson, 1962; Harrison, 2006) have described the need for a means to adjudicate between rapidly emerging doctrines and theories, which created a context for a new understanding of the hypothesis and its role in the pursuit of knowledge. The first shift in conceptualizing hypotheses, Harrison argues, was a shift towards understanding hypotheses as true propositions about the universe. One of the earliest proponents of such an understanding of hypotheses was Copernicus, who dismissed the idea that hypotheses are merely convenient conjectures used to "save the phenomena" of interest. Instead, he proposed that from a true hypothesis, a scientist could deduce further truths, which can then be verified. He argued, "if the hypotheses assumed by them were not false, everything which follows from their hypotheses would be verified beyond any doubt" (Copernicus, 1959, p. 28). Copernicus thus introduced a shift from previous understandings of hypotheses as fictional accounts that fit observed phenomena but whose truth or falsity was irrelevant, to an understanding of hypotheses as true propositions. Copernicus also provided a definition of a true hypothesis that implied a process by which its truth can be verified. A true hypothesis, according to Copernicus, was one from which true principles could be deduced. By implication, the truth of a hypothesis could be verified by verifying the truth of the principles deduced from it. In this way, the assumed direction of the relationship between observations and phenomena in Antiquity and the Middle Ages was reversed during the Renaissance. Hypotheses had previous merely served as convenient ways

to account for phenomena. Beginning in the Renaissance, however, phenomena were thought to provide a means to adjudicate the veracity of a hypothesis. In the Renaissance, phenomena thus became primary evidence in the adjudication between competing hypotheses.

The Emergence of Natural Laws: Francis Bacon (1561-1626)

Along with epistemological shifts to the empirical study of nature and the belief in hypotheses that truly represent nature, the Renaissance also witnessed an ontological shift in how nature was thought to operate. In Aristotelian and subsequent Scholastic philosophy, nature itself was granted as having a variety of efficient causal powers that governed its operations. It was thought that elements of nature moved toward their final cause based on their essences. With the advent of both the doctrine of occasionalism and a voluntaristic conception of God, however, such an account was no longer plausible. Instead, God himself came to be viewed as the direct volitional, causal operator of nature. Yet God's will did not operate arbitrarily on an ad-hoc basis. Rather, it was thought that God operates the universe according to uniform principles (Daston & Stolleis, 2009). The notion of imperfect human knowledge provided a further means to explain irregularities in the operations of nature: such irregularities were attributed to imperfections in human knowledge, rather than irregularities in the manner in which God governed the universe (Harrison, 2006). In this context, the idea of laws of nature emerged.

A particularly clear and very early account of nature as governed by natural laws can be found in Francis Bacon's (1620/2005) Novum Organon, which sometimes is considered to be one of the foundational works in the establishment of the scientific method (e.g., Blunt, 1904; Muntersbjorn, 2003). Bacon, who was trained as a lawyer, thought that human law was based on divinely ordained natural law and that similar laws also underlie the behaviour of all elements of the natural world. He viewed the discovery of these laws as the ultimate goal of science. Bacon rejected the value of investigating final and formal causes; the former, he thought, led scientists astray in search of purposes and goals; while the latter had proven to be futile as goals of scientific inquiry. Bacon thus seemed to reject the first principles and essences of Antiquity and the Middle

Ages. He saw greater promise for science in the investigation of material and efficient causes, but disparaged the superficiality with which such causes had been studied, arguing that true efficient and material causes often are hidden and not immediately apparent. These hidden efficient and material causes, he conceptualized as invisible laws of nature, as illustrated by the following passage:

> Earlier in this work, I noted as an error of the human mind the opinion that to understand what exists you have to look at forms. It's true that nature really contains only individual bodies, performing individual pure actions according to a fixed law; but in science this law is what we inquire into, discover, and explain; it is at the root of our theorizing as well as of our practical applications. (Bacon, 1620/2005, trans. Bennet, 2:26)

Bacon was concerned with various human biases (idols) that he thought had obscured the discovery of laws of nature. He proposed a new method of inquiry that he thought would minimize the influence of such biases, thus setting the stage for the scientific method that was to emerge in the following century.

"The Emergence of Probability":[14] Pascal (1623-1662) and Leibniz (1646-1716)

A final important development towards the end of the Renaissance was the emergence of a mathematical conception of probability. This topic has been discussed extensively by several scholars (e.g., Franklin, 2001; Hacking, 1975). Both Hacking and Franklin argue that conceptions of probability have existed in various forms in all civilizations. However, only at the end of the 17th century were attempts made to quantify uncertainty using the tools of mathematics. Franklin (2001) distinguishes between three conceptualizations of 'probability.' The first is probability as an unconscious response to uncertainty at a cortical level, which he situates outside the realm of historical investigation. The second is probability in an ordinary language sense, as in the legal concept of proof beyond a reasonable doubt, which figures in the discourse of most civilizations. Thirdly, he identifies probability as a formal mathematical concept. This last conception emerged at the end of the 17th century. Franklin argues that there were mathematical expressions to quantify factual probabilities in the contexts

[14] Title borrowed from Hacking (1975).

of insurance and other aleatory contracts prior to this time. What emerged at the end of the 17th century was a mathematical expression of chance. It is generally agreed that the emergence of mathematical probability occurred when mathematical methods were applied to the quantification of outcomes in games of chance, particularly dice throwing (Franklin, 2001; Hacking 1975). Many people have been credited in this development, including de Mere, Roberval, Caramuel, Huygens, Fermat, and Pascal (Franklin, 2001), and there is no way to credit any one of them over the others. However, in a famous correspondence between Pascal and Fermat, where the two discuss ways to divide the stakes in a gambling game that had to be interrupted prematurely, the most thorough mathematical accounts of probability calculations are articulated and what has been referred to as a "geometry of chance" emerged (Franklin, 2001). The application of mathematics to the realm of knowledge occurred earlier, in the geometric accounts of divine illumination provided by the Scholastics. However, with the emergence of mathematical calculations of chance, uncertainty became firmly situated in the realm of reasonable scientific inquiry.

Certainty in Uncertainty: Empiricism, Hypotheses, Natural Laws, and Probability

Towards the end of the Renaissance, then, several conceptions had emerged that provided a new framework for 'evidence,' its objects, and the relationships between evidence and its objects. All three of these conceptions underwent significant change during the Renaissance. Hacking (1975) argues that probability is the relationship between evidence and a hypothesis. In the context of the present discussion, it can be argued that by the end of the Renaissance, empirical data had become legitimate evidence in scientific investigation, and the establishment of natural laws or regularities had become the main objects of scientific investigation. In this context, it was possible to conceptualize hypotheses as means of understanding and identifying laws of nature. The goal of a hypothesis was to provide a true account of nature and its laws. Empirical data provided a means of adjudicating between competing hypotheses and of examining their truth. Finally, variations and uncertainty in empirical data came to be explained and quantified in rational mathematics with the emergence of probability; previously such uncertainty and variation had been explained through the fallibility of human knowledge relative to divine omniscience. The stage now was set for an account of empirical

evidence, the objects of which are hypotheses that seek to represent the true laws of nature, and the expression of which is quantifiable uncertainty in the form of probability.

The Enlightenment:[15]
Experiments, Inductive Science, and a Mechanistic Universe

Above, I argue that by the end of the Renaissance the stage was set for an account of 'evidence' in science that is quite similar to contemporary accounts. However, the concept of 'evidence' itself had not yet made its appearance on the scientific stage. A concept of 'evidence' still did not play a major role in the science of the Renaissance. In fact, one is hard pressed to find many explicit references to the concept of 'evidence' in Renaissance science and philosophy of science. The Enlightenment was to change all of this. The time period from the mid-17th to the early 19th Century brought further intellectual developments that contributed significantly to the ascendance of a notion of 'evidence' to a prominent role in practices of science.

Evidence in the Enlightenment

During the Enlightenment, the manner in which scientific data or evidence was generated became central to decisions about what constitutes evidence. Of course, how data for inferences were arrived at or known was epistemologically significant prior to this time. For example, the evidence of Antiquity was arrived at through logical demonstration, the evidence of the Middle Ages was gained from authority, and in the Renaissance evidence was apprehended through the senses as enabled by divine authority. What was novel in the Enlightenment, however, was that the process of apprehension or generation itself began to be viewed as central in giving evidence its legitimacy. In Antiquity, it was not so much the demonstrative process, but rather the fact that demonstrations occurred from first principles that made them legitimate. Similarly, in the Middle Ages it was not how authoritative testimony was known but the

[15] I use the term Enlightenment solely for reasons of historical convenience and convention, not to portray this time as an emergence from intellectual darkness.

fact that it was from authority that mattered. Again, in the Renaissance, although evidence of nature was increasingly thought of as apprehensible through the senses, what made such apprehension credible was that it was provided by divine authority or was divinely ordered. In the Enlightenment, however, increasing skepticism about what made evidence credible (e.g., derivation from first principle, authority, or divine order) became widespread and processes through which data were collected, treated, and adjudicated gained great importance. This shift is evident in multiple trends and events in science and philosophy of science during this time period, including the debates between empiricist and rationalist epistemologies, the increasing importance of experiments, and the emergence of science as a public venture.

Empiricism vs. Rationalism: Bacon (1561-1626), Descartes (1596-1650), and Locke (1632-1704). In the Middle Ages and into the Renaissance, the notion that the senses can provide legitimate data for science, in addition to rational inference, emerged. Indeed, during this time period, most thinkers argued for the legitimacy of both sense experience and reason. In the seventeenth century, however, the relative utility of these two sources of evidence became a matter of heated debate among philosophers. Empiricism and rationalism became two competing and opposed epistemologies.

Rationalists such as Descartes and Spinoza, inspired by skepticism about certain knowledge, advanced reason as a sole source of such knowledge. Of course, the rationalist epistemology of Aristotle was dominant until well into the Renaissance. However, the rationalism of Enlightenment thinkers such as Descartes was based not on certain, assumed first principles, but began instead by radically doubting all knowledge. In his *Meditations on First Philosophy,* Descartes questioned all knowledge, in order to establish only those things that can be known with certainty. For him, the route to such inquiry lay in reason rather than perception.

> Since I now know that even bodies are not, properly speaking, perceived by the senses or by the faculty of imagination, but by the intellect alone, and that they are not perceived through their being touched or seen, but only through their being understood, I manifestly know that nothing can be perceived more easily and more evidently than my own mind. (Descartes, 1641/1998, Meditation Two)

Descartes posited that the mark of a true idea is that it is clear and distinct to the mind. Through his method of rational doubt, he arrived at a few such clear and distinct ideas, ideas that he thought were indubitable, including his own existence (in his famous cogito) and the existence of God. From these clear and distinct ideas, he then went on to deduce a system of philosophy. Of note here is the fact that Descartes used reason to provide evidence for the existence of God, when previously during the Renaissance God's divine illumination or testimony was used as evidence for the reasonability of a claim.

The empiricist strand, on the other hand, had its origin in the Renaissance, particularly in the work of Francis Bacon who advocated experience as a source of data. In the Enlightenment, however, the term empiricism was formally coined and elaborated by John Locke. Locke's (1690/1998) account of knowledge was laid out in his *Essay Concerning Human Understanding*. Locke famously rejected the notion of innate ideas, argued that the mind at birth is a 'tabula rasa,' and claimed that all knowledge is based on sensations (direct sensory perception) or reflection (which he conceptualized as internal perception of the contents of the mind). Although Locke allowed for the possibility that a small number of ideas can be grasped by reason or intuition (the existence of self and the existence of God), he was committed to the view that knowledge of the external or natural world is based on empirical data. He wrote:

> The knowledge of the existence of any other thing we can have only by sensation. For there being no necessary connection of real existence with any idea, a man has in his memory, nor of any other existence but that of God with the existence of any particular man, no particular man can know the existence of any other being, but only when, by actually operating upon him, it makes itself perceived by him. For the having the idea of anything in our mind no more proves the existence of that thing than the picture of a man evidences his being in the world, or the visions of a dream make by this means a true history. (Locke, 1690/1998, Book IV, Chapter XI)

Locke and other Enlightenment empiricists, driven by the skepticism of the time, rejected the certainty of reason and its value as a means to obtain knowledge of the external world. They privileged empirical data as the sole source of knowledge. Evidence in science, for the empiricists, had to be empirically apprehended.

Rationalist and empiricist approaches in the Enlightenment were diametrically opposed. Locke wrote in response to Descartes, which is apparent in many parallels between their works. Descartes reified clear and distinct ideas, which he took to be innate. He posited that all perception could ultimately be reduced to reflection and that reflection was the true source of knowledge. In contrast, Locke rejected innate ideas, argued that humans are born with minds that are like "blank slates," and that all reflection could be understood as perception of the mind's content. During the Enlightenment, empiricist epistemology was not yet dominant, and an active debate between rationalists and empiricists raged about reflection versus perception as legitimate and primary sources of evidence. For both rationalists and empiricists however, what ultimately determined whether something could serve as evidence in science depended on how one arrived at it: through reflection or through perception.

Measurement and mathematics: The example of Isaac Newton (1642-1727). The ascent of mathematics to a position of importance in the realm of science began well before the Enlightenment, as is evident in the account of the Middle Ages and Renaissance provided earlier. However, towards the end of the Renaissance and at the beginning of the Enlightenment, major shifts in the understanding of the role of mathematics in science occurred; namely, it has been argued (e.g., Henry, 2001) that during this time, a realist understanding replaced an instrumentalist understanding of the role of mathematics in science (Henry, 2001). Prior to the Enlightenment, mathematics had been viewed as a convenient, heuristically useful tool that can be employed to describe and predict phenomena but the calculi (expressed in formulae) were ultimately considered to be fictitious. In contrast, beginning in the Enlightenment, mathematics began to be thought of as a way to provide epistemological access to the real world -- mathematical calculi were thought to represent real-world relations between phenomena. This shift parallels the shift in understandings of hypotheses during the Renaissance, as described previously. This new understanding of mathematics is partially evident in the earlier work of Copernicus, who thought of calculi as accurate real-world representations of planetary motion. However, according to Henry (2001), the primacy of mathematics in the acquisition of scientific knowledge becomes fully evident in Isaac Newton's (1678) *Mathematical Principles of Natural Philosophy* – in his account of gravity and various other natural phenomena, Newton assumed, instead of justified,

the mathematical approach. With Newton, mathematics came to be understood as essential to natural philosophy, where previously it was merely a possible tool to be employed in natural philosophy.

A second shift during this time period, which was closely related to the ascendance of mathematics, occurred with respect to the manner in which empirical data were conceptualized, particularly with the rise of measurement and instruments of measurement in science. The use of sense data in the pursuit of knowledge was legitimized during earlier times. However, the sense data of the Renaissance differed from those of the Enlightenment in a major way (Butterfield, 1957; Dear, 2008; Henry, 2001). The sense data of earlier times were those of ordinary common experience, accessible to all. Observation also was considered to be central to some sciences (e.g., astronomy and meteorology) but did not play a major role in all scientific practices (Daston & Lunbeck, 2011). Towards the end of the Renaissance and the beginning of the Enlightenment, however, mathematics began to be applied to the natural world in a different manner. Observations were quantified in increasingly detailed ways and newly devised technical instruments were used to obtain what were considered to be more accurately quantified observations of the world in all areas of science. Daston and Lunbeck (2011) argue that during this time period observation also came to be reconceptualised as a skilled practice undertaken by multiple observers who have received special training in appropriate methods of observation.

Shifts towards greater use of measurement and instruments of measurement were also related to changes in the manner in which matter was understood and conceptualized (Henry, 2001). From the time of Aristotle, matter had been thought to be alive and imbued with various qualities, which were of interest in relation to a search for causes. During the late Renaissance and Enlightenment, however, matter began to be viewed as homogenous, uniform, and differing only in terms of extension and motion (Reill, 2003). This change in the conceptualization of matter made quantities a major focus of science. Quantification, measurement, and mathematics thus came to play a central role in scientific observation (Henry, 2001). Mathematical forms of representation came to dominate the paradigm system of representation within physical science. Instruments of quantification proliferated (e.g., thermometers, barometers, the telescope, the microscope). Of course, there were mathematics and instruments prior to

this time. However, it was during the Enlightenment that mathematics and instruments of natural philosophy became widely accepted as means of uncovering truths about the natural world that cannot be ascertained without such methods and instruments.

With the emergence of instrumentation and measurement, a major epistemological shift occurred. Prior to this time, the self-evident knowledge of experience reigned supreme and was taken to be the indubitable gold standard against which all claims had to be tested. In the Enlightenment, however, the authority of ordinary experience was replaced by the authority of trained observation, quantification, and measurement through instruments, which at times contradicted ordinary experience (e.g., the claim that disease is transmitted through micro-organisms rather than the much more easily observed and previously held theory that it stems from bad air emitted by rotting matter and identifiable through the sense of smell). Reill (2003) provides a useful interpretation:

> Science was directed towards establishing a comprehensive system of measure and order, a universal mathesis. Mathematics became the privileged language of natural philosophy; more than that, it was assumed to be its ideal form of exposition. In the hierarchy of knowledge, the place occupied by any specific form of knowledge was established by the degree to which its subject matter was capable of being treated in a manner guided by mathematical principle. (p.27)

This methodological and epistemological shift contributed to the increasing acceptance of mathematics as a means of arbitrating truth or knowledge. It also contributed to the increasing importance of experiments as a source of scientific knowledge.

Experiments and public demonstrations in science: The Royal Society. There were documented experiments in science in Middle Eastern societies dating back to ancient times. During the 17th and 18th century, however, experiments emerged in a new way as the primary means of establishing truth in science. Dear (2008) argues that this time period witnessed a shift in the manner in which experiments were conceptualized -- from naturalistic sense experience to the systematic observation of manipulations of conditions in an experiment. Aristotelian science, Dear points out, was primarily concerned with final causes, the purposes or ends towards which elements of nature were taken to move naturally. Experiments in which nature is manipulated thus

had no place in Aristotelian science. Any such manipulations would obscure the natural causes or trajectories of nature. In addition, Serjeantson (2008) argues, individual instances were not particularly relevant to Aristotelian science, which was concerned with tendencies or final causes towards which elements of the physical world move invariably by virtue of their nature. By the end of the Renaissance, experiments began to be defined as interventions into the course of nature that included trained observation. Prior to this time, any observation of nature was considered to be an experiment (Daston & Lunbeck, 2011)

The shift in conceptions of legitimacy of experiments is clearly evident in the famous 17th century work of Harvey, which established the circulation of blood by means of dissection of a human body (Dear, 2008). Following his experiments, Harvey had to defend his method to the scientific community. He argued that his dissecting left unaffected those aspects that were under direct observation in his studies (the blood flow), despite his modification of adjacent elements (e.g., veins). From Harvey's defense of his methods, it is obvious that the method of experimental manipulation, at least in certain contexts, was still highly controversial at the beginning of the 17th century. Shapin and Schaffer (1985) document the active debate between proponents of a worldview based on natural laws, which was inconsistent with experimentation, and proponents of a more mechanical worldview, who viewed experiments as a way to manufacture knowledge or fact during the 17th century. They cite the work of Thomas Hobbes as an example of the former view and describe the opposing perspective as advanced by Robert Boyle, who considered experiments to be a way to resolve social conflict over facts. Boyle argued that experiments, which can be repeated and observed by the public, allow the public to witness the production of facts in a manner that eliminates disagreement about them. By the end of the 18th century, experimental manipulations were commonly cited as privileged sources of knowledge (Dear, 2008) and had become well established.

In comparison, the experiments of earlier times, including those conducted by Bacon, were intended to establish not what happened in a particular instance (during an experiment) but rather "how things behave" (Dear, 2008, p.110) in general. Even the famous tower of Pisa experiment of the Renaissance, conducted by Galileo, historians now agree, may never have been conducted outside of Galileo's thoughts. Accounts of

Renaissance experiments, Dear argues, were generally vague and focused on general principles that were taken to have been established, rather than on details concerning who conducted an experiment, with what methods, and under what circumstances. The principles established by scientists (e.g., that identically shaped objects of different mass fall at the same rate of acceleration) were intuitive and easily commanded assent from the public that they needed to be invariably the case.

The emergence of mathematics as part of natural philosophy brought with it the ability to make contestable scientific claims that ran counter to commonly accepted principles with the support of mathematical and experimental proof. The theories and laws proposed often could not be easily observed by all. Subsequently, self-evidence as a criterion of truth was replaced by credible and reliable reports (Serjeantson, 2008). Because scientific principles were no longer self-evident, scientists, in order to gain credibility, had to find ways to make these newfound theories, that were experimentally verified under certain specifiable conditions, empirically demonstrable to the public (Dear, 2008). In this way, science became a public venture.

As scientific principles were no longer taken to be self-evident in nature, scientists had to find means to make them accessible to the educated public. One such means was the narrative account of an experiment, which occurred in the context of the proliferation of scientific journals and monographs published at this time (Johns, 2004). Although members of the public could not witness newly confirmed scientific principles in their daily lives, they could experience them through the narrative accounts in these publications. A second means was provided by the public demonstrations and experiments that occurred in the proceedings of scientific societies (Serjeantson, 2008). Here, experiments became directly observable and accessible to non-scientists, taken out of the laboratory and enacted in front of an audience. The emergence of scientific societies such as the Royal Society in England and the Academie Royale in France provided an institutional framework for the popularization of science.

‘Evidence’ of the Enlightenment: A Summary

Based on this very brief and cursory account, what constituted ‘evidence’ in the Enlightenment, or the criteria by which one could determine whether something can

serve as 'evidence' in science, shifted during this time period. At the outset of the Enlightenment, rationalists and empiricists still debated the supremacy of reason versus empirical observation. Increasingly, however, knowledge based on that which is empirically observable, not necessarily through naturalistic sense observation but, rather, in terms of mathematical quantification/measurement by instrument became the most valued scientific evidence. Such evidence was generated through experimental examination and manipulation, and, ideally, through public demonstration in narrative accounts or public experimental demonstrations. Although 'evidence' during this time period became more abstruse and esoteric in particular ways (quantification and measurement), it also became increasingly important for such 'evidence' to be observable and witnessed by the wider public

The object of 'Evidence' in a mechanical world: The examples of Harvey (1578-1657) and Descartes (1596-1650). Along with epistemological shifts in how evidence was to be generated, the 17th and 18th centuries also witnessed an ontological shift with respect to the nature of the natural world. During the Renaissance, the world was thought to be governed by laws and regularities. In contrast, during the Enlightenment, the world was thought to be governed by mechanical principles. Some (e.g., Butterfield, 1957) argue that this development was inspired by the invention of a wide variety of machines and instruments. The world was thought to operate according to the same principles as machines, and causation came to be viewed in terms of physical contact or force. Relatedly, the natural world was increasingly taken to consist of matter, which can be reduced to small, indivisible particles (in early accounts, Corpuscles). The mechanical worldview became predominant in society by the end of the 17th century. According to the mechanical worldview, mechanical principles (governing, for example, shape, size, quantity, motion) were the only legitimate principles to be used to explain phenomena in the natural world (Henry, 2001).

Mechanical principles were applied to both inanimate and animate parts of the natural world, including animals and humans. Although dualistic accounts tended to posit non-material aspects of human nature (e.g., a soul), the functioning of the human body was increasingly explained through mechanical principles. This shift is apparent when one contrasts the respective works of Harvey and Descartes on blood circulation and the heart (Henry, 2001). Harvey had famously demonstrated that blood circulates

through contraction of the heart. The contraction of the heart itself however Harvey attributed to "pulsific faculties" of blood, consistent with an Aristotelian model. In contrast, Descartes drew on Harvey's principles but provided a purely mechanical account, proposing that a source of heat in the ventricle of the heart vaporizes blood as it enters the heart, causing the heart to expand and rapidly push out the vaporized blood as it then contracts when blood cooled by the lungs enters again. Many of the mechanical accounts of the natural world proposed during this time will seem as foreign to the contemporary reader as Descartes' account of blood circulation. The general principles of a mechanical worldview, however, remain influential in science.

It was during the Enlightenment that mathematical principles of mechanical physics first began to be thought to underlie the manner in which the universe operates. The goal of science was now to explore and discover these principles. As such, the evidence of science needed to be couched in terms of physical and mathematical principles. It is thus not surprising that 'evidence' increasingly came to be conceptualized as mathematically quantified observation, which provided a means to access the often unobservable mechanical principles according to which the world operates. The 'evidence' of the Enlightenment was thus quantifiable, empirical, and publically observable. Its objects were the mechanical principles underlying the operations of the natural world. The relationship between 'evidence' and its object was thought to be mathematical in nature.

Fact and Induction:
The Concept of 'Evidence' Emerges

Aristotelian science, the first subject of this chapter, clearly distinguished between logic and rhetoric, placing science firmly in the former realm. Serjeantson (2008) rather convincingly argues that in the seventeenth century this division was bridged as concepts such as 'fact,' which had formerly played a role only in rhetoric and law, became an increasingly important part of scientific practice. This connection also becomes apparent in Newton's works (see Harper, 2011), where one comes across explicit references to 'evidence' that is inductively connected to conclusions, and which

seems to play much the same role in Newton's science as the evidence of law plays in a trial. It is clear from the previous discussion of experiments and public demonstrations that the scientists of the 17th and 18th century were concerned with establishing themselves as legitimate purveyors of knowledge, equipped with powerful tools they could utilize to establish their legitimacy.

According to Serjeantson (2008), prior to the 17th century, the concept of fact played a role only in legal contexts, which dealt with the establishment of volitional human behaviour in a particular instance. Facts, understood as temporal-spatial individual instances, were irrelevant to science, as they precluded logically necessary demonstration. As science shifted away from demonstration and towards experiment, however, it was the establishment of temporal-spatial fact that became relevant for the credibility of an account of an experiment. Whether or not the scientist was telling the truth about an experiment had to be established in much the same way as whether or not the eyewitness in a trial was telling the truth. Similarly, the only tool the scientist had in order to establish the truth of his account was that of supporting facts, or evidence.

Although the concept of 'evidence' itself was used only rarely among the scientists of the Enlightenment, the stage was set for its emergence. In the works of Isaac Newton, one discovers references to 'evidence' that mirror the manner in which this concept is at times utilized at present. Faced with the challenge of establishing conclusions from experiments and observations, in his *Opticks* (1730), Newton argues that induction from observation and experiment provides the soundest means of scientific inference. In his *Principia*, Newton states his fourth principle as follows:

> In experimental philosophy, we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate or liable to exceptions. (Newton, 1687, Rule IV)

In Newton's science, as an increasing number of experiments confirm a proposition certainty about that proposition increases. Again, in his *Opticks*, Newton refers directly to 'evidence': "with the help of philosophical geometers and geometrical philosophers, instead of the conjectures and probabilities that are being blazoned about everywhere,

we shall finally achieve a natural science supported by the greatest evidence" (cited in Dear, 2008, p. 128). To Newton, evidence supports a science and the best science is that which is supported by the greatest evidence. The extent to which science is a matter of support by evidence, a concept adapted from legal frameworks, is clearly apparent here.

Evidence in the Enlightenment: A Summary

During the Enlightenment, the world increasingly came to be viewed as mechanical. The mechanical principles according to which the world operates became the object of evidence. Substantial shifts occurred in how these principles were thought to be known, and in how what constitutes 'evidence' in inference was delineated. Increasingly, whether something constitutes evidence was determined by how it was generated. The evidence of the time was generated through quantification and measurement in controlled and publically observable experiments. Evidence was empirical, quantitative, measurable, publically observable, and generated through experimental procedures. Within this new experimental context, the concept of 'evidence' emerged along with the notion that such 'evidence' inductively and rhetorically supports scientific achievements. Concepts such as fact and 'evidence,' which had previously played a role only in rhetoric and courts of laws, became firmly established in the practices of science.

Evidence in Modern Science

By the end of the Enlightenment, practices related to establishing scientific evidence had emerged along with the concept of 'evidence.' The concept of 'evidence' gained increasing importance in the science of the following centuries. In fact, major debates in the philosophy of science during the 19th and 20th centuries to the present day have centered on conceptions of 'evidence' and its relationship to facts and scientific knowledge. Competing and contradictory accounts have been advanced by positivists, logical positivists, empirical realists, and post-modernists, and debates surrounding these competing accounts remain alive and well in current philosophy of science. These

accounts and debates will be outlined in the end of the second part of this work, in Chapter 7, which deals with contemporary conceptions of 'evidence' in philosophy of science. In that chapter, the implications of these varying conceptions of evidence in modern philosophy of science for psychology and particularly evidence-based practice in psychology will be examined. To facilitate such an examination, it is necessary to turn to a discussion of the role of evidence in the history of the discipline of psychology. The following chapter (Chapter 4) provides a very brief overview of the role of evidence in the development of psychology as a discipline. The following chapters (Chapters 5-6) examine more closely the role of evidence in the evidence-based practice movement in modern psychology, which is the main focus of the present work. I return to a discussion of modern philosophy of science in Chapter 7.

Evidence in Law and Science: A Few Conclusions

The very cursory review of the history of evidence in law and science contained herein is neither exhaustive nor definitive. My goal, rather than to provide an exhaustive account of all practices related to evidence in these contexts, has been to highlight the plurality of roles that evidence has served across historical time in science and law. In addition, the previous two chapters have highlighted the extent to which the conceptualization and utilization of evidence have been contingent on a variety of social and contextual factors throughout history. It is impossible to understand the practices surrounding a particular employment of evidence in a particular social context without understanding the larger social context in which the employment occurs.

In the following chapters, my discussion of evidence in psychology and especially in the EBP discourse describes a much more narrow conceptualization of evidence in these more contemporary contexts. It is important to recognize however, that evidence, as defined in common parlance, is a much broader concept than the technical definitions of evidence that are employed in modern science, conceptions which will be described in the following chapters. For now, it suffices to keep in mind the many and varied roles that evidence has played since Antiquity. Although, it may be difficult for may contemporaries to conceive of a science that does not employ the sorts of evidence commonly utilized in modern science or to imagine legal practices that do not rely on

physical evidence and testimony with respect to facts, the previous chapters have highlighted that for most of human history, extant scientific and legal practices served important social and epistemological functions that, despite some convergences often were very different from contemporary scientific and legal practices. It thus will be important to keep in mind the diverse roles and conceptualizations of evidence described in the previous chapters, when considering the more narrow contexts of employment and conceptualizations of evidence in modern psychology.

Chapter 4:
An Introduction to the
History of Evidence in Psychology

In his book, *Inventing Our Selves: Psychology, Power, and Personhood*, Rose (1998) argues that the subject matter and content of psychology cannot be separated from the methods and technologies used to arrive at psychological truths. He argues:

> In the construction of psychological truths, the technical means available for the materialization of theory have played a determining and not a subordinate role. The technical and instrumental forms that psychology has adopted for the demonstration and justification of theoretical propositions have come to delimit and shape the space of psychological thought itself. (Rose, 1998, p. 57)

He identifies two techniques that psychology adopted from other domains of knowledge as part of its establishment as a legitimate discipline: statistics and experiments. His account meshes well with current conceptions of the methods of psychology and seems to suggest that experiments are the primary method of evidence-generation in psychology. In the present chapter, however, I argue that evidence generated through experiments has played an important, though not exclusive role in psychological research. I aim to show that the role of experiments in the development of psychology as a discipline and the emergence of clinical psychology as a sub-discipline may be much smaller than commonly thought. I support this view by tracing developments in the evidence of psychology related to the emergence, growth, and revision of psychological treatments and interventions over historical time. Experiments, I argue, did not play a very central role in the emergence of psychology as its own discipline and in the early days of psychological research. Experiments also played a very minor role in the early days of clinical practice and the development of clinical psychology as a sub-discipline. The role of experiments as a source of evidence in psychological practice

only became prominent in the middle of the 20th century, after the Second World War, when the scientist-practitioner model emerged. This overview, although brief and broad in nature, will highlight the manner in which not only evidence but also the questions to be addressed through evidence have changed substantially over the history of psychological practice. Experiments have not always been the main source of evidence for psychological practice.

Psychotherapy's Pre-history: The Birth of Psychology as a Scientific Discipline

In many 'History of Psychology' textbooks (e.g., Brysbaert & Rastle, 2009; King, Viney, & Woody 2009), the birth of psychology is told in some variant of the following story. Questions about psychology (e.g., selfhood, mind, etc.) had long been addressed by philosophers and theologians. However, in the latter part of the 19th century the methods of experimental science began to be applied to these questions and the first psychological laboratories were established. The birth of psychology can be dated to the establishment of the first psychological laboratory by Wilhelm Wundt in Leipzig in 1879. This version of the early history of psychology, along with the implied view that psychology began with the establishment of experimental psychological laboratories, is consistent with Rose's view that experiments are a necessary, essential, and defining feature of psychological inquiry. Contrary to these accounts, however, psychology as a sub-discipline of philosophy, rather than an independent discipline, was well established by the late 18th century, long before Wundt's time. In addition, it is also interesting to note that Wundt was not a proponent of the sole use of experiments, and that much of post-Wundtian psychological inquiry did not rely on experimental methods of evidence generation in a contemporary sense. Rather, the manner in which 'experiments' were defined in the early days of psychology differed significantly from contemporary definitions of experiments. It would therefore be misguided to assume that psychology has always or primarily relied on experiments as currently understood and practiced.

Psychology, as a sub-discipline of philosophy, became part of the mandatory curriculum in teacher education in Germany by the early 19th century (Brysbaert &

Rastle, 2009). The psychology of this time was not based on mere mental contemplation, but utilized various methods of inquiry. In fact, the first discussions of types of evidence used in psychology date back to 1730, when Wolff, a German philosopher, proposed a distinction between rational and empirical psychology, where the former was based on rational deductive inquiry while the latter was based on individual experience (as cited in Brysbaert & Rastle, 2009). Thus, both individual experience and logical deduction were forms of evidence associated with psychology over one hundred years before Wundt's laboratory in Leipzig. By the time of the establishment of Wundt's laboratory, tensions between traditional philosophers and those who were pursuing empirical studies in philosophical psychology had reached a high point (Lamiell, 2012). Philosophers were concerned with protecting their traditional areas of scholarship and viewed psychology as a lower form of inquiry, a "philistine art" (Wundt, 1913, p. 9). They argued that psychologists were not qualified to teach courses in traditional areas of philosophy. Psychologists on the other hand viewed many traditional areas of philosophy as a bothersome distraction from the work that was of direct interest to them (Lamiell, 2012).

In his 1913 essay *Psychology's Struggle for Existence,*[16] Wundt takes the following stance on this issue:

> In the opinion of some, philosophy and psychology should divorce from each other. Now, it is well known that when a married couple seeks a divorce, both members usually are at fault. In these pages it will be shown that the same is true in this instance, and that if this matter takes the course that both parties want, philosophy will lose more than it will gain, but psychology will be damaged the most. Hence, the argument over the question of whether or not psychology is or is not a philosophical science is, for psychology, a struggle for its very existence. (Wundt, 1913, foreword)

Surprisingly then, Wundt, who is often credited as the founding father of the scientific discipline of psychology, was concerned about a separation of psychology from philosophy and thought that philosophical inquiry informs sound psychological investigation in a central and essential manner. In fact, Araujo (2012) argues that Wundt

[16] My awareness of this essay stems from the work of James Lamiell (2012), to whom I am indebted for the content of this section.

positioned his psychological investigations and theories in the context of a broader philosophical project, concerned with issues of metaphysics, epistemology, and logic. Wundt's psychological work was directly influenced by his philosophical commitments and can in fact only be understood in this context. Danziger (1979) also highlights the impact that Wundt's opposition to positivism as a philosophical framework had on his work. It therefore seems that Wundt may have been as committed to philosophy as he was to psychology.

In addition, Wundt viewed experiments as only one source of evidence, suited to the investigation of only a small part of potential areas of psychological interest. Wundt posited that the psychology that philosophers sought a divorce from was merely a small branch of psychology, namely experimental psychology or what Wolff would have termed empirical psychology. Wundt, despite identifying as an experimental psychologist, was convinced that such an approach was suited to rudimentary processes such as perception and consciousness but that many areas of psychological inquiry, including child psychology, animal psychology, social psychology, and cultural psychology were only partially or ill-suited to the experimental method and instead had to rely on other sources of evidence (Lamiell, 2012). Thus, Wundt was not a pioneering champion of a solely experimental, scientific approach to the study of psychology, despite the fact that he is considered to be the founding father of this approach. Rather, for Wundt, legitimate evidence in psychology was much broader and varied than that garnered from experiments, and was determined by that for which it was taken to be evidence (its object or its subject matter). It would be erroneous to characterize Wundt as a proponent of an exclusively experimental approach to psychological investigation.

In addition, as mentioned earlier, it is important to note that conceptualizations of experiments have shifted significantly over the course of psychology's history, and that what was considered to be an experiment in the early days of psychological research may not presently be considered to be an experiment. Danziger (1985) describes three models of experiments that existed in the last two decades of the 19th and the first two decades of the 20th century. The Leipizig model came out of Wundt's laboratory and involved observation of individuals without clear separation between research participants and those conducting the experiments. Often the researcher himself would serve as the data source in the experiment. On the other hand, the Paris model, which

emerged out of the hypnosis research conducted by Binet and his colleagues in France, more clearly conceptualized participants in a passive manner as data sources, somewhat akin to contemporary conceptualizations of participants. Often, participants were patients of the doctors conducting the experiments, and as such had existing close relationships with the experimenter. Finally, the Clark model which was developed in the United States by G. Stanely Hall and his followers, involved the investigation of large groups of anonymous participants who were thought to be representative of a certain population of interest, using a variety of measurements, but no experimental manipulation of conditions. In the early days of psychology then, experiments differed greatly from their contemporary counterparts. Winston (1990) argues that well into the later part of the 1920's psychologists did not distinguish between experiments that involved simple measurement, and experiments that involved manipulations of experimental conditions. In 1938 however, Robert Sessions Woodworth published his influential textbook *Experimental Psychology,* and in this work formally distinguished between correlational and experimental research. He defined experiments more narrowly as the manipulation of an independent variable and the measurement of a dependent variable, and defined research that only involves measurements of variables as correlational rather than experimental. This definition of experiments was speedily and widely adopted by psychologists. Morawski (1988) demonstrates that experimental models have undergone further changes since this time, based on the need to promote experiments as a way to remedy social problems. It is apparent that experiments in the contemporary sense of the word were not essential to early psychological research, which utilized a broader definition of experimentation and very different models of experimental research. Definitions and models of experiments have undergone substantial changes since the early days of psychological research.

Consistent with the previous claim, when one examines some of the famous psychological findings of the early days of scientific psychology, one finds a distinctive lack of experimental evidence. Indeed, the studies conducted in Wundt's laboratory employed a very restricted kind of self-report that typically was accompanied by more objective psychophysical data such as reaction times and other objectively recorded responses and could be considered experimental. However, other research conducted by Wundt, especially in the context of his interests in "Voelkerpsychologie" (folk

psychology), utilized diverse methods and sources of evidence including historical and archeological evidence, and did not incorporate any experimental evidence at all. A second famous German example from this time period can be found in Ebbinghaus' memory research, in which he sought to establish patterns and limits of human memory and forgetting. His studies, although attempting to show that experimental methods are suited to the investigation of human psychological processes, were based entirely on a lengthy series of attempts to learn and relearn lists of nonsense syllables. Ebbinghaus himself was the sole subject in his studies. Such self-study, however rigorously conducted and recorded, certainly would not satisfy the criteria for an experiment in the contemporary sense.

In early American psychology, the popular structuralist school, defined primarily by work conducted in Titchener's Cornell laboratory, employed broad range of introspective methods, some conducted in ways that could be considered experiments (Benjamin, 2007). William James, however, who has been called by some the founding father of American psychology, relied on diverse methods of introspection, physiological research, and deductive inference in his investigations. His disdain for experimental evidence can hardly be more clearly illustrated than through the following quotation from his *Principles of Psychology:*

> But psychology is passing into a less simple phase. Within a few years what one may call a microscopic psychology has arisen in Germany, carried on by experimental methods, asking of course every moment for introspective data, but eliminating their uncertainty by operating on a large scale and taking statistical means. This method taxes patience to the utmost, and could hardly have arisen in a country whose natives could be bored. Such Germans as Weber, Fechner, Vierordt, and Wundt obviously cannot. (1890, p. 192, as cited in Brysbaert & Rastle, 2009)

Other pre-eminent early American scholars like G. S. Hall and James McKeen Cattell typically also did not make extensive use of contemporary experiments, but preferred to collect large bodies of data based on the observations of teachers, parents, and child-care workers (Hall) and the mass testing of college students using a variety of psychophysical measures (Cattell).

In short, the emergence of psychology as a separate discipline at the end of the nineteenth century in both Germany and America was not marked by a sudden or a complete turn to experimental evidence in the contemporary sense of the word. Rather, psychology as an identifiable area of inquiry emerged well before Wundt's time as a sub-discipline of philosophy, and utilized both rational and empirical evidence. Wundt advocated for an integration of philosophy and psychological inquiry and viewed such inquiry as a subset of his philosophical projects (Araujo, 2012). Furthermore, psychologists during Wundt's time and immediately afterward did not rely exclusively or even mostly on evidence generated through methods that would now be considered experimental, but defined experiments much more broadly and employed a diverse set of methods for securing evidence, including case studies, self-report, introspection, and historical and archaeological studies. Early psychology therefore employed different models of experiments as one source of evidence. Experiments defined in the contemporary sense, as the manipulation of an independent variable among a group of anonymous subjects and the measurement of the outcome of the manipulation as a dependent variable, did not become the predominant source of evidence for psychology until the 1950's.

Evidence in Early Clinical Practice

Three main areas can be identified in the early emergence of clinical practice in psychology: the mental testing movement, the development of psychoanalysis, and the development of Lightner Witmer's clinical psychology clinic. The types of evidence used in each of these areas diverge greatly, so a separate treatment of each is warranted. Notably, evidence generated through experiments did not play a major role in any of these developments, and only in the mental testing movement was research conducted in a manner that could presently be considered as empirically scientific.

The Mental Testing Movement

The mental testing movement, a major predecessor of the assessment branch of clinical practice, has its roots in the work of Galton in the late 19th century (Brysbaert & Rastle, 2009). Galton sought to establish the heritability of intellectual abilities. In his anthropometric laboratory, he measured a wide variety of sensory and motor skills,

developed the first mental tests, and posited that sensory acuity provided a means to assess individual differences in intelligence (Galton, 1884). He intended to correlate various sensory measurements with indices of intelligence such as occupation and school achievement. However, the correlational component of his work never materialized and what remained of his research was primarily physical measurements such as height and weight and measurements of sensory discrimination (Brysbaert & Raestle, 2009). Although Galton was central to the development of the idea and early methods of mental testing, including a focus on individual differences as an area of study, he employed no experiments in the contemporary sense of the word. He did no manipulating of experimental conditions or comparisons of experimental groups.

The further development of the mental testing movement occurred in France, with the work of Alfred Binet at the end of the nineteenth, and into the twentieth century. Binet developed a series of tasks that could be used to measure outcomes that were highly correlated with teachers' ratings of children's school performance. These tasks, which grouped together were considered to measure intellectual capability, were adapted in English language measures developed by American psychologists like Lewis Terman. Robert Yerkes later developed both verbal/written and non-written intelligence tests for use by the American army during the First World War to evaluate and place recruits. The evidence used in the development of these tests involved measurements of various mental attributes of large samples of individuals and the calculation of correlations between these measures.. The early testing movement however was not based on data from experiments where variables were manipulated and compared across groups. The early mental testing movement, which constituted an important contribution to the development of clinical practice, was primarily based on descriptive, correlational rather than experimental evidence.

The Development of Psychoanalysis

A second major contribution in the emergence of clinical practice was the articulation and development of psychoanalysis. Freud, likely the most well-known figure in the history of psychology, whom many consider to be the father of psychotherapy, developed the practice of talk therapy during the late 1800s, as

psychology was emerging as a distinctive scientific discipline. Drawing on previous work by Breuer and Charcot, who had developed practices based on hypnosis to treat various forms of psychopathology, Freud, a physician, posited that unconsciously repressed conflicts were the cause of psychopathology, which could be cured if these conflicts could be brought into the conscious mind through free-association, remembrance, catharsis (a kind of emotional re-enactment and purging), and discussion in the context of psychoanalytic therapy. However, as Brysbaert and Rastle (2009) argue, Freud's approach was principle- and theory-driven rather than based on experimental evidence. The evidence employed by Freud was not generated through scientific experiment, but based on his own and others' clinical observations. In fact, Freud rejected the value of statistical research in psychotherapy and was also opposed to the recording of psychotherapy sessions, which he thought would alter their nature (Strupp & Howard, 1992). Instead, Freud used informal observations of his patients along with introspection and rational contemplation as the main sources of evidence in support of his theories. The evidence he and later proponents of psychoanalysis offered in support of their approach was primarily in the form of case studies. Freud and his followers described in detail the manner in which their approach helped particular patients and then provided a theoretical case formulation that stipulated the reasons why the cases proceeded as they did (Brysbaert & Rastle, 2009). There was clearly both an observational and an interpretive, rational component to Freud's case studies but experimental data played no role in them. Interestingly, psychoanalytic theories have since been criticized by many psychologists for being untestable through scientific experiment (e.g., Gruenbaum, 1984; Popper, 2004), although such critiques have in turn been called into question in recent years (e.g., Shedler, 2010; Westen, 1998). [17]

Witmer and Clinical Psychology in America

Experiments also did not play a major role in the early development of clinical psychology as a formal sub-discipline of psychology. Lightner Witmer, who has been

[17] The focus on Freud as the father of psychoanalysis is emphasized here from a historical perspective. It has to be acknowledged that contemporary psychoanalysis has changed substantially since the days of Freud and in fact has been examined scientifically in recent years in an effort to establish it as an evidence-based practice (e.g., Shedler, 2010).

called the father of clinical psychology, played a central role in the establishment of clinical psychology as a sub-discipline of psychology. Witmer established the first psychological treatment clinic at the University of Pennsylvania in 1896 (McReynolds, 1997). He was the first to propose an applied branch of psychology and coined the term clinical psychology to describe it. An experimental psychologist who had received part of his training in Wundt's laboratory, Witmer became interested in developmental pathologies of children, especially related to learning difficulties. A case of a boy who had difficulty with spelling, which Witmer was able to remedy quickly through a recommendation of vision correction by means of glasses, prompted Witmer to pursue the application of psychological principles to the assessment and treatment of learning and behavioural difficulties in children (O'Donnell, 1979). He eventually established a clinic where he assessed and treated numerous children (and later adults) with these difficulties. The individuals seen at his clinic usually received a thoroughly multi-disciplinary assessment (by a physician, social worker, and Witmer himself or one of his students), sometimes followed by educational treatment.

Contrary to mental testing approaches in the late 1800's, Witmer was concerned with determining causes of deficiencies and remedying them, rather than correlating deficiencies to low intelligence, which was more common approach at the time (O'Donnell, 1979) Based on records he kept of his cases, it appears that he employed procedures he had used in his psychophysical laboratory research as part of his assessment methods, and developed his treatment methods using creativity, trial and error, and his knowledge of psychological theories (e.g., some of his treatments relied on elements of early learning and memory research) (McReynolds, 1997). The evidence used by Witmer in the establishment of his treatments was therefore certainly not experimental in the sense of the use of controlled experiments involving comparisons of groups but rather was mostly based on general principles or learning and theories of psychopathology prevalent at the time. Witmer established the first clinical psychology journal, *The Psychological Clinic*, as a means of disseminating the findings of case studies from his clinic. His goal, in addition to remedying deficiencies, was also educational and activist in the sense that he wished to oppose genetic determinist accounts of human behaviour that were popular at the time and to advocate for the treatability of human deficiencies (O'Donnell, 1979).

Nonetheless, Witmer was a staunch supporter of experimental scientific psychology. He advocated for the separation of the American Psychological Association from more philosophical psychology, which was still part of the organization at the time, and for which he had little regard. He also proposed a program of training for clinical psychologists, which was based in academic settings and was taught by those familiar with experimental scientific research and its applications. However, Witmer's own experimental research was limited to basic psychophysiological inquiries (McReynolds, 1997). His experimental work did not directly connect to his clinical work, which was instead based on rational deductions from his own knowledge and experience and case reports. Overall then, early clinical practice in psychology relied on observational studies of individual cases, together with the measurement of physical and mental attributes of large samples of individuals, in the case of the mental testing movement. The history of experimental evidence as a basis for psychological practice and treatment certainly does not date to the early days of clinical practice. It has a much shorter history, which is detailed in the following section.

The Emergence of the Scientist-Practitioner Model

Scientific evidence from experiments conducted in a controlled laboratory setting began to play a more central role in clinical practice when clinical psychology became part of American academic psychology. In the decades following the establishment of Witmer's clinic, training programs in clinical psychology were established at various U.S. universities. Consistent with Witmer's proposed training model and an APA resolution in 1924, the PhD became the agreed upon standard of qualification for a clinical psychologist (Routh, 2000). However, no further consensus or accreditation mechanism to ensure standardization of training models existed. By the 1940's, lively discussion with respect to training standards for clinical psychology programs was wide-spread (e.g., Gregg, 1947; Morrow, 1946).

In the wider socio-political context, the treatment of mental health problems came into national focus following World War I, when veterans returned from the war with war-related mental health problems to an absence of mental health care facilities and personnel to respond to their needs. This lack of necessary services and service

providers became more pressing with the onset of the Second World War, at which time it was hoped that a similar fiasco could be avoided. Recognizing a major discrepancy between supply and demand for mental health care, the department of Veteran Affairs and the United States Public Health Services allotted funding to expand training in mental health care. They asked the APA to assist with strategies to increase training of professional psychologists in June, 1947, with a deadline of September, 1947 (Baker & Benjamin, 2000). The American Psychological Association formed a Committee on training in clinical psychology, chaired by Dr. David Shakow, which published a report recommending standards for training in clinical psychology that involves an integration of research and clinical training (APA, 1947). Shakow (1976) later clarified that the recommended training would involve a true integration of clinical and research training, such that clinically relevant research would be conducted in applied clinical settings by trainees in clinical psychology programs.

By 1949, accredited training in clinical psychology had expanded and clinical psychologists, along with other mental health professionals, were in high demand. However, given the rapid expansion of clinical psychology, concerns developed about communication and coordination across Veteran Affairs hospitals, mental health centers, and university programs. The United States Public Health Services and the APA sponsored a 15-day conference in Boulder, Colorado to discuss important issues in clinical psychology related to the education and delivery of clinical interventions by psychologists. From these proceedings came the 1950 report on *Training in Clinical Psychology* which endorsed the scientist-practitioner model of training wherein clinical psychologists were to be trained equally in research and practice in the areas of clinical research, diagnosis, and therapy (Raimy, 1950). The equal emphasis on research and practice was consistent with the demands and interests of the United States Public Health Service (USPHS).

Beyond merely appeasing the USPHS, the Boulder model of training had far-reaching implications and shaped the manner in which training in clinical psychology has been conducted since the inception of the model in 1949. The additional funding devoted to clinical psychology programs through the public sector became a viable source of research funding for university psychology departments, whose expanding clinical programs soon constituted a major component of graduate training in psychology

(Rosner, 2005). The Boulder model also cemented the status of clinical psychology as a scientific, research-based discipline. With the advent and widespread adoption of the Boulder model of training, psychologists began to be viewed as both scientists and practitioners, and financial support and accreditation for clinical training became contingent on the adaptation of such a training model. Thus, evidence generated through scientific research in the laboratories of graduate programs in universities across the United States came to play a more central role in clinical psychology (Baker & Benjamin, 2000).

The Clinical versus Actuarial Prediction Debate

In addition to the establishment of psychology as an academic discipline, the 1950's also brought a questioning of the merits of clinical judgment and a turn towards a greater emphasis on mathematical models of prediction of human behaviour, in opposition to models based on the judgment of the practicing clinician. In 1954, Paul Meehl published his book *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. In this work, Meehl called into question the utility of using clinical data in an attempt to predict human behaviour. Meehl argued that a statistical approach and more objectively gathered research data ought to be favoured over an approach that involves any non-statistical judgment and that clinical judgment is more prone to error than is a sound statistical approach.

Meehl's work was widely debated among clinical psychologists and compelled a significant divide between those who agreed and disagreed with Meehl. Rosner (2005) argues that this divide persists into the present day. Irrespective of the fact that some agreed with a more experimental and statistical model favoured by Meehl, many clinicians remain wedded to models based on clinical cases, judgement, and inference (Rosner, 2005), Nonetheless, Meehl's work further advanced the conceptualization of clinical psychology as a scientific discipline, and propelled the accumulation of research findings in various areas of clinical work, especially assessment, with the aim of developing the sort of statistical algorithms that were believed to be required in order to make sound clinical decisions.

A Half-Century of Accountability: Psychotherapy Research since Boulder

After Boulder, clinical psychology was firmly established as a discipline that required an integration of science and practice, and as a discipline that owes its existence to the needs and funding provided through the public health sector. Indeed, true to these roots, early reports on clinical training emphasized that university psychology departments were to discourage their clinical trainees from engaging in independent practice, which was considered to be inferior to and less desirable than practice that occurs in the context of the public health care sector. As clinical psychologists became accountable to funding sources and the public sector, clinical research became a means of establishing the legitimacy and value of clinical practice and psychotherapy. The second half of the 20th century witnessed increasing attempts to produce evidence that supported the utility of clinical psychology and psychotherapy.

The questions addressed by psychotherapy research (or the objects of evidence) have changed substantially over the last fifty years and along with them, the methods used to address them. Originally, psychotherapy research tended to focus on whether psychological treatment works (outcome research). However, research soon shifted to include investigations of what components of psychological treatment work and how they work (process research). In more recent decades, the object of evidence of such research has again shifted to concerns about what treatments work for particular conditions (focused outcome research). Psychotherapy evidence now can be classified in terms of the specific questions it seeks to answer concerning particular conditions and interventions.

Does Psychological Treatment Work?

Some of the earliest research on psychological treatment sought to establish simply whether psychotherapy works, that is, whether it leads to improvement in those conditions it seeks to ameliorate. Freud and some of his followers opposed the quantitative evaluation of psychological treatment, which they thought was too simplistic for a practice that was as complex as psychoanalysis. Despite this opposition, in an

effort to justify their work, psychoanalytic institutes began to encourage the collection of data on the outcomes of cases (primarily in terms of categorical data, using categories such as uncured, improved, much improved, and cured). Published results began to appear in the 1930's (Lambert, 1992). Around the same time similar data were collected and published on outcomes associated with other therapeutic approaches. Many of these studies had serious methodological shortcomings and did not include comparisons to untreated control groups.

Such shortcomings led Eysenck (1952) to conduct a famous study, which summarized findings from extant psychotherapy outcome research and compared them to spontaneous remission rates. Eysenck concluded that remission rates following treatment were comparable to spontaneous remission rates and called into question the efficacy of psychological treatment. Many aspects of Eysenck's study were questioned, including the methods he used to synthesize research data, but most prevalently Eysenck was criticized for the lack of a control condition in his study. For example, Luborsky (1954) criticized the fact that the studies summarized by Eysenck did not include random assignment to treatment and no-treatment conditions. He argued that those who sought treatment might differ from those who didn't in some systematic manner. Eysenck's findings thus spurred the publication of numerous treatment outcome studies that directly compared treated groups to wait-list or untreated control groups.

Over time, randomized assignment to treatment groups became increasingly valued. The randomized-controlled trial became the main method used to investigate psychological treatment. Randomized-controlled trials involve the comparison of a treatment group with a control group and the random assignment of participants to either the treatment or the control group (with the assumption that such random assignment eliminates systematic differences between the two groups). In response to Eysenck's study, the randomized-controlled trial therefore became the most widely accepted standard for outcome research, and an important method of evidence-generation in the evaluation of psychotherapeutic effectiveness.

Three decades later, Smith, Glass, and Miller (1980) pioneered the data synthesis methods of meta-analysis in order to statistically analyze and synthesize the

results of different outcome studies of psychological treatment. Meta-analysis provides a set of statistical techniques for the systematic review and synthesis of effect-size data from a variety of studies. By using these newly developed methods of statistical data aggregation, Smith and colleagues were able to calculate an overall effect size for psychological treatment from a sample of 475 outcome studies. They concluded that on average those who have received psychological treatment are 80% improved over similar groups of individuals who have not. This result provided tremendous relief for psychologists who had been disconcerted by Eysenck's study and subsequent questions about the efficacy of psychological treatment. However, the Smith and colleagues meta-analytic study was soon criticized for various methodological limitations of its own and for what some regarded as its overly broad and undifferentiated findings. Eysenck himself (1978) criticized Glass and Smith's method, arguing that their inclusion of studies irrespective of their quality called into question their findings. Gallo (1978) re-analyzed Glass and Smith's data and arrived at the conclusion that the effects of psychotherapy were weak and that the comparisons of heterogeneous methods, treatments, and measures made any conclusion from their data questionable. Despite such concerns, meta-analysis became an important tool in the legitimization of psychological treatment and meta-analytic evidence was widely used to support the efficacy of psychological treatments.

How Does Psychological Treatment Work?

A further question, that came into focus for researchers during the 1950's and 1960's concerned the mechanisms whereby psychological treatment works and the processes that occur during psychotherapy. This question became particularly relevant once some evidence for the general efficacy of psychotherapy had been established (Lambert, 1992). Psychotherapy process research was pioneered by client-centered therapists, particularly Carl Rogers during the 1950's (Strupp & Howard, 1992). Because video and audio recording of therapy sessions were employed by Rogers himself, such recordings were widely accepted by many client-centered therapists (recall that the recording or therapy sessions had been strongly rejected by proponents of psychoanalytic psychotherapy who argued it interfered with the therapeutic process). In the 1950's however, increasing social values related to democratization of knowledge

and free and objective inquiry prompted psychoanalysts to also begin to use videorecording as a way to study their methods in what they considered to be an objective manner (Rosner, 2005). Funding for mental health research also become contingent on the use of such methods (Rosner, 2005). Methods of content and process analysis, which were in part borrowed from communications, media, and political studies, allowed researchers to identify unique and common ingredients in successful psychological treatments. Various ways to code and quantify therapist-client interactions in therapy sessions were developed and used to analyze recordings of psychotherapeutic sessions in a detailed and laborious manner (Strupp & Howard, 1992).

Findings from process research were fruitful but also occasionally disconcerting to clinical psychologists. Common factors that played an essential role in successful treatment were identified. They included the establishment of a therapeutic alliance and the appropriate use of empathy and unconditional positive regard. However, attempts to separate the effects of such common factors from the effects of interventions that were particular to specific approaches to psychotherapy were unsuccessful (e.g., Norcross, 2002; Imel & Wampold, 2008). Findings that attributed the outcomes of psychotherapy to common factors rather than specific interventions or treatments called into question the need for specialized treatment approaches and training in specific models of psychotherapy, as had been advocated by the Boulder Model. Their professional expertise threatened by such findings, psychologists turned to the possibility that reliable effects of particular approaches to psychotherapy might become evident for specific kinds of client concerns and difficulties.

Which Psychological Treatment Works for What?

By the 1970's and 1980's, more elaborate and standardized systems to classify and diagnose mental health problems that had emerged during the 1950's and 1960's achieved widespread use amongst mental health professionals, including clinical psychologists. From basic categories of psychosis and neuroses, elaborate taxonomies of psychopathology had been developed and publicized by the American Psychiatric Association in its Diagnostic and Statistical Manuals of Mental Disorders. The DSMs

encouraged the development of medical treatments for a wide variety of particular disorders. Gradually, significant diversity in approaches to the treatment of different forms of psychopathology became the norm in both psychology and psychiatry. Within clinical psychology, numerous models of treatment had been developed -- psychoanalytic treatment, psychodynamic approaches, behavioural therapy, cognitive therapy, client-centered approaches, and several others. These psychological treatments relied on models that were sometimes theoretically incompatible. Around this same time, the pharmacological industry began to develop medications for different mental health problems. This posed a threat to clinical psychologists, whose treatments were more time-intensive and difficult to implement. The greater variety of psychological and pharmaceutical treatments available motivated clinical researchers to investigate questions concerning what kinds of psychological and/or pharmaceutical treatment worked well for different psychological disorders.

In addition, randomized-controlled trials had been established in medicine as the standard for the evaluation of medical treatments and there was increasing pressure to apply these same standards to the evaluation of psychological interventions (Rosner, 2005). In this context, researchers in clinical psychology and psychotherapy began to examine the efficacy of particular treatments for particular forms of psychopathology, not just in comparison to no treatment, but also in comparison to other treatments and medications. The National Institute for Mental Health became actively involved in promoting the investigation of psychological treatments using randomized controlled trials by allocating funding to such studies (Rosner, 2005). The first major study with this aim was funded by the National Institute of Mental Health in 1977 and compared various approaches to the treatment of depression, including interpersonal psychotherapy, cognitive-behavioural therapy, and a psychopharmacological treatment (Rosner, 2005). A main finding of this study was that all three approaches to treatment were effective and few differences between the treatments could be identified. Once again, it seemed as if psychotherapy was generally effective, but that evidence in support of matching particular therapeutic treatments to particular client difficulties was lacking. Nonetheless, the fact that psychological treatments like cognitive-behavioural therapy seemed equally as effective as pharmacological treatment provided some encouragement.

As experimental studies comparing different treatments for particular disorders became more widespread, additional difficulties became apparent: lengths of treatments varied, individual providers delivered treatments in various ways, and there were many different ways to measure the outcomes of treatment. These problems, although posing little difficulty to the practitioner of psychotherapy, were challenging to the scientists who relied on experimental designs that involve randomized selection and assignment of clients or patients to treatment conditions that are strictly controlled or standardized in some manner. In order for treatments to be directly amenable to examination in a scientific experiment, therapeutic processes had to be standardized and held constant across treatments. Therapeutic outcomes also needed to be defined and measured in similar ways. Manualized treatments with clearly specified protocols and standardized ways of measuring consistency and appropriateness of treatment delivery were thus developed and became increasingly widespread. A proliferation of standardized measures of outcomes that could be used reliably and efficiently for various treatment effects also became available.

In an ironic twist, the attempt to generate evidence to support the use of particular psychological treatments shaped and altered the very treatments that were tested and delivered. When evidence for a treatment's efficacy in terms of superiority to other treatments through a randomized-controlled trial became the standard by which treatments were judged, a treatment's amenability to such experimental investigation became a standard for the treatment's demonstrable efficacy. Soon, standardization and manualization that allow for investigation through randomized-controlled trials began to be viewed as necessary features of efficacious treatments, as much or more than features that allow for the more general investigation of therapeutic efficacy. This development set the stage for the emergence of the EBP movement, in a manner that is explicated in the next chapter.

Evidence and Psychological Treatments

The story of evidence in the development of psychological treatments presented in this chapter can be summarized as follows. In the early days of psychology, before the development of psychological treatments, psychology came to be defined as a

scientific experimental discipline. However, scientific experiments only came to play a prominent role in the development of psychological treatments at a much later point. In historically early types of clinical practice, including the mental testing movement, Freud's psychoanalysis, and Witmer's development of the methods used in his psychological clinic, relevant evidence was based on non-standardized methods of psycho-physical measurement, theoretically-informed practices, and trial and error, and used mostly descriptive and case summaries as evidential support. Only in the middle of the 20th century did clinical psychology formally come to be defined as an integration of practice and experimental research. However, the questions addressed by such research continued to change significantly over time, as did the kind of evidence used to answer them. In the early days of experimental research on psychotherapy, the overall efficacy of psychotherapy was of central interest. Evidence used included broad counts of treatment successes and failures. Eysenck's (1952) comparison of treatment success rates to spontaneous remission rates eventually prompted an effort to demonstrate the efficacy of psychological treatments relative to no treatment, leading to the development of randomized-controlled trials and meta-analysis. A second question addressed in experimental psychotherapy research concerned the specific components of psychotherapy that are effective. Evidence used to address this question was generated through both quantitative and qualitative analyses of psychotherapy sessions. Although such research was able to identify factors that distinguished successful from unsuccessful treatment, it failed to identify specific factors in particular treatments that contributed to their relative success. In more recent decades, studies have sought to compare the efficacy of various treatments for particular conditions. In order to allow for such comparisons, standardized diagnostic systems, manualized treatments, and standardized outcome measures were needed. Once psychotherapy research acquired an experimental form, the manner in which psychotherapy was conducted changed as a result of the way in which it was studied. Psychotherapies that were easily amenable to experimental study became more widely practiced. At present, whether or not a treatment is amenable to experimental study and comparison has become part of the evidence of its efficacy. In this context, the evidence-based practice movement has emerged. However, it should not be forgotten that for most of the history of psychological treatment and intervention, scientific experiments did not play a major role in the development or evaluation of treatments.

Chapter 5:
The History of Evidence-based Practice

The first articles on evidence-based practice (EBP) that can be located in a PsycInfo search appeared in psychiatric journals in 1995 (e.g., Furukawa, 1995; Sharpe, Gill, Strain, & Mayou, 1996). These early articles primarily focused on introducing evidence-based practice, which in the preceding years had entered the world of general medicine as a potential new paradigm. In the following years, 19 such papers on EBP appeared primarily in journals with a psychiatric/medical focus. Clinical psychologists, however, were quick to follow suit, at first with just a single article on the topic (Shapiro, 1996), but by 1998 numerous articles on EBP in psychotherapy had appeared in the psychological literature (e.g., King, 1998; Sanderson, 1998; Treasure & Kordy, 1998).

These early articles marked the beginning of a decade of striking proliferation in the literature on EBP, both in terms of geographic reach and the number of publications on the topic. Articles on EBP in psychology came out of Europe (e.g., Dent, 1999), Australia (e.g., King, 1998), and the United States (e.g., Somerfield & Costa, 1999), while psychiatric articles also were published from Japan (e.g., Furukawa, 1995) and many other countries. The rapid proliferation of EBP becomes apparent when one examines the number of publications on this topic on PsycInfo. The title search term "evidence-based" yields 65 results prior to the year 2000. The same search term, restricted to publications between 2000-2005, yields 1078 hits, and 3565 hits since the formation of the APA Task Force in 2006 (1694 of which were published since 2010). Thus, it seems fair to conclude that: (1) the notion of EBP is a relatively recent entrant into the field of psychology, dating back a mere 17 years, and can be traced to its roots in medicine as evidenced by the fact that the earliest publications on the topic entered psychology from psychiatry; (2) the term EBP slowly entered psychological discourse between 1995 and 2000, but quickly rose to prominence, resulting in an exponential increase in publications on the topic in the past decade; and (3) at present, publications

on EBP continue to be produced and there is no evidence of a decline in this trend in recent years.

The above conclusions, particularly that EBP has its roots in medicine and entered psychology more formally around the middle of the first decade of the new millennium, are supported and further contextualized by a more formal review of the history of EBP, as provided below.

The Evidence-based Turn in Medicine

Precursors of Evidence-based Medicine

As already noted, the EBP movement has its roots outside of psychology, in the discipline of medicine. The phrase "evidence-based" was coined in the context of medicine (Eddy, 1990), the premises of the EBP movement were first advanced in medical contexts (Evidence-Based Medicine Working Group, 1992; Guyatt, 1991), and the guidelines that have informed much of the psychological discourse on EBP were drafted for the purpose of EBP in medicine (Sackett et al., 2000). A confluence of intellectual developments in the century preceding the emergence of EBP in medicine paved the way. These included changes and developments in the way in which medical treatments were evaluated, along with the development of new statistical methods that allowed for quantitative comparisons of treatments in the 20th century.

At the end of the 18th century, new techniques and methods were developed in order to scientifically examine various medical practices. Different medical practitioners also proposed a growing number of models of disease and treatments for disease. Treatments for disease at the time were evaluated primarily by means of public demonstrations of their efficacy. These demonstrations often relied heavily on charisma and showmanship, which heavily influenced the audience's judgment of the efficacy of the treatment demonstrated. For example, in the late 18th century, Anton Mesmer proposed a theory of disease, called animal magnetism, and a treatment for diseases that involved the use of magnets and iron rods. He performed his new treatment in public demonstrations to large audiences, who attended in part for the entertainment

value. His aim was to convince them of the treatment's effectiveness. Concerns arose about the legitimacy of such practices in establishing the value of a treatment. These concerns prompted King Louis XVI of France to establish a Royal Commission led by Benjamin Franklin to specifically examine Mesmer's theory, which had become widely popular and influential (Wampold & Bhati, 2004). Out of this commission, two practices used to evaluate medical treatments emerged: the comparison of various medical treatments to each other and the comparison of a treatment group to a placebo or control group. Both of these techniques were further refined and developed in the following century and became vital to the development of the randomized controlled trial. Nonetheless, the placebo control group comparison came to be widely used in medicine only in the mid-20th century (Wampold & Bhati, 2004), primarily because mathematical methods to evaluate such comparisons were unavailable until that time.

In addition, statistical methods had to be developed to provide the quantitative means to evaluate placebo-controlled trials. Fisher provided such statistical techniques in the early 20th century, when he utilized the concept of randomization, which had previously been used to determine the order of administration of stimuli in psychophysical research, and proposed it as a means of constituting experimental and control groups (Dehue, 1997). Fisher also provided statistical tools provided a means to disentangle random error from the systematic variance that can be attributed to the effect of a treatment. This allowed for the quantification of changes attributable to specific interventions, usually referred to as treatment specificity (Wampold & Bhati, 2004). Fisher popularized his experimental and statistical methods in his 1935 book *The Design of the* Experiment (Dehue, 1997). Propelled by the new feasibility of comparisons between treatment groups, by the mid-20th century, treatment specificity came to be the standard used in the evaluation of medical treatments. In addition to demonstrating that a treatment leads to improvement, it also had to be demonstrated that it exerts its effect through its specific ingredients rather than more general psychological factors. The double-blind randomized-placebo-controlled trial provided the means by which the specific effects of a treatment could be established. Against the backdrop of increasing numbers of randomized-controlled trials of medical treatments conducted and published in the later part of the 20th century, evidence-based medicine arose.

The Birth of Evidence-based Medicine

Although randomized-controlled trial designs began to be widely used in medicine by the 1950's (and shortly thereafter in psychology), the Food and Drug Administration did not require randomized-controlled trials of the efficacy of a drug in order to approve its use until the 1980's (Wampold & Bhati, 2004). This requirement, however, provided the final push to establish the randomized-controlled trial as the gold standard in the evaluation of medical treatments. In the second half of the 20th century, randomized-controlled trials were widely conducted and provided the medical community with a growing body of research findings about various medical practices. Instead of providing definitive answers to medicine's great questions or changing the way medicine was practiced, however, medical training continued as an apprenticeship model based on authority and established tradition and expertise. Amidst a series of studies that highlighted the wide discrepancies between the findings of randomized-controlled double-blind studies of treatments and the treatments that were actually being used, growing concern about the lack of integration of research into medical practice was voiced in the medical community (Wyer & Silva, 2009). This discourse created a fertile environment for the establishment of evidence-based medicine.

The earliest articulation of the principles of evidence-based medicine was provided in the American Cancer Society's guidelines for cancer-related health check-ups in 1980 (Eddy, 2005). These guidelines suggested that medical practice should be based first on evidence supporting its efficacy, and secondly on a consideration of costs, benefits, and feasibility. In response to these guidelines, a proliferation of literature reviews and guidelines published by various commissions and organizations on a large number of medical practices and methods occurred. The term 'evidence-based' itself, however, was not coined until a decade later in a 1990 article in the journal of the American Medical Association (Eddy, 2005). The article (Eddy, 1990) appeared in the context of a series of articles that dealt with the shortcomings of expert judgment in medicine and called for greater utilization of research evidence in medical practice. It examined various ways of generating guidelines for practice, including "consensus based," "preference based," "outcomes based," and "evidence based" methods.

The term ‘evidence-based medicine’ was formalized shortly after in 1990 and 1991 in two articles that came out of the deliberations of an evidence-based-medicine working group formed by the American Medical Association (Evidence-Based Medicine Working Group, 1992; Guyatt, 1991). The following years saw publication of a large number of summaries of available research evidence on various clinical topics and the formation of organizations dedicated to the generation of such research syntheses. Most famously, the Cochrane Collaboration, an international initiative aimed at creating and maintaining updated systematic reviews of research findings on numerous medical practices and treatments to aid physician decision-making, was formed in October 1993 (Cochrane.org, n.d.). The intuitive appeal of the notion that medical practice should be based on the conclusions and knowledge generated by scientific study facilitated the success of evidence-based medicine as a standard for medical practice and training.

In 2001, the Institute of Medicine, a non-profit health care policy advisory arm of the National Academy of Sciences published a report by its Committee on the Quality of Health Care in America. In this report, the committee described widespread disparities in medical training and health care delivery, and suggested that evidence-based medicine should be used as a training and practice standard that would ameliorate these disparities. The report touted evidence-based medicine as a complete revision of the health care delivery system that promised to reduce error and increase patient and provider satisfaction. The committee suggested ten principles that were to guide the revision of the health care system, including that decision-making was to be evidence-based, that knowledge and information were to be shared freely, and that care was to be customized to patient preferences. They also adopted the following definition of evidence-based medicine from Sackett and colleagues (2000): “Evidence-based practice is the integration of best research evidence with clinical expertise and patient values” (p. 147). This definition was to become highly influential in the deliberations on EBP in psychology in a manner discussed further below. By the turn of the millennium, evidence-based medicine had become a catch-phrase, although its central premises, feasibility, and practical applicability were also beginning to come under attack (e.g., Eddy, 2005). Nonetheless, various other health care disciplines followed medicine’s lead and began to develop their own guidelines for EBP.

APA's Early Response:
Treatment Guidelines for Best Practice

In the early days of the EBP movement, APA (2006), according to a report by its Task Force on Evidence-based Practice, responded to calls for greater integration of evidence into practice by focusing its attention on guidelines for best practice. Sparked by concerns that guidelines for practice that were emerging out of medicine might be used inappropriately to unduly restrict practice by commercial health care organizations, APA formed a joint task force[18] constituted by three of its professional boards and committees (Board of Scientific Affairs, Board of Professional Affairs, and Committee for the Advancement of Professional Practice). The task force considered standards for guidelines, particularly with respect to the kinds of evidence on which guidelines ought to be based, culminating in a policy document that recommended that guidelines be based on both research data and clinical expertise. The policy document emphasized that guidelines need to evaluate treatments based on both their efficacy, in terms of the causal effect of the treatment on a specific disorder, and clinical utility (i.e., generalizability, feasibility, costs, and benefits; APA, 2002). Subsequently, various psychologists evaluated existing practice guidelines on the basis of these recommendations. Concerns arose over guidelines that recommended the use of medications over the use of psychological interventions for the treatment of mental disorder without clear evidence to support such recommendations. Further action was deemed necessary and spurred psychology's more formal entry into the arena of the EBP debate.

[18] For the purpose of this document, Task Force will be capitalized to designate the Task Force on Evidence-Based Practice, which will be discussed at length throughout this document. Any references to other task forces formed by APA will not be so capitalized.

Clinical Psychology's Early Response: Empirically Validated/Supported Treatments

Psychology, in the words of John Norcross (personal communication, April 28, 2011) [19], was "not an early adopter" of EBP but also "wasn't late to the table." Norcross (April 28, 2011) argues that psychology was "an early adopter of the spirit of evidence-based practice" but "a mid-adopter of the terminology and having a formal policy statement." Before psychology entered the EBP arena, however, the notion of "empirically validated treatments" was formulated. This development was to be a fateful turn that both inspired the later formation of a Task Force on Evidence-based Practice in Psychology (APA, 2006) and continues to be responsible for the great confusion about the distinction between the two approaches that persists into the present day. A clarification of the two approaches and their differences follows.

The empirically validated treatment movement was developed in the context of a task force on the promotion and dissemination of Empirically Supported Treatments (EST) of Division 12 (the Society of Clinical Psychology) of the American Psychological Association during the presidency of David Barlow. In Barlow's (2011) own words:

> The purpose of that task force was to really make all the mental health professions and the public at large and policy makers more aware that we have very strong treatments for a variety of disorders that perhaps weren't widely recognized and weren't being widely administered. (personal communication, August 6, 2011)

Barlow's self-stated goal was "to really increase awareness and therefore greater dissemination of some of these techniques to give psychologists more tools with which to work, to increase recognition that what they were doing was effective by the public" (personal communication, August 6, 2011). The task force was chaired by Diane Chambless. Its members included academics and private practice/hospital clinicians

[19] In the present study, five of the members of the APA Task Force were interviewed. They all responded to requests to be interviewed and agreed to schedule an interview. All 18 Task Force members were contacted but the remaining Task Force members did not respond to requests for interviews or did not agree to be interviewed.

who ascribed to various theoretical orientations. The task force focused on training and promotion of psychological treatments. Task force members generated a stipulative definition of empirically validated treatments based on a number of specific attributes (Chambless et al., 1996). They defined two types of empirically validated treatments, well-established treatments and probably-efficacious treatments. They stipulated that well-established treatments have to be supported by studies utilizing a treatment manual and clearly specifying characteristics of the client population used. At least two studies conducted by separate investigators that show superiority to a placebo or equivalence to a well-established treatment or a large number of well-designed single case studies with comparisons to another treatment were required in order for a treatment to be considered well-established. The task force also established a second category of treatments, which they called "probably-efficacious treatments." This category included treatments supported by two efficacy studies conducted by the same investigator, two studies showing superiority to a wait-list control, two studies flawed by heterogeneity in the client samples, or a series of single case design studies.

One of the most controversial components of the task force's work was a list that they compiled, of various psychological treatments for specific psychological problems that fit the criteria for well-established and probably efficacious treatments. The rationale for the creation of this list, according to Chambless (1996), was the sense that time or lack thereof was a major obstacle to the incorporation of research into clinical practice. It was thought that most clinicians did not have the time to locate, review, and consider relevant research findings for their practice and that, subsequently, providing them with lists of treatments that are supported by research would be a valuable shortcut that would facilitate clinical practice supported by research findings. The list the Task Force created was preliminary, based on Task Force members' knowledge, and was not intended to be exhaustive. In fact, the Task Force recommended that a more exhaustive list be generated and updated regularly. Such updates indeed occurred roughly annually in the following years (e.g., Chambless et al., 1996; Chambless & Ollendick, 2001) through the continued work of the task force, which became a permanent part of Division 12's committee structure as the 'Dissemination Subcommittee of the Committee on Science and Practice' (APA Division 12, n.d.). The task force clearly acknowledged in all of its documents that this list was non-exhaustive and that the absence of a

treatment from the list did not suggest that said treatment was not effective (Chambless et al., 1996). The intended audience, which included the general public, practitioners, and third-party payers, however, did not apply the list in that spirit. Instead, they tended to take the view that treatments can be either effective or ineffective and that only those treatments listed were effective. For example, managed care and health maintenance organizations began to use the list as a basis for decisions about treatment funding. Clinicians at times could no longer be reimbursed for treatments they offered if they were not deemed effective based on inclusion in these lists. Attempts by the Division to ameliorate these effects by emphasizing the tentative nature of their list, including changing from discussions of 'empirically validated treatments' to 'empirically supported treatments,' did little to change such practices.

Other divisions of APA responded to these concerns by forming their own task forces. Division 29 (Psychotherapy) formed a task force on empirically supported psychotherapy relationships to highlight the important role of non-specific factors in treatment outcomes. Division 17 (Counseling Psychology) also formed a task force on empirically supported counselling psychology treatments (Goodheart et al., 2005). Growing concerns about the social and disciplinary implications of the empirically validated/supported treatment movement set the stage for the formation of the broader APA Task Force on Evidence-based Practice.

EBP in Psychology

EBP became evermore central to health-care policies and practices in the Unites States and Europe by the end of the millennium. Amidst growing concerns about the rising costs of health care as well as inconsistencies in health care delivery, third-party payers (e.g., insurance companies, health maintenance organizations, government health services) and government agencies increasingly called for services to be justified by research evidence for their reimbursement (D. Barlow, personal communication, 2011; APA, 2006). In the absence of clear standards for EBP in psychology, lists generated by the Division 12 task force on empirically validated treatments began to be used as standards for reimbursement in a manner that many of the Task Force members, including David Barlow (who had formed the Division 12 task force on

empirically validated treatments), perceived as abuses by managed care companies. Barlow summarized these trends as follows:

> There were some on the Task Force who, quite reasonably, felt that unless they employed a very small narrow range of procedures they might not, that practitioners might not be reimbursed for their practice, because there had been some abuses by managed care companies leading up to this point. The managed care companies basically said you can only do so many sessions and it has to be something on our list, anything else is not reimbursable so it was that kind of fear. (personal communication, August 6, 2011)

There was a growing sense among psychologists, especially many who eventually came to serve as members of the APA Presidential Task Force on Evidence-based Practice, that urgent action was required. Several of the Task Force members who were interviewed for this project commented on these sentiments: "I think there was concern that livelihoods were moving out of their control as opposed to the society trusting a licensed clinical psychologist" (S. Hollon, personal communication, April 26, 2011). "Bluntly stated we have plentiful and convincing research for psychological treatments. Nonetheless we did not have a formal position paper on Evidence-based practice" (J. Norcross, personal communication, April 28, 2011).

The 2005 formation of the APA Task Force on Evidence-based Practice thus constituted the APA's response to these developments and marked the formal entry of 'Evidence-based practice' into the discourse surrounding and policies governing psychological practice. Dr. Ronald Levant, the 2005 president of APA, was a key figure in the formation of this Task Force.

Levant compelled the formation of the Task Force in the year preceding his presidency, while he was president-elect of APA, prompted by his self-described concerns about the empirically-validated treatment movement. His perception was that this movement diverged from evidence-based movements in medicine that emphasize more general definitions of EBP instead of the generation of lists of treatments that meet particular criteria (specifically the 2001 Institute of Medicine statement). Levant described his motivation for the formation of the Task Force as follows:

> I formulated the idea that we needed to develop a policy analogous to that of the IOM policy, which we would call evidence-based practice in psychology, so that's kind of how it originated. I felt that the Division 12 lists were potentially harming psychology, although I understand that Dave Barlow who is a friend and a colleague did not have that intention. In fact, his intention was very clearly in setting up that task force, was to counteract the, at that time, 1995/1994 the overwhelming trend towards viewing medications as far more effective than psychological treatments and the treatment of mental illness. So, that was his intent, and that's a valid one and one I support, but I felt that it was having the unintended impact of forcing psychologists into procrustean beds that really didn't fit effective psychological practice, so that's really how it got started. (personal communication, June 27, 2011)

From a review of the documents of the Task Force as well as interviews with Task Force members, much information can be gathered about the formation and proceedings of the Task Force, including how members were selected, deliberations occurred, documents were generated and reviewed, and how members felt about their work on the Task Force.

Member Selection

Members for the Task Force were selected by APA president-elect (at the time) Ronald Levant in close consultation with Geoffrey Reid, the senior staff member in the practice directorate assigned to the Task Force project and APA staff member Lynn Bufka (R. Levant, personal communication, June 27, 2011). A main objective, according to Levant, was to select members that represent all constituencies active in in the caucuses of the APA. In the Task Force charge, diverse representation in the Task Force was described as follows:

> The Task Force incorporates scientists and practitioners from a wide range of perspectives and traditions, reflecting the diverse perspectives within the field: Clinical expertise and decision-making; health services research; public health and consumer perspectives; RCT science; full time practice; clinical research and diversity; non-RCT clinical research; health care economics; EBP research/training and applications. (APA, unpublished document, 2004)

In addition to representation of all caucuses, Levant described two further criteria that he used in selecting members for the Task Force: "I wanted people who met two criteria:

that they were respected by their caucus or constituency, probably broadly constituency, and that they were statesman-like people who could hear other points of view and could compromise" (personal communication, June 27, 2011). The process through which members were chosen according to Levant was rigorous and included the review of CV's, collection of recommendations, and careful consideration of the criteria laid out.

Various Task Force members had different recollections and rationales with respect to their invitation to serve on the Task Force. For example, John Norcross recalled that he was invited in the early days of the formation of the Task Force and, as a member of the Board of Directors, was also consulted in member selection along with other members of the APA Board of Directors and Counsel of Representatives. He described that in addition to the diversity criteria employed in member selection outlined above, potential members' views on EBP based on their previously published opinions were also considered. The aim was to represent various differing views and opinions (personal communication, April 28, 2011). Most members were also able to identify reasons why they were personally invited to serve on the Task Force. Bruce Wampold, for example, posited that he was selected to represent Division 17 (Society for Counseling Psychology), for which he was a board member and which had lobbied to have him represent their interests on the Task Force (B. Wampold, personal communication, February 22, 2011). Steve Hollon identified his previous work as a member of the task force on treatment guidelines as a main reason for his invitation to the Task Force. He describes this as follows:

> I'd been involved probably about five years earlier with something that's kind of like this and probably on that committee hadn't embarrassed anyone too badly and seemed to do my work and had some views which probably would have represented the kind of science wing of the psychotherapy community along with several of the other folks on the Task Force. That's probably how my name got brought up, because of both the work I'd done, and the role I played on earlier committees. (S. Hollon, personal communication, April 26, 2011)

Finally, David Barlow identified his role on Division 12's task force on empirically validated treatments as a main reason for his invitation to the Task Force. He thought he was "invited on this Task Force as a clear principal proponent of evidence-based practice" (D. Barlow, personal communication, August 6, 2011). Based on their own

recollections and reports by Dr. Levant, Task Force members were thus invited to represent caucuses, various viewpoints, and because of their reputations and expertise. No information about the number of members who were invited or how many invitations were declined is available.

Purpose of the Task Force

The officially stated purpose of the Task Force based on the Draft Policy Statement submitted to the APA Council of Representatives was as follows:

> In this report, The Task Force hopes to draw on APA's century-long tradition of attention to the integration of science and practice by creating a document that describes our fundamental commitment to sophisticated evidence-based psychological practice and takes into account the full range of evidence that policy makers must consider. We aspire to set the stage for further development and refinement of EBP for the betterment of psychological aspects of healthcare as delivered around the world. (APA Presidential Task Force on Evidence-based Practice, unpublished document, August, 2005)

The official Task Force charge started with a provision of the definition of EBP from the Institute of Medicine's statement, as the most widely accepted definition at the time. Task Force members, despite their diverse background and perspectives, were asked to assume that the Institute of Medicine's definition is "generally acceptable" and to use it as a starting point in their deliberations. They were asked to expand on this definition by incorporating greater attention to patient preferences, multicultural perspectives, and a "practitioner scientist" role for clinicians. The Task Force was tasked with examining: (1) the broad range of evidence and its role in practice, (2) the role of clinical expertise in clinical decision making, and (3) the role of patient values in treatment decision-making (APA Task Force Charge, unpublished document, 2004). The Task Force deliberations on these topics were to result in a set of recommendations for APA governance action as well as a policy document to guide wider social practice.

Proceedings

The deliberations of the Task Force occurred relatively quickly over the course of approximately one year. At the outset of deliberations, APA staff members assembled a list of relevant references based in part on recommendations by Task Force members and other potentially interested parties, which all Task Force members were then asked to review. Task Force members met face-to-face on two occasions in October 2004 and January 2005 (Presidential Task Force on Evidence-based Practice, Spring Consolidated Meeting Agenda Item, unpublished document, March 2005). APA President at the time, Ronald Levant described that at the first meeting members were given a short amount of time to present their perspectives on EBP (they were limited to 3-5 slides to reduce "speechifying") and a chance to debate their respective views to begin the deliberations of the Task Force. A deliberate effort was made to group people with opposing perspectives and to seat them close to each other in order to encourage open and free discussion (R. Levant, personal communication, June 27, 2011). During the remainder of the proceedings, the Task Force broke off into small writing groups to work on various parts of the document (e.g., clinical expertise, research, patient characteristics, introduction, and integration) and reconvened to discuss and review drafts as a whole group (APA Presidential Task Force on Evidence-based Practice Meeting Agenda, unpublished document, January 2005). The Task Force alternated between such small-group writing sessions and large-group discussions in an iterative process until consensus on wording of particular sections had been achieved (D. Barlow, personal communication, August 6, 2011). Between the meetings of the Task Force, phone meetings and email discussions were used to revise and rework various sections of the Task Force documents.

Major Debates

Most of the Task Force members recalled debates and disagreements that occurred during the deliberations of the Task Force. The differences in opinion between various Task Force members, according to the members interviewed, mostly centered around the three components of EBP: evidence, clinical expertise, and patient preferences. More specifically, the way in which these three components ought to be

weighted during clinical decision-making was a highly contested area during the deliberations of the Task Force. One of the areas of divergence of Task Force members' views concerned the importance of research evidence compared to clinical judgment. In the words of Task Force member David Barlow:

> I think there was a component on the Task Force who clearly adhered to the supremacy of clinical prediction over statistical prediction and there was another group, including myself, who were much more confident in the more empirical statistical prediction, so that was kind of the implicit divergence on the Task Force but explicitly everybody agreed that you know the evidence is basically empirically controlled observation. (personal communication, August 6, 2011)

Most Task Force members who were interviewed commented on this divide on the Task Force but agreed that ultimately the Task Force came to an agreement that research evidence was to be weighted more heavily than clinical expertise. The Task Force incorporated this weighting into its policy by conceptualizing respect for research evidence as part of clinical expertise, thus making expertise ultimately dependent on a consideration of relevant research evidence. For example, Bruce Wampold recalled the following:

> So you can't be a clinical expert if you ignore or are ignorant of research evidence. And that was the way their perspective was accommodated, and I think it satisfied the practitioners too, because you know how could they argue against, well you know, we should be at least aware of what the evidence is, and the other part of this is then, it's the psychologist then who needs to interpret how the evidence applies to a particular patient. (personal communication, February 22, 2011)

Other members of the Task Force suggested more strongly that the Task Force agreed to privilege research evidence over clinical expertise and patient preferences, as is evident in David Barlow's statement above.

A second area of debate on the Task Force more specifically related to the evidence to be used in EBP. In particular, there was disagreement with respect to the extent to which evidence generated through various methods ought to be weighted in clinical decision-making as expressed in the following description by a Task Force member:

> I believe I'm an ardent proponent of methodological pluralism, that's largely dependent upon the question being addressed. The entire Task Force embraced that notion, though the devil was lurking in the details. To what extent do we price a controlled outcome study over a naturalistic effectiveness? What are the relative advantages and disadvantages of say systematic case studies. (J. Norcross, personal communication, April 28, 2011)

Within this same debate, the adequacy of current evidence as a means to guide clinical decision-making was also an area of debate as illustrated by the following:

> I think there was an attempt by some of the Task Force to minimize the evidence we already had with the goal of, to state that you know the kind of evidence we have is very very limited in terms of its applicability to practice and then there were others such as me who said that's not true, the evidence actually is quite good. It's not certainly everything we need to know but it's actually quite good and we can focus on how we can improve that so it was that tension I think throughout I think that showed up. (D. Barlow, personal communication, August 26, 2011)

In fact, based on the expressed views of those interviewed for the present study, there was no consensus among Task Force members interviewed with respect to whether the list of various types of evidence provided in the Task Force statement was to be interpreted in a hierarchical fashion as some members argued or, instead, as a list of multiple methods that could be used to generate evidence for difference purposes.

Resolutions

Despite the divergences in Task Force members' views on the issues described above, the draft policy statement on EBP includes the following statement:

> Perhaps the central message of this task force report, and one of the most heartening aspects of the process that led to it, is the consensus achieved among a diverse group of scientists, clinicians, and scientist-clinicians from multiple perspectives that EBPP requires an appreciation of the value of multiple sources of scientific evidence. (APA, unpublished document, 2005)

Most of the Task Force members interviewed echoed this sentiment and emphasized the extent to which discussion and collaboration facilitated the formation of friendship

and greater agreement amongst members. A Task Force member expressed this as follows:

> I fondly recall the good faith efforts of people to discover common ground. For me it was a template of how to proceed on fairly contentious matters. And third, that Task Force initiated some brand new friendships and deepened several collaborations with people with whom I'm still now working, so I look back on it quite fondly. (J. Norcross, personal communication, April 28, 2011)

A number of Task Force members also described that over the course of the deliberations of the Task Force their own views changed slightly:

> A lot of us came to really like and respect the other folks that were sitting in the room from different perspectives. I'd be very surprised if Bruce Wampold changed his perspective on the basis of our conversation. I'd be very surprised if I changed my perspective on the basis of our conversation although I gotta say I think I did. I think I have more respect for stuff that I would have dismissed until I worked with those folks closely on the Task Force. (S. Hollon, personal communication, April 26, 2011)

Most Task Force members said that they agreed with the final documents generated by the Task Force. They identified several aspects of Task Force deliberations that facilitated consensus building and ultimate agreement on the documents produced.

Proceedings were kept diplomatic in nature as illustrated by the following statement by APA president Ronald Levant: "the brilliant beauty of the Task Force was that it focused on the gap you know between the two perspectives and the recognition that you know my truth and your truth while very meaningful to each of us respectively are only a small portion of the whole truth" (personal communication, June 27, 2011). Many Task Force members also commented that more philosophical topics, such as debates about the value of specific scientific methods, were avoided by the Task Force in the interest of promoting agreement between members:

> We purposely avoided that [philosophical discussion] because we were gonna quickly get to some irreconcilable differences you know and the different kinds of research evidence, you even see qualitative research there right? Because there were some people on the Task Force that said well qualitative is the only way research should be done, well you know once you take that position you're never going to come to an agreement

> with people like (names of particular Task Force members). I think so; I think we tried to avoid those things. (B. Wampold, personal communication, February 22, 2011)

Finally, Task Force members also commented that the Task Force purposefully kept its documents relatively general and broad to allow for agreement among members. David Barlow described the writing process of the Task Force as follows: "Many of the sentences that one person or another would narrate would be too specific to agree on so we had to back off and become more general in how we said it that it'd cover all" (personal communication, August 6, 2011). Task Force members conceded that the general nature of the Task Force documents has been received with some criticism by members of the psychological community, who thought that the generality of the guidelines detracted from their value in informing practice in concrete ways. Nonetheless, at the conclusion of deliberations, "the Task Force achieved unanimous consensus in support of its two primary work products, the proposed APA policy statement and the position paper" (APA, Council Item, unpublished document, August 2005).

Documents and Policies

Revisions of the Task Force's draft documents occurred between March and August of 2005. In March 2005, the documents were presented at the APA Consolidated Board and Committee meeting, sent out to presidents, president-elects, and executive directors of all APA divisions and state and provincial psychological associations, as well as the members of the APA board of representatives for review. Reviews from the public were also invited through publication of the documents' availability on the APA's website and in the March edition of the APA monitor (APA, Council Item, unpublished document, August 2005). Several hundred comments were submitted and reviewed by Task Force members through conference phone calls prior to the Task Force's final meeting in June 2005 at which revised versions of the documents received their final approval. The Council of Representatives approved the final documents at their meeting in August 2005.

In their final documents, the Task Force formally defined EBP as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA Presidential Task Force on Evidence-based Practice, 2006, p. 3). The Task Force also provided definitions of, and elaborations on, the various components of this statement: best available evidence, clinical expertise, and patient characteristics, culture, and preferences. They summarized the state of research in each of these areas and indicated directions for future research. First, they elaborated their conception of the best available evidence, pointing out that research expertise places psychologists in a unique position to be able to generate research evidence and examine existing treatments. They conceded that treatments must balance external and internal validity, and raised a number of considerations that apply when research data are applied to a particular setting. Finally, they also cautioned that the absence of investigations of a particular treatment or intervention does not warrant a judgment that the treatment is ineffective, but rather that further investigation is called for, and barriers to such investigation must be identified.

The Task Force also provided a more detailed discussion of what constitutes evidence in the context of this statement. Best research evidence, they argued, "refers to scientific results related to intervention strategies, assessment, clinical problems, and patient populations in laboratory and field settings as well as to clinically relevant results of basic research in psychology and related fields" (APA, 2006, p. 273). They endorsed the use of multiple types of evidence and evidence produced by various research designs and provided a list of various types of evidence and the types of questions each type of evidence is suited to address (e.g., clinical observation, qualitative research, systematic case studies, single-case experimental designs, public health and ethnographic research, process-outcome studies, effectiveness studies, randomized controlled trials (RCT's)/efficacy research, and meta-analysis). They also argued that evidence should be evaluated along two dimensions: treatment efficacy and treatment utility. With respect to efficacy, different types of evidence can be ranked for their

relative contributions based on a hierarchy established in the APA's (2002) "Criteria for Evaluating Treatment Guidelines."[20]

Secondly, the Task Force elaborated the meaning of clinical expertise. They cited research on the importance of expertise in clinical treatment and other performance outcomes. They also provided a list of competencies that are encompassed by clinical expertise including:

> (a) assessment, diagnostic judgment, systematic case formulation, and treatment planning; (b) clinical decision making, treatment implementation, and monitoring of patient progress; (c) interpersonal expertise; (d) continual self-reflection and acquisition of skills; (e) appropriate evaluation and use of research evidence in both basic and applied psychological science; (f) understanding the influence of individual and cultural differences on treatment; (g) seeking available resources (e.g., consultation, adjunctive or alternative services) as needed; and (h) having a cogent rationale for clinical strategies. (APA, 2006, p. 276)

The Task Force then described some of the avenues to expertise including training, theoretical understanding, experience, self-reflection, knowledge of research, and continuing education and development.

Finally, the Task Force's policy document provided a discussion of the importance of patient characteristics, culture, and preferences. In this section, the Task Force acknowledged that findings based on groups may not apply to all individuals and that individual characteristics may be relevant to the outcome of interventions. In particular, they conceded that psychologists are faced with difficulty when deciding whether research findings apply to individuals who differ from the individuals from whom the research findings were derived. They described the multitude of ways in which patients may differ on individual characteristics, culture, and personal preference, and pointed out the importance of considering such factors in EBP.

[20] Notably, there is substantial disagreement among APA Task Force members and the general public with respect to whether the list of evidence provided should be treated as a hierarchy from least to most credible.

As may be apparent from the above summary, the APA statement on EBP provides a set of very general guidelines rather than specific rules. The exact manner in which one should go about integrating the various components of EBP and their relative importance to psychological practice remain unspecified. In fact, nothing in any version of these statements suggests a hierarchical organization of the three facets of EBP. Despite this fact, however, proponents of EBP tend to privilege evidence over clinical expertise and a consideration of patient characteristics. Since the publication of this statement, the privileging of evidence in EBP has become a virtual imperative among applied psychologists who are expected to use evidence as the main determinant of clinical decision-making.

Responses to EBP in Psychology

In the decade since its entrance into psychological practice, the EBP paradigm has been widely discussed and examined. The debate surrounding EBP is well documented in numerous articles in academic journals and many edited volumes on the topic (e.g., Goodheart, Kazdin, & Sternberg, 2006; Norcross, Beutler, & Levant, 2005). Despite the intuitive appeal of the notion that practices should be supported by scientific evidence and the obvious utility of such a standard in preventing the use of potentially harmful practices, calls for EBP in psychology have not been uncontroversial. The majority of papers published on EBP have whole-heartedly and at times uncritically endorsed its adaptation. Nonetheless, criticisms or concerns about numerous facets of EBP have been raised, and can broadly be divided along a few main lines, namely practical, theoretical, and conceptual. Practical concerns entail issues such as the applicability of research evidence to applied clinical settings, the incremental value of evidence-based treatments over treatments without research-evidence support, and the applicability of research evidence to diverse individuals and groups. Theoretical critiques of the assumptions of EBP, including its reliance on empirical epistemology and modernist/realist ontology, also abound. Finally, a few conceptual examinations of EBP have called into question the nature of the evidence in EBP.

Trumpets and Fanfares: Endorsements of EBP

EBP was rather warmly received by the psychological community at large. Any discussion of responses to EBP would be remiss not to include an overview of some of the endorsements of EBP that have been well-aired in academic journals. In a more indirect way, one could argue that much of the endorsement of EBP occurred implicitly and quietly, through the speedy proliferation of applied articles with various permutations of titles such as "The evidence-based practice of…" or "Evidence-based practice in…." These articles, despite refraining from explicit participation in the EBP debate, simply bypass questions about the value of such practice by implicitly assuming its desirability. A full review of these articles and their implicit assumptions about the value of EBP is not feasible within the scope of this work.[21] Thus, this section will focus on the more vocal participants in the EBP debate, who explicitly and loudly voiced their appreciation and endorsement of EBP.

A number of endorsements of EBP occurred even before the formation of the APA Task Force and, in fact, may have prompted it. The earliest such endorsement by Sanderson (1998) raised concerns about the use of highly structured, empirically supported treatments (EST), but stressed the importance of grounding psychological practice in empirical evidence. Sanderson responded to a number of concerns about the reliance on evidence, including arguments that research may not generalize to applied settings, neglects therapist and client variability, shows that all therapy is effective, and tends to be limited to the treatment of DSM diagnoses. He argued that many of these concerns apply to the EST (Empirically Supported Treatment) movement, but not to practice that is merely grounded in evidence. He voiced a strong sentiment that reliance on evidence is essential if psychological intervention is to survive in the current health care environment.

Fonagy (1999) advanced a narrower vision of EBP based on the use of manualized treatments, a position more in line with EST. He explored reasons for clinicians' resistance to EBP from a psychodynamic perspective, equated it to a phobia,

[21] The use of the search term "evidence-based practice" on PsycInfo or Google Scholar yields a rather large sample of such applied articles.

and attributed it to erroneous cognitions. He proposed that the value of manualized treatment lies in the fact that it limits various iatrogenic factors in the therapeutic relationship. Around the same time, Hunsley (2000), on the other hand, advocated a broadening of EBP. He provided a Canadian perspective on EBP in his review of the report of the task force on Empirically Supported Treatments by the Clinical Psychology Section of the Canadian Psychological Association. He commended APA's work on EST and described CPA's endorsement of EST, but also commented on the extent to which the Canadian report on EST recognized the need to place EST in a broader context of practice based on the best available evidence in a manner that anticipates the later EBP movement.

Reynolds (2000) provided a detailed discussion of components of EBP and its applications. She argued that EBP could facilitate the translation of research into practice and cement psychology's role in the health care system. However, she also validated concerns that EBP may lead to a neglect of less investigated, and more innovative, treatments, and argued that conceptions of evidence must be broadened to include various kinds of evidence, including qualitative and case evidence from clinical practice. Barlow (2004) went as far as to suggest that the title "psychological treatments" should be used to describe only evidence-based treatments to separate them from the more general category of psychotherapies. Silverman (2005), another member of the APA Task Force on EBP, advocated the applicability of EBP to various theoretical orientations, including psychodynamic practice. He aimed to dispel the myths that equate EBP with manualized treatments or ESTs. He provided a brief preview of the stance of the Task Force, and expressed great hope for an EBP approach that integrates evidence, clinical expertise, and contextual and client factors.

The publication of the APA Task Force Statement on EBP also stirred up much less reserved enthusiasm among a number of psychologists. Brooke (2006), in his commentary on the Task Force, proclaimed that "something remarkable has happened" (p. 23) and expressed enthusiasm about his perception that this document constitutes a return to the recognition of the importance of clinical expertise and context on the part of APA. Hunsberger (2007) echoed this sentiment, and voiced his hope that EBP may bridge the rift between psychological research and practice that had been created through the neglect of subjective experience in psychotherapy. He viewed EBP as an

explicit endorsement of the centrality of clinical expertise and subjective experience in psychotherapy, and hoped that this endorsement might facilitate greater exchange of knowledge and experience between practitioners and researchers. Spring (2007) described EBP as a "trans-disciplinary, idiographic approach that promotes lifelong learning" (p. 611) and endorsed the extent to which it expands EST.

Soon after the publication of the Task Force Statement of EBP, however, unbridled enthusiasm was replaced by more moderated responses to criticisms of EBP. Hunsley (2007a) addressed a number of challenges to EBP, including the difficulty in translating nomothetic research into idiographic practice, the extent to which research participants are representative of clinical populations, the availability of adequate evidence to inform practice, and whether evidence-based treatments work in applied settings. Hunsley provided research evidence and suggestions to address some of these challenges. He concluded that despite the need for much additional clinical research, the research that is currently available supports the use of EBP. Similarly, Kazdin (2008), who also shared some of the concerns about EBP, expressed hope that EBP may provide a means to bridge the gap between researchers and practitioners while acknowledging their separate but equally valuable roles. Continued support for the use of EBP also is apparent in views such as those by Anderson and Cuijpers (2009) who endorsed Barlow's (2004) proposal to separate evidence-based psychological treatments from general psychotherapy in order to aid the dissemination of EBP and the integration of psychology into the healthcare system.

Practical Concerns about EBP

From the debates that preceded the publication of the Task Force report and the responses that followed, several general lines of criticism of EBP and its premises are evident. They focus on the limitations of existing evidence, the external validity of research evidence, the feasibility of EBP, and the theoretical assumptions of EBP.

The Evidence for EBP

Many commentaries on EBP lament the shortage of available evidence on a number of issues of clinical concern. In the early years of discussions of EBP, concerns were expressed about the implications of EBP for treatments that have not been evaluated. Shapiro (1996) stressed that "absence of efficacy evidence is not evidence for ineffectiveness" (p. 257). Reynolds (2000) voiced concerns that misperceptions of the sort Shapiro decried may quash the development of newer, innovative treatments and the practice of longer-term treatments that may be less amenable to quick empirical investigation or clear symptom reduction.

Similarly, a number of psychologists have called into question the extent to which sufficient evidence exists to inform most routine psychotherapeutic practice. Addis (2002) argued that evidence was lacking on a number of aspects of practice, including a lack of research on the effectiveness and process of treatments, on the utility of research products to practitioners in their present form, and on education in EBP at the undergraduate, graduate, and post-graduate level. King (1998) cautioned against overgeneralizations in interpretations of evidence in support of a particular treatment. He used the example of data from a large-scale study on the effectiveness of cognitive-behavioural therapy in the treatment of depression to demonstrate how research data, even in large-scale randomized-controlled trials, often require much more complex interpretations than mere dichotomous judgments of efficacy. He argued that present research allows for conclusions regarding the efficacy of a treatment "for very specific conditions under circumscribed conditions" (p. 87) and expressed doubt that, given the specificity of current research on even treatments that are considered to be well-established such as CBT for depression, there truly is enough research to allow for EBP and questioned whether the employment of such a term may not be a misleading rhetorical device.

Finally, many have argued that, aside from treatments which have not been investigated and treatments where research is limited to circumscribed populations and contexts, there are also some treatments and practices where research evidence is inconsistent or even contradictory (De Los Reyes & Kazdin, 2008; Westen & Bradley, 2005). In part, outcome data are contingent on and vary with the way in which outcomes

are measured. Subsequently, proponents of this criticism have argued that EBP may be difficult when research findings are varied and inconsistent. De Los Reyes and Kazdin (2008) suggested a more complex approach to the evaluation of evidence whereby evidential support is viewed as dimensional rather than categorical. Westen and Bradley (2005) argued that a more nuanced, multi-dimensional approach to EBP is needed.

Evidence in Clinical Settings

A number of discussions of EBP have focused on differences between treatment research and most treatment settings, and the extent to which these differences might affect the applicability of research to applied practice. In particular, treatment research tends to differ from applied treatment in terms of the way in which patients are selected or referred to treatment; the experience, skill, adherence, and caseloads of therapists; and, the kinds of interventions and treatments commonly employed (Bower, 2003). All these aspects tend to be more uniform or closely controlled in research than they realistically can be in applied settings. Many have thus questioned whether research evidence is likely to generalize to applied practice (e.g., Bower, 2003; Franklin et al., 2006; Tannenbaum, 2003; Westen & Bradley, 2005). Tannenbaum concluded, from a qualitative investigation of health professionals' views on EBP, that a number of barriers make it unlikely that EBP will be implemented in applied health settings and that the value of EBP lies primarily in its ideological and political potential rather than in its practical utility. Westen and Bradley (2005) reflected on the manner in which research treatments tend to be "brief treatments for discrete disorders" (p. 268) and argued that they differ in important ways (e.g., the settings in which they are delivered, the uniform populations that are treated in research, etc.) from the kinds of treatments commonly provided by practitioners.

Evidence and the Individuals in Treatment

A key question often raised with respect to EBP is the extent to which research evidence can be applied to particular clients. Most generally, Hayes, Kaoholokula, and Watkins (1999) pointed out that research usually provides nomothetic information, while individual treatment requires idiographic judgment. This, they argued, presents a major

challenge to clinicians, who are concerned with finding a treatment that works for a particular client rather than for clients on average. Hayes and colleagues (1999) stated that "the applicability of empirically supported, nomothetically based treatment to a client is a function of the degree of convergence between the causal relations relevant to the client's behaviour problem and the causal relations targeted by the treatment" (p. 456).

An additional question related to this issue is the extent to which EBP is useful in the treatment of diverse clients. Particularly, many have argued that much research is conducted with culturally homogenous patients and that little research considers the extent to which diversity in patients (e.g., in terms of symptoms, diagnoses, cultural factors) may impact the efficacy and applicability of treatments (e.g., Bernal, Jimenez-Chavey, & Rodriguez, 2009; Ingraham & Oka, 2006; Munoz & Mendelson, 2005; La Roche & Christopher, 2009; Stanley & Zoe, 2006). Some (e.g., La Roche & Christopher, 2009) have pointed out that EBP may be an improvement over EST in this respect in that it encourages the consideration of client factors in the application of evidence. Evidence does in fact suggest that cultural and contextual factors impact most aspects of treatment (Bernal et al., 2009) and the clinician who wants to implement EBP with a client from an understudied population is challenged by a scarcity of evidence on the manner in which treatment ought to be applied to such a client (Ingraham & Oka, 2006).

Obstacles to the Dissemination/Implementation of EBP

A large body of literature also addresses various challenges that may obstruct the dissemination and implementation of EBP. Factors that have been identified as such potential obstacles include clinicians' views and (mis)perceptions about EBP (Addis, Wade, & Hatgis, 1999; Lehman, 2010; Pagoto et al., 2007; Wolfe, 1999), difficulties in establishing training in EBP (Hunsley, 2007b; Pagoto et al., 2007), and organizational barriers (Rosenberg, 2010). Ultimately, this literature usually assumes that widespread dissemination of EBP would be desirable, but acknowledges the pragmatic factors that may make such dissemination challenging and, thus, may limit the utility of EBP. The same is true for most of the literature that deals with the practical concerns about EBP discussed above. Most of these authors did not advance arguments against EBP in theory but express concerns about the feasibility of EBP in clinical practice.

Theoretical Concerns about EBP

The hesitations about EBP reviewed above primarily focused on the feasibility of EBP and, at least implicitly, appear to be premised on the assumption that EBP would be desirable if it were possible to somehow evade the various practical matters seen as problematic.. Some responses to EBP, which can be broadly classified as theoretical in nature, have expressed concerns about the EBP paradigm more whole-heartedly by questioning the desirability, not just the utility, of such practice through a challenging of some of the theoretical assumptions on which EBP is premised. Most of the theoretical concerns, however, pertain to empirically supported treatments rather than EBP more broadly, and reflect the confusion of these two movements.

Criticisms of Empirical Evidence and Empirically-Supported Treatments

A number of psychologists (e.g., Addis & Waltz, 2002; Nathan, 2004; Wampold, Ollendick, & King, 2006) have outlined the assumptions of the empirically supported treatment movement and raised questions about the manner in which they might be adjudicated. Primarily, these responses concern EST's, rather than the broader EBP paradigm. Wampold and colleagues (2006) questioned whether empirically supported therapies outperform non-empirically supported therapies. They argued that often the latter are simply untested rather than refuted and that empirical evidence showing that empirically supported treatments are superior to these untested treatments is lacking. Addis and Waltz (2002) put forth similar arguments with respect to the utility of standardized or manualized empirically examined treatments. They pointed out that the use of such treatments is premised on the assumptions that research findings generalize to clinical settings, the use of these treatments improves treatment outcomes, and that these treatments are feasible and teachable. They call for evidence to be collected to examine these assumptions. This sentiment is echoed by Nathan (2004), who argued that proponents of empirically supported treatment must establish that efficacy studies are more valuable than effectiveness research, specific factors in treatments are more important than the common factors shared by all treatments, all treatments are not equal, and evidence is more important than clinical judgment. These critiques all rely on the more narrow EST model and have been responded to in some way or other by the

EBP movement, which advocates the use of various types of evidence and the consideration of contextual and individual factors as well as clinical judgment.

The Epistemology of EBP

A much smaller number of theoretical critiques focus on EBP and challenge the epistemological assumptions that underlie it. Some concerns have emerged out of evidence-based medicine and raise the question of whether medical (and, by extension, therapeutic) practice based solely on evidence is practical or possible. For instance, a paper by a number of Dutch bioethicists (Molewijk, Stiggelbout, Otten, Dupuis, & Kievit, 2003) provided a qualitative analysis of the presentation of medical facts in an evidence-based medical decision making project. The authors found it was impossible for them to present descriptive facts based on evidence in isolation from prescriptive, normative judgments as they had originally planned. They thus came to realize that many implicitly normative judgments are inevitably inherent in the use of descriptive factual evidence, and that a strict separation of description and prescription in medicine is impossible. Subsequently, they argued that medical practice cannot be grounded in descriptive evidence alone. Similarly, an orthopaedic surgeon (Michelson, 2004) pointed out that evidence-based medicine limits the physician to reliance on statistical evidence to the neglect of deductive reasoning (which he referred to as common sense) that can often account for causal relationships that have not been empirically established (e.g., the causal relationship between hitting oneself on the hand with a hammer and experiencing pain in one's hand). He argued that much of the available empirical evidence does not account for the complexity of human physiology and that sound medical practice requires deductive reasoning rather than sole reliance on empirical evidence. The same argument could, of course, be made for the complexity of human psychological functioning, which may be even more idiosyncratic than human physiology. Thus, this argument may certainly be applied critically to EBP in psychology.

Psychologists themselves have pointed critically to some of the epistemological and ontological commitments of EBP. In particular, EBP has been aligned with positivist and empiricist epistemology and realist ontology. The commitment of EBP to an empiricist epistemology is relatively explicit, given that a reliance on empirical evidence

is advocated. A commitment to empiricism itself, however, is a normative value-governed choice (Slife, Wiggins & Graham, 2005; Wendt & Slife, 2007) and, thus, a commitment to empiricism cannot be based on evidence. Slife, Wiggins and Graham (2005) and Wendt and Slife (2007) questioned the extent to which such a commitment to empiricism in EBP neglects alternate epistemological positions (e.g., post-modernism, pragmatism, hermeneutics) and methods that may be based on such epistemologies (e.g., qualitative research methods). They note that EBP uncritically assumes this epistemological stance without considering alternative options. Ramey and Grubb (2009) similarly pointed out the tension between the positivistic commitments of EBP and alternative conceptions of knowledge. EBP has as its aim the attainment of objective truth or knowledge about psychotherapy and thus neglects the possibility that truth in psychotherapy may not be entirely objective but rather co-constructed by clients and therapists in ways that reflect their experiences, interests, aims, and interactions.

Some of the members of the APA Task Force such as Wampold (2002) and Wampold, Goodheart, and Levant (2007) argued that the APA statement on EBP deliberately avoided the discussion of some of these philosophical issues by advocating a very open approach that relies on some type of evidence without unduly restricting the nature of such evidence or positing that evidence must be objective or used without exception. Therein, they argue, lies the strength of EBP, which must be distinguished from EST or any paradigms that advocate reliance on a particular type of evidence or the prescription of particular treatments for particular disorders. The only necessary component of EBP as advocated by the APA Task Force, according to these authors, is a reliance on some form of evidence, with the nature of this evidence left to be determined by the relevant scientific and professional communities.

Conclusions

Overall, then, the history of EBP can be summarized as follows. EBP emerged out of medicine. EBP arose in response to concerns that new methods developed to examine and compare medical treatments were having little impact on the practice of medicine, which continued to be informed by tradition rather than scientific evidence. Medical associations provided several articulations, statements, recommendations, and

definitions of evidence-based medicine. The most influential contribution of medicine to EBP in psychology was the definition of evidence-based medicine advanced in a statement by the Institute of Medicine in 2000. Various disciplines responded to medicine's call for EBP with their own statements.

Psychology's response to evidence-based medicine originally focused on practice guidelines in an attempt to assure that psychological treatments would not be dismissed in the face of guidelines regarding the use of pharmacological treatments put forth by medical organizations. Shortly thereafter, Division 12 of the APA also formed a task force, which lead to the birth of the empirically-supported treatment movement and generations of lists of treatments which were deemed to meet particular standards laid out for "empirically supported treatments." Misapplications of these lists by third-party payers and health management organizations prompted widespread concern among psychological practitioners. In response, the APA formed its own Task Force on Evidence-based Practice during the presidency of Dr. Ronald Levant. One year of deliberations by this task force culminated in the generation of a policy document and a position paper, which was unanimously approved by the Task Force and APA Council of Representatives.

Responses to the documents generated by the Task Force varied, and debates about EBP that had started before the formation of the Task Force have continued beyond its conclusion into the present day. Although widespread acceptance of EBP is evident in much of the published psychological literature, concerns about the feasibility of EBP, particularly the availability of necessary and externally valid evidence for clinical practice, have been voiced. Theoretical discussions of the assumptions of empirically-supported medicine and empirically-supported treatments are also well aired in the literature. Far fewer theoretical discussions of the assumptions of EBP exist. The following chapters of this work (Chapters 6 & 7) contribute to remedying this current limitation of the literature by providing a preliminary examination of debates and philosophical issues related to EBP in psychology.

Chapter 6:
Conceptual Issues in EBP

In 2006, the APA Presidential Task Force on Evidence-based Practice (EBP) by consensus of all of its members defined such practice as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture and preferences" (p. 273). A treatment of conceptualizations of 'evidence' and 'best evidence,' two concepts that lie at the heart of the EBP discourse, requires explication of all three dimensions of the conceptual framework described in the first chapter of this work: what is evidence, what it is evidence for, and what is the relationship between evidence and that for which it serves as evidence. The policy document drafted by the Task Force provides only limited guidance with respect to these three dimensions. In the present chapter, the three dimensions of 'evidence' as defined in the context of EBP are clarified by means of a qualitative analysis of references to 'evidence' in a large sample of journal articles on EBP and through interviews conducted with Task Force members Alas, consensus on any of the three dimensions was not found in the expressed views of Task Force members or in the implied, assumed, or expressed perspectives of authors of journal articles that dealt with EBP. Rather, a great diversity of perspectives was apparent on all three dimensions. In addition, the very definition of EBP, contrary to expectations, turned out to be a rather contentious issue, to some degree among members of the Task Force on EBP, but to a disconcertingly larger degree among the authors writing on the topic of EBP in academic journals.

The purpose of the present chapter is to highlight diverse perspectives on definitions of 'evidence', as represented in the EBP discourse. Chapter 7 will highlight inconsistencies and incompatibilities among conceptions of evidence in the EBP discourse, and explore ontological and epistemological issues related to them. No attempt will be made to resolve the conflicts between these various competing and contradictory conceptions. Instead, I focus on providing a survey of conceptions and

their implications for EBP in psychology. I hope to provide a starting point for contemplation, discussion, and, hopefully, eventual clarification of the nature of evidence in EBP in psychology.

Some Notes on Method

The evidence for many of the conclusions put forth in this chapter was derived from qualitative analysis (loosely based on a grounded theory method; Glaser & Strauss, 1967) of transcripts from interviews conducted with members of the APA President Task Force on EBP and journal articles on EBP published in major academic journals. I use the word 'evidence' in a guarded manner here, for it will soon become very clear that the qualitative data that informs this chapter would be disqualified from playing the role of evidence in psychological inquiry based on some of the views that will be outlined below, and would be considered merely weak evidence under alternative conceptualizations of evidence. Nonetheless, to provide a context for the claims made, a description of the methods used to arrive at these claims is warranted.

Interviews with members of the APA Task Force provided indication that they espoused different conceptions of 'evidence'. For the purpose of the present chapter, interview transcripts were examined for references to interviewee's conceptions of evidence, its objects, and the relationship between evidence and its objects. Such references were extracted and descriptively labeled (coded) using MAXQDA qualitative data analysis software (Verbi Software, 2013). In the second stage of analysis, findings from interview transcripts were then integrated with data from other documents reviewed for the purpose of this chapter. Documents generated by the Task Force were also coded for references to the Task Force's views on these three dimensions of 'evidence.'

The major sources of data for the present chapter were academic journal articles on EBP. Although the manner in which articles were located and analyzed was described earlier in this work (see Chapter 1), a more detailed description is warranted at this point. A sample of journal articles on EBP was collected from PsycINFO in January 2011 using a search restricted to peer-reviewed journal articles that contained the phrase "evidence-based" in the title (a similar search for the phrase "evidence

based" returned identical results). A total of 2,240 articles matched these search criteria. All articles were published between the years 2000 and 2010, consistent with the history of the emergence of EBP discussed in the previous chapter, and the majority of articles were published after the formation of the APA Task Force on EBP in 2006. Titles and abstracts for all 2,240 articles were reviewed and articles pertaining to EBP in clinical psychological assessment and intervention rather than other areas of psychology (e.g., education, industrial-organizational) were selected, resulting in a sample of 850 relevant articles. Of these articles, only 425 articles were available in an electronic format that permitted the planned analysis using qualitative analysis software. A total of 410 articles could be converted to text format for the planned analysis and 15 that could not be thus converted had to be eliminated from the sample. An additional 40 articles became unusable during the conversion to a format that permitted the required text-search in MAXQDA. A sample of 360 journal articles, therefore, constituted the final sample of articles analyzed.

Using the search function of MAXQDA, I located a total of 2,728 paragraphs that contained references to 'evidence' in the article sample and extracted them for further analysis. I then reviewed these 2,728 paragraphs and descriptively labeled and highlighted particular conceptions of evidence, its object, the relationship between evidence and its objects or other conceptions that appeared noteworthy and relevant to the current project contained within them. In a third round of analysis, I grouped labeled conceptions of 'evidence' into broader categories based on the perspectives on evidence they represented (e.g., more objective vs. more subjective conceptions of evidence) based on prior knowledge of broad divisions in accounts of evidence in epistemology and philosophy of science. Within each category, I then selected excerpts that represented or exemplified the larger category and extracted them verbatim into a Microsoft Word document, to generate a document that contained sample quotations representing various categories of conceptualizations of 'evidence.' In a final step, I divided categories of conceptualizations of 'evidence' into four broad divisions, based on whether they pertained to evidence, its object, or the relationship between evidence and its objects, or other relevant topics. At that point, I also grouped references and quotations from interviews and Task Force documents into the existing categories of conceptualizations to allow for their inclusion in the final write-up and facilitate a

comparison of official documents, Task Force members' perspectives, and journal articles. The resulting document contained representative paragraphs from various articles on EBP along with paragraphs from interview transcripts and official documents and was the basis of the summary of findings provided below

The Ontology of Evidence in EBP

With respect to views on the nature of evidence, a vast number of perspectives seem to be represented in the EBP discourse. Especially in the EBP journal articles, perspectives on the nature of evidence are mostly assumed and implied rather than described explicitly. Few authors provide an explicit definition of what they mean by 'evidence.' Rather, they speak about evidence with the assumption that something they present can constitute evidence for something else. 'Evidence' seems to play multiple, and at times contradictory, roles in one and the same article. Often, very different types of things are treated as one and the same in the sense that they are all treated as constituting evidence. Differences in perspectives on the nature of evidence concern both what is taken to constitute evidence and beliefs as to where evidence comes from or how it is produced.

In its official report and policy documents, the APA Task Force defined 'evidence' as follows:

> Best research evidence refers to scientific results related to intervention strategies, assessment, clinical problems, and patient populations in laboratory and field settings as well as to clinically relevant results of basic research in psychology and related fields. A sizeable body of evidence drawn from a variety of research designs and methodologies attests to the effectiveness of psychological practices. (2006, p. 273)

Based on this definition, 'evidence,' in the context of EBP, is defined much more narrowly than in general parlance, and becomes something of a technical concept. This definition betrays, potentially, a few assumptions about evidence that are relatively consistent with the remainder of the Task Force documents: (a) scientific results or the results of scientific studies are the evidence to be used in EBP, (b) evidence can be drawn from a variety of research designs, and (c) there are bodies of evidence. A

review of the EBP literature highlights some of these same overlapping assumptions but also many additional assumptions and conceptual controversies with respect to the nature of evidence and reveals perspectives that are at times consistent with the Task Force's position but that also sometimes deviate from the definition of evidence advanced by the Task Force. These various assumptions and controversies will be outlined in the remainder of this chapter.

The Controversy of the APA Task Force's "Three-legged Stool"

The APA Task Force, in its definition of EBP, stipulated that EBP involves the integration of best evidence, clinical expertise, and patient characteristics/culture/preferences. The three components of this definition have commonly come to be referred to as the "three-legged stool" of EBP. The Task Force, however, also goes on to define best evidence as the results of scientific studies. At least implicitly, this definition seems to make a conceptual distinction between the evidence to be used in EBP and the two remaining legs of the stool—clinical expertise and patient characteristics/culture/preferences—with which the evidence is to be integrated. At the same time, however, the Task Force document also implies that all three legs are to serve as the basis or justification of clinical decision-making. In fact, in a review of the EBP literature, both interpretations are represented and some contention between these perspectives is apparent.

Some proponents of EBP explicitly argue that all three legs of EBP ought to be conceptualized as sources of evidence. For example, Wilson and colleagues (2009), in their study of psychologist's attitudes towards EBP, write "many of the participants initially believed that EBPP[22] was the same as ESTs...However, once they learned that their clinical expertise and the context of the client were also considered sources of evidence equivalent to research, they had more positive reactions" (p. 407). Similarly, in a discussion pre-dating the Task Force that focused on the framework of evidence-based medicine and its application to psychology, Tannenbaum (2003) placed clinical

[22] The abbreviation EBPP for Evidence-based psychological practice is at times employed instead of the briefer abbreviation EBP in the EBP literature.

expertise at the bottom of medicine's hierarchy of evidence, implying that it is in fact a form of evidence if a less well-respected one.

In addition, a number of authors discuss the role of clinical experience and expertise in a manner that explicitly or implicitly implies that it is a form of evidence to be considered in EBP (e.g., Milton, 2002; Pollio & MacGowan, 2010; Rohricht, 2009).[23] To provide just a few examples, claims such as the following are common in the EBP literature.

> In everyday clinical practice, in medicine as well as in psychotherapy, wise practitioners draw from their own experience, intuition, commonsense, perhaps from available published evidence, and from their immediate sense data—the client in front of them. They see what is evident (*videre*) and act on it. (Feltham, 2005, p. 136)

Evidence in this article is defined using its original etymological meaning (see Chapter 2), and clinical experience is implicitly included among the evidence to be used by clinicians. Some more explicit claims are also common in the literature. "For this pilot-project, evidence regarding need for a group and type of intervention focus came not only from publications, but also from consultations with practitioners and from the researcher's own interview findings" (Home & Biggs, 2005, p. 46). The idea of "practice-based" evidence is also often promoted in a manner that implies that such evidence is part of clinical expertise.

> For many researchers the meaning of evidence aligns with the traditional notions often associated with randomized clinical trials. Many now reject this type of evidence as ''sterile'' and unrealistic, promoting instead what some now call ''practice-based evidence,'' or information that comes from actual clients in real clinical settings. (Sexton & Kelly, 2010, p. 81)

In addition, clinical expertise is at times rebranded as a form of evidence through the use of terms such as 'experiential evidence.'

[23] For many of the arguments made in this chapter, a multitude of examples was found within the large number of articles reviewed. In cases where there were numerous articles that could be cited as examples of a certain view, a few examples were selected and cited with the understanding that many additional articles that espouse the same view exist.

> Experiential evidence encompasses the knowledge gained from the direct care of patients. The practising doctor may rely on personal experience or attempt to learn from the personal experience of others. With experiential knowledge, more is generally considered better than less. Hence, expert opinion, when based on extensive experience with large numbers of patients with a particular disease, may be viewed as the highest form of experiential evidence. (Tonelli, 2006, p. 252)

In some articles on EBP, clinical expertise is thus conceptualized as part of the evidence to be considered in clinical decision-making. In addition, some Task Force members endorsed (during their interviews) the view that clinical expertise is to be conceptualized as part of the evidence for EBP.[24]

> DB: …You know there were some real abuses in the early days of EBP, restricting the practitioners' ability to determine what's best for the patient. But I think some of these concerns have been allayed and there's a realization that these principles can be incorporated into practice, but there has to be you know obviously good clinical expertise in how best to implement them, that's all part of the evidence.
>
> NL: So you view the clinical judgment and all those things as part of the evidence?
>
> DB: Well clinical expertise is very much part of the evidence!
>
> (D. Barlow, personal communication, August 6, 2011)

In at least some of the EBP discourse, clinical expertise is therefore considered to be part of the evidence on which to base clinical decision-making.

Similarly, but to a lesser extent, a small number of authors also conceptualize the client's attributes and preferences as part of the evidence to be used in EBP. For example, Wifstad (2008), commenting on evidenced-based medicine (EBM), argued the following:

> The doctor has to get to know the patient and the situation he is in. This is not the kind of knowledge we can get from randomized, controlled clinical trials. But still it is knowledge. And it is because the doctor has this kind of knowledge she is able to judge how likely it is that the patient in front of her will benefit from a chosen treatment. But from the

[24] Recall from the previous chapter that five Task Force members agreed to be interviewed and only their views will therefore be represented in this chapter.

> perspective of EBM, it might be difficult to remember that this kind of knowledge based on internal "uncontrolled" evidence has to be presupposed as a bridge between group data and the individual case. (p. 139)

At times, the third leg of EBP is referred to as "internal evidence" as in the following quotation:

> …even though we are unable to decide with certainty how susceptible the individual patient is, we can make a judgment. And this judgment is more than a guess. It is based on knowledge, but not the kind of knowledge we get through randomized trials. In contrast with the external evidence of such trials, we can call this knowledge "internal evidence. (p. 138)

In fact, client feedback also has been conceptualized as a form of empirical evidence.

> It is important to acknowledge that EBTs (Evidence-based Treatments) are not the only manner in which psychotherapists can incorporate empirical evidence into practice. For example, EBP could involve utilizing such strategies as client feedback to clinicians regarding progress, which has consistently been found to enhance outcomes. (Spielmans, Gaitman & McFall, 2010, p. 234)

Some authors advocated for the inclusion of both clinical expertise and client preferences. For example, Edwards (2007) argued:

> Episodes within cases may provide evidence that is just as important as that gleaned from multivariate group comparison research. To call such evidence "anecdotal", with its connotation of amusing and perhaps far-fetched tales over dinner is not helpful to science. It is effectively a rhetorical device designed to exclude clinical experience or case-based evidence from scientific debate, under the guise of a scientific principle. (p.11)

Both clinical expertise and patient preferences are thus conceptualized as part of the evidence in EBP in at least some of the published literature on EBP.

In contrast, at least some of the Task Force members in their interviews explicitly endorsed the view that best evidence is evidence produced by research and that the remaining aspects of the three-legged stool, albeit possibly useful, are separate from the

evidence to be considered in EBP. For example, S. Hollon (personal communication, April 26, 2011) stated:

> Well I think we had a lot of time and attention paid to the second leg which I think is alright but I think it doesn't really add up to much and, but it kept my colleagues happy and I think that we paid attention as well to the third leg, the patient preference, which I think you can summarize with a single statement. You know, patient preferences trump whether the evidence is, whether what they prefer is consistent with the evidence or not. In my mind it's, the first leg is the big leg and then the second one is the frosting on the cake and the third one is the, nothing more than the recognition that people have a right to control what happens to them.

He also explicitly stated that only the best evidence is directly relevant in establishing whether a treatment works and that clinical expertise and client preferences and that evidence is "controlled evidence" which takes primacy over other elements of the Task Force statement.

Similarly, in the published literature on EBP (e.g., Gray, 2002; Lewis, 2009; Levant & Hasan, 2008; Shlonsky & Gibbs, 2004) many writers clearly distinguish between best evidence and the remaining two legs or EBP, arguing or implying that only the former constitutes evidence. For example, Shlonsky and Gibbs (2004) argued, "in the absence of relevant evidence, the other two elements are weighted more heavily, whereas in the presence of overwhelming evidence the best-evidence component might be weighted more heavily" (p. 138). They clearly implied that neither clinical expertise nor patient preferences constitute "relevant evidence." Levant and Hassan (2008) also distinguished between evidence and "the other two pillars of what the Institute of Medicine (2001) has defined as the foundation of EBP in health care, namely, clinical expertise and patient values" (p. 658). Lewis similarly argued that "opinion based on clinical experience is an insufficient basis for practice, positive evidence trumps lack of evidence, and students should learn to base treatment on the highest empirical ground possible" (p. 108). In a more nuanced view, Tonelli (2006) advocated the inclusion of clinical expertise as a kind of evidence, which he calls experiential evidence, but also opposed the treatment of experiential evidence and empirical evidence as one and the same.

> Both the decision to include clinical experience and physiologic understanding under the umbrella of 'evidence' and the subsequent assignment of these kinds of knowledge to the lower rungs of the hierarchy of evidence are, of course, not themselves evidence-based, but rather represent philosophical (more precisely, epistemological) assertions. Both assertions, I will argue, are incorrect. (p. 250)

The multiple uses of the term 'evidence' in the EBP literature is therefore highlighted by the fact that Tonelli both espouses clinical expertise as a kind of 'evidence' and simultaneously contests the inclusion of clinical experience "under the umbrella of evidence."

To summarize, there is substantial ambiguity with respect to the use and definition of the concept of 'evidence' in the EBP literature. Some define 'evidence' as the outcome of research (e.g., Shlonsky & Gibbs, 2004). Other writers conceptualize research, clinical expertise, and patient preference as different kinds of evidence (e.g., Tonelli, 2006). Finally, some view research evidence, clinical expertise, and patient preferences as all part of the same overarching concept of 'evidence' to be employed in EBP. Whether one conceptualizes evidence as merely research findings or as the incorporation of the three different struts of the three-legged stool has substantial implications for what can and cannot be considered EBP and the kinds of roles that evidence can play in it.

What is Research Evidence?

Unfortunately, even if one puts aside the idea of the three-legged stool and stipulates that 'evidence' in EBP is defined as research results or scientific results, many questions remain. What exactly constitutes a research finding or a scientific result? Are research findings the data generated by such research, or the communicated findings in the form of research literature and journal articles? Neither the Task Force's documents nor the EBP literature provide very clear answers to these questions. In fact, based on a review of the EBP literature, it would seem that evidence consists of any or all of these components of the research process: the research design, the research data, and the communicated research findings in the form of academic literature. It is difficult to make sense of how evidence can come from designs in the form of data, and literature, all at

the same time. The following paragraphs will outline each of these three conceptualizations.

Research as evidence. Whatever research evidence is, there appears to be at least some agreement in most of the EBP literature that it is derived from research studies employing certain kinds of research designs. Research designs that are commonly referred to as a source of evidence include randomized-controlled trials, case studies, single-case designs, and qualitative designs. In addition, many authors argue that all of these various kinds of designs can provide evidence in the context of EBP.

Randomized-controlled trials. The research design that is most commonly associated with EBP and very frequently mentioned in the EBP literature is the randomized-controlled trial (e.g., Falkum, 2008; Patterson et al., 2004; Sanderson, 1998). Most commonly, randomized-controlled trials are treated as a source of a certain kind of evidence that is valuable to EBP, such as evidence of causal relationships:

> For a particular intervention to be judged efficacious, level 4 causal evidence is required. The question to be answered here is whether the intervention itself has caused the outcome. Ideally, a study at this level also reveals the elements are responsible for attainment of the desired outcome with a particular population of clients should. There are two research designs that can produce this level of evidence. The RCT is usually considered the most powerful design, as the internal validity of the results is maximized (i.e., the results lead to a high degree of certainty that the intervention is responsible for the observed effects). (Veerman et al., 2007, p. 218)

Many articles on EBP imply that randomized-controlled trials are a superior source of evidence to other research designs or, in fact, as previously mentioned, are the only useful source of evidence for EBP. For example, Michie and Abraham (2004) argued that "Behavior change only indicates intervention success when naturally occurring change is controlled. Therefore, non-randomised studies using 'before and after' designs provide weak evidence of change" (p. 30). Even more strongly, Baz (2007) argued that non-randomized studies do not provide reliable evidence, "Most of the studies they found were uncontrolled 'before–after' studies and no randomized controlled studies were found. The authors came to the conclusion that 'these reports do not provide reliable evidence and publication bias is highly likely'" (p. 920).

Interestingly, in some discussions, randomized-controlled trials are described in a manner that appears to imply that they themselves in fact *are* evidence, rather than a source of evidence. For example, in a discussion of limitations of randomized-controlled trials, Tucker and Roth (2006) conclude that "over-reliance on the RCT as evidence risks driving the evidence base in the direction of treatments that are relatively easy to study using this method, in particular the use of medication" (p.924). Implicit in the language of other authors is the belief that the research designs themselves are the evidence; it is not entirely clear, however, how a research design can be evidence. A charitable interpretation may be that such phrases reflect momentary inattention in the use of language, rather than actual conceptual confusion. However, in light of an additional area of confusion between EBP and the empirically supported treatment movement that will be outlined at the end of this chapter, such use of language may warrant greater attention. Specifically, throughout the EBP literature, many authors commonly refer to criteria whereby one can identify evidence-based treatments[25] and posit that the existence of a certain number of randomized-controlled trials supporting a treatment approach makes that treatment approach evidence-based (e.g., Ginsburg, 2006; Keel & Haedt, 2008; Silverman et al., 2008).[26] Based on such conceptions, an alternate interpretation may be that that the use of a certain design in a scientific study itself provides evidence for the value of the study or the legitimacy of its findings. For example, if randomized-controlled trials are thought of as superior means of establishing evidence for a certain purpose, the use of a randomized-controlled trial itself could be provided as evidence for the value of the findings generated over and above the value of findings generated by other less-esteemed methods. Similarly, if the evidence-base of a treatment approach is a dichotomous matter based on the existence of a certain number of randomized-controlled trials, the mere existence of such trials supporting an approach is evidence for the fact that the approach is evidence-based. It is unclear whether these authors intended such an interpretation. If so, however, this interpretation has important implications for the manner in which EBP is conducted. If randomized-controlled trials or

[25] Whether treatments can be evidence-based and if so, what part of the treatment the evidence pertains to (e.g., its efficacy, a particular treatment method, etc.) is an additional issue that will be further discussed later in this chapter.

[26] This issue and the theoretical questions attached to it will be further discussed later in this chapter.

other research designs provide evidence, a clinician would review these trials and contemplate the evidence generated by them as a guide to decision making. If, on the other hand, the mere existence of such trials provides evidence for the utility of a treatment approach, a simple count of existing randomized-controlled trials that support a certain treatment approach may suffice as a basis for a decision to utilize said approach. This issue will be discussed further in the final chapter (Chapter 8) of this work.

Other designs. In response to the espoused supremacy of the randomized-controlled trial, many authors have put forth and advocated the use of various other research designs as sources of evidence in EBP. For example, the role of case studies in EBP is often highlighted (e.g., Cooper et al., 2008; Richards & Worthington, 2010; Tannenbaum, 2003). In this context, writers (e.g., Messer, 2004) describe case studies as sources of evidence: "HSCED (Hermeneutic single-case efficacy design) attempts to use quantitative and qualitative information to create a rich case record, which provides direct and indirect evidence for the causal influence of therapy on client outcome" (p. 586). Many (e.g., Edwards. 2007) also advocate for the greater use of case studies as sources of evidence: "There is still a need for advocacy in the promotion of case study research because there has been insufficient appreciation of its role as a source of evidence relevant to the development and evaluation of practice in psychotherapy" (p. 6). Other types of research designs that are frequently described as sources of evidence in the EBP literature include qualitative designs.

> An expanded view of evidence would most certainly include qualitative research. Qualitative research has a long tradition within psychology, dating back to its introspectionist roots. More important (sic), qualitative methods have allowed psychologists to give voice to underrepresented individuals and groups and identify psychological phenomena and psychotherapeutic processes that would not otherwise be available. (Chwalisz, 2003, p. 504)

Many authors (e.g., Edwards et al., 2004; Molloy, 2007; Veerman et al., 2007) and some of the Task Force members interviewed (e.g., D. Barlow, personal communication, August 6, 2011) also address the value of series of single-case designs. Often, it is argued that the evidence produced by such designs is as convincing as the evidence of a randomized-controlled trial. For example, "When carefully designed,

repeated single case studies can provide evidence that is just as convincing as the evidence from RCTs" (Veerman et al., 2007, p. 218).

Multiple designs. The most commonly espoused views of research designs found in the EBP literature in the present review, however, emphasize the importance and role of multiple and varied designs as sources of evidence (e.g., Rohricht, 2009; Spielmans, Gaitlin & McFall, 2010; Westen et al., 2005). For example, Nelson and Nelson (2010) argue, "best available research can include evidence from a diverse set of study designs (ranging from randomized clinical trials to systematic case studies) and provides the scientific foundation for clinical practice" (p. 305). Similarly, La Roche and Christopher (2009) comment on the APA Task Force as follows:

> The Presidential Task Force noted that the development of psychological treatments is a complex process that requires clinical and research attention to multiple interacting sources of evidence. This can include evidence gleaned from all types of scientific studies ranging from RCTs to clinical observation and qualitative research. This broader set of guidelines allows findings from various types of studies, including emic and etic, to be considered relevant in the development of intervention strategies. Multiple research designs can provide a more complex and complete portrayal of reality than research from only one source and can be used to address different types of questions. (p. 400)

Many of the accounts that espouse the value of various research designs in generating evidence also stress that different designs can answer different types of research questions and produce evidence for different purposes (e.g., Falzon et al., 2010).

Overall, then, most accounts of evidence in the EBP literature emphasize that randomized-controlled trials, case-studies, qualitative designs, or series of single-case designs can produce evidence, or, most commonly, that all of the above designs are legitimate sources of evidence. Whatever evidence is, it must therefore be something that arises from studies using certain research designs. Two potential candidates for evidence are therefore data generated by research designs, and research literature that communicates research findings.

Data as evidence. A good portion of the literature on EBP seems to imply that the evidence to be used in EBP are data (i.e., functions of quantitative measurements), which are collected or generated in research (e.g., Goldenberg, 2006; Joinier et al.,

2005; McCabe, 2006). In particular, articles reviewed in the present study conceptualized data and statistics (including effect sizes, standardized measured data, cost data, completion rates, and outcome data) as evidence. According to such accounts, evidence can be represented quantitatively, summarized and synthesized in meta-analyses, and presented in graphs and tables. For example, Goldenberg (2006) writes, "Evidence as accumulated data has been made widely and easily available to clinicians and educators by evolving information technologies, and EBM aficionados, such as those found in the Cochrane Collaboration" (p. 2621). McCabe (2006) posits, "knowing how and where to locate relevant research data is obviously a pre- requisite for clinical work based on current evidence" (p.54). Of course, in this example, again careless use of language may be at play and the authors may have intended to refer to accumulated findings rather than data. However, some authors implicitly assign attributes to evidence that presume that evidence is quantitative; for example, Hunsley and Mash (2005) refer to "psychometric evidence," a common reference in literature that deals with evidence-based assessment. It is difficult to conceive of a way that evidence could be anything other than quantitative or statistical if it is to be psychometric (i.e., concerned with measurement). Often, it is also stipulated that EBP requires the use of data measured in a standardized, replicable manner, For example, "Clinicians often ask their patients to subjectively determine if they are getting or feeling better, but EBP requires them to use a standardized metric to determine treatment outcome" (Singh & Oswald, 2004, p. 133). Again, the use of a standardized metric as a prerequisite for something to count as evidence implies the belief that evidence is necessarily quantitative in nature.

Some specific types of data are also conceptualized as evidence in the EBP literature. For example, Veerman and colleagues stipulate that outcome data can be used as evidence for the effectiveness of a treatment: "Pre-post studies that monitor the outcomes of treatment are also used to gather evidence of effectiveness at this level. Goal attainment data provide essential information for empirically establishing the effectiveness of interventions" (2007, p. 217). Other types of data that have been conceptualized as evidence include completion rates (e.g., Andrews & Titov, 2009) and cost data (Singh & Oswald, 2004). The statistics that are most commonly conceptualized as evidence in the EBP literature are effect size estimates (e.g., Gorman

& Conde, 2010; Veerman et al., 20078; Weisz et al., 2008). Gorman and Conde (2010) argue that in EBP "isolated statistically significant effects provide sufficient evidence for inclusion on the major lists of evidence-based drug prevention programs and practices, since their primary outcome criterion is typically just a single effect on one behavioral outcome from one evaluation" (p. 220). Often times, effect sizes are presented as the results of analyzing data obtained from randomized-controlled trials, and are taken to be the preferred format in which evidence can be presented or represented: "Most of the evidence presented in this article is based on RCTs and summarized in terms of effect size (ES), an index of the magnitude and direction of treatment impact" (Weisz et al., 2008, p. 54).

Despite the fact that research data are treated as evidence by some writers in the EBP discourse, many proponents of EBP also argue that data from one study alone cannot provide adequate evidence to support a particular decision. Rather, they argue that only meta-analyses of data (particularly the effect size estimates) from multiple studies provide adequate evidence for EBP. There is substantial diversity (and sometimes confusion) with respect to the way in which the relationship between meta-analysis and evidence is conceptualized. First of all, some authors argue that meta-analysis itself constitutes a form of evidence, and indeed the only legitimate type of evidence to be considered in clinical decision making.

> We have mentioned the need for studies into the outcome of different psychotherapeutic interventions and what they are good for. These begin to form a bridge from practice to evidence. However, individual studies count for little, unless their methods have been so rigorous that their results are beyond question. In practice, reliable evidence that a treatment is effective (or harmful) depends upon the pooling of results from several studies. There are well-established methods by which this can be done, which in practice involve the rejection of findings from more studies than those they include. This process leads to 'systematic reviews' that, together with research reports, constitute 'evidence.' (Mace, 2006, p. 308)

Contrary to some of the accounts summarized above which posit that data are evidence, these accounts suggest that effect sizes and other statistics from single studies only constitute evidence once they are pooled in a meta-analysis. Other accounts, however suggest that meta-analyses generate or "provide" evidence rather than constituting

evidence: “Evidence-based medicine is underpinned by this paradigm. In its purest form, it raises experimental evidence to rule over other forms of knowledge. Certain study designs, such as meta-analyses and randomized trials, are thought to be less vulnerable to bias and, therefore, provide ‘superior’ evidence” (Broekart et al., 2010, p. 230). Those who conceptualize meta-analyses as constituting evidence or providing evidence would disagree with those who maintain that results from single studies can provide evidence, a divide that will be taken up again later on in this chapter, during the discussion of conceptions of the relationship between evidence and its object.

On the other hand, meta-analyses are also commonly conceptualized as “summaries of evidence” or “bodies of evidence” in a manner that is more consistent with the conceptualization of effect sizes and data from single studies as evidence. For example, Weisz and colleagues (2008) propose that:

> ES (Effect Size) values can be averaged across multiple outcome measures within a single study, to reach conclusions about the mean impact of a tested treatment within that study. Such study mean effects can, in turn, be averaged across multiple studies to generate a picture of the mean ES for a body of evidence — say, on treatment of depression, or even treatment across multiple conditions. When ES values are averaged across studies in this way, the process is called meta-analysis. (p. 55)

At times, in such accounts, meta-analyses is described in a manner that appears akin to a kind of statistical alchemy (a phrase borrowed from McDonald & Viehbeck, 2007) in the sense that it appears not only to summarize the evidence but also augments it in some manner:

> For most clinicians, it also requires ‘systematic reviews’ – syntheses of what qualified reviewers consider the best evidence. Reviews may include meta-analysis, a complex and controversial statistical technique by which the data from different studies are combined to strengthen the evidence from any one RCT. (Tannenbaum, 2003, p. 288)

Meta-analyses in these accounts are thus ways to both summarize and strengthen or improve upon existing evidence from single studies.

Lastly, some authors treat meta-analysis both *as* evidence and as a summary of evidence.

> The centerpiece of this new literature is a body of evidence called *systematic reviews*. Most psychologists are familiar with traditional reviews where authors search the literature and generate conclusions based on their findings. Unlike these traditional reviews, systematic reviews follow highly specific, predetermined methods for capturing the evidence, appraising it, and synthesizing it in a manner that is easily accessible to clinicians. (Walker & London, 2007, p. 637)

Meta-analysis is thus called upon to serve a variety of functions in the EBP context. It both constitutes evidence and summarizes evidence, and also strengthens or improves upon evidence. Such diversity in conceptualizations raises questions about the extent to which one statistical method can in fact serve such a multitude of functions. These questions will be discussed further in the following chapter. Suffice it to say, as a general conclusion here, that evidence according to some EBP literature is provided in the form of data, statistics (particularly effect sizes), and syntheses of statistics (in the form of meta-analyses).

Literature as evidence. According to a final view that is commonly espoused in the EBP discourse, the findings communicated in research literature are conceptualized as evidence (e.g., Sturmey, 2009; Upadhyay, Cockrill & Freedman, 2010; Singh & Oswald, 2004). An implication of this is that some authors lament the lack of training that clinicians have in "locating" evidence: "Unfortunately, many practitioners lack cofidence in being able to search for evidence using electronic databases such as Medline, PsycLit and Cinahl, and have not acquired skills in the critical appraisal of research" (Ramchandani, Joughin & Zwi, 2001, p. 61). Other authors laud the role of technology in allowing clinicians to access evidence:

> Evidence based practice provides these advantages because it has harnessed the powerful tool of information technology. In this context the power of information technology lies in the fact that research evidence can be disseminated beyond hospital and university libraries directly to clinicians at their desks or in their clinics. Electronic communication makes it possible to link up with libraries, journals and research institutions via the Internet. Literature searches can now be performed without leaving the clinic, and selected information about high quality re-

> search can now be directly accessed on CD- ROM and on the World Wide Web. (Reynolds, 2000, p. 259)

References to the publication of evidence or even a requirement that evidence be published are also not uncommon: “Evidence of this positive outcome needs to have been published in a peer-reviewed journal or documented in a comprehensive evaluation report” (Gorman & Conde, 2010, p. 220).

Although one could argue that in the above references the evidence is thought to be contained in the published research, and thus may not consist of the research literature but rather is communicated within it (e.g., in the form of data or statistics), such an interpretation is inconsistent with the fact that there are quite frequent references within the EBP literature to evidence as being contained in databases, and usually in that context these are databases of research findings rather than databases of data or statistics (e.g., McDonald & Viehbeck, 2007; Patterson et al., 2004; Spring, 2007). For example, Spring (2007) describes such databases as follows: “These tools, based upon continually updated evidence reviews, offer pithy evidence synopses, clinical practice guidelines, and structured abstracts. Examples are UpToDate, MDConsult, ACP Journal Club, *BMJ* Clinical Evidence, InfoPOEMS, and Clineguide” (p. 618). All of the databases mentioned here provide summaries of published research. McCabe (2006) makes additional claims about the value of one such research database:

> Clinical Evidence is a monthly, updated directory of evidence on the effects of common clinical interventions published by the BMJ Publishing Group. It provides a concise account of the current state of knowledge, ignorance, and uncertainty about the prevention and treatment of a wide range of clinical conditions, including mental health, based on thorough searches of the literature. It deliberately does not make recommendations but summarizes the best available evidence. (McCabe, 2006, p. 59)

According to McCabe, then, research summaries provide a fairly high degree of epistemic certainty. In the EBP literature, the ability to locate research findings and read and review research literature is emphasized and highly prized. At least some writers seem to conceptualize findings communicated in research literature itself as the evidence to be utilized in EBP. Findings, compared to research data, usually involve an interpretation of data and statistics and their relevance with respect to a certain theory or

hypothesis. 'Evidence,' conceptualized as findings communicated in the literature is therefore the result of a more inferential process that may make such inference more subjective or contingent on the researchers themselves relative to objectively measured data. Whether evidence is indeed better conceptualized as objective or subjective may therefore be closely tied to whether it is conceptualized as objective data or findings based on data, an issue that will be further discussed below and in Chapter 7.

Summary. Based on the foregoing considerations, research data, research literature, and research designs are at times conceptualized as evidence in the EBP literature. Whether evidence is conceptualized as quantitative data, or the findings reported in published written articles may make a difference with respect to the attributes that can be credited to evidence and the roles that evidence can meaningfully be said to play in the EBP discourse. A discussion of these issues will follow, in the next chapter of this work.

Physical versus Propositional Evidence

Irrespective of whether any or all parts of the three-legged stool are seen to count as evidence, several additional issues with respect to the nature of evidence can also be raised. More specifically, borrowing a distinction from philosophy, some authors of the EBP literature appear to characterize evidence in physical terms, implying that it is material, and capable of being observed (similar to fingerprints or DNA in a criminal case) while others tend to portray it as propositional, implying that it consists of knowledge and supports particular claims or theories. In physical accounts, evidence is commonly referred to as "a body of evidence" (e.g., Weisz & Simpson Gray, 2008; Zazalli et al., 2008), and conceptualized as something that exists to be found in the world (e.g., Patterson, Miller, Carnes & Wilson, 2004; Walker & London, 2007). In contrast, propositional accounts conceptualize evidence as information, knowledge, or theories produced in a particular social context with the aim of establishing facts (just like in the testimony of a court case). These accounts tend to focus more on the relationship between evidence and that which it is evidence for, and emphasize the processes of reasoning, evaluation, and generation of facts whereby information becomes evidence (e.g., Pollio & Macgowan, 2010; Wifstad, 2008).

Physical entities as evidence. Certain metaphors used within the EBP discourse imply that evidence is conceptualized as physical or object-like in nature. One of the most common of these metaphors is that of the 'body of evidence.' For example, "The body of evidence on intervention programs includes at least 1,500 clinical trials" (Kazak et al., 2010, p. 91) or "We have a sizable body of evidence based on different research methods" (Silverman, 2005, p. 310). Another such commonly employed metaphor is that of the 'evidence base': "The research evidence base relevant to health care proliferates at an astonishing rate" (Spring, 2007, p. 618) or "RCTs with lengthy follow-up intervals comprise an important piece of the relevant evidence base, especially for establishing treatment efficacy" (Tucker & Roth, 2006, p. 918). Although as already conceded the use of these terms may well be strictly metaphorical, the metaphors are extended quite far in suggestions that the body of evidence is "sizable," "proliferates," and can be divided into "pieces." The impression created by the use of such metaphors, that evidence is physical and observable, gives it the appearance of being objective and indubitable or given, in the same way that physical evidence (e.g., fingerprints, documents, physical objects) is in a courtroom. As will become more clear in the next chapter and the discussion below, however, the evidence used in psychology may not be as tangible or objective as such metaphors suggest.

The impression created by language described in the previous paragraph, that evidence is a physical entity, is furthered by the fact that evidence is sometimes discussed using language that suggests its objective existence, (e.g., "As noted above, often where evidence does already exist, it is not used or is ignored completely," Hunter, 2009, p.584) and can be "located" (e.g., "Clinicians do not generally have the time and/or the skills to synthesize all the evidence available to answer their questions; rather, they seek articles in which the evidence has already been located, appraised, and synthesized," Walker & London, 2007, p. 637). In a slightly tongue-in-cheek, although in parts quite accurate, manner, McDonald and Viehbeck describe physical conceptions of evidence as follows:

> Research users have their own version of evidentiary alchemy. In their view, relevant, easy-to-understand evidence should be summarized and placed in convenient repositories where it can be retrieved when required. Research is viewed as a commodity like bread or automobile tires. When you need it, visit your local repository and pick out the object of your

> desire. An entire industry concerned with the creation of best practice reviews and online repositories of meta-analyses has emerged to fill the perceived void. (2007, p. 141)

This account, although possibly slightly exaggerated, does seem to describe the way in which evidence is often talked about, at least by many of the proponents of EBP. Of course, there is a sense in which evidence exists and can be located and this sense need not imply that evidence is physical or object-like at all since non-physical entities (e.g., ideas, theories) can be said to exist or be discovered. Nonetheless, the sense in which physical objects can be said to exist differs from the sense in which non-physical entities exist, and therefore psychology may benefit from clarification of the manner in which evidence can be said to exist or be located. After all, the manner in which one may go about locating a rock or plant in one's backyard differs greatly from the manner in which one may go about locating a memory or idea in one's mind and therefore, how evidence is conceptualized will have implications for the manner in which psychologists may embark upon its exploration.

Beliefs/knowledge as evidence. In contrast to conceptualizations of evidence as physical entities, evidence is also often discussed in distinctly non-physical terms. Evidence in these accounts is conceptualized as information, knowledge, theory, and fact, all of which cannot be thought of as physical since they are propositional in nature. Some writers quite explicitly conceptualize evidence in theoretical terms. For example, Veerman et al. (2007) provide a definition of theoretical evidence, in contrast to descriptive evidence, which appears to mirror Aristotle's demonstrations: "Theoretical evidence goes beyond descriptive evidence in that a sound rationale or theory (e.g., program theory, change theory) is specified along with why and how intervention activities with a particular target group will lead to the intended outcomes" (p. 216). Some authors also suggest that evidence can be evaluated in terms of existing theory:

> In reviewing this history, it is apparent that among any group of like-minded practitioners, the standard of evidence for validity of a clinical practice has often been whether it fits the theory, whether the advocates strongly believe in the truth of their theories, and whether they appear to be sincere in advocating the value of clinical experience in support of their beliefs. (Beutler, 2000, cited in Silverman, 2005, p. 306)

Theories are usually defined as sets or relations of facts, beliefs, or principles that explain observed facts. If evidence is theoretical in nature, it appears that it explains physical or observable facts but is itself distinctly non-physical in the sense that there is no meaningful way to speak of a theory as a physical entity in a way that is consistent with the way in which the word 'theory' is usually employed. To some extent, this has implications for the certainty which one attributes to evidence. If evidence is in fact theoretical in nature, a degree of uncertainty may be attached to it as it is to all theories whereas a similar degree of uncertainty may not be attached to physical evidence.

Consistent with the conceptualization of evidence as theoretical, 'evidence' is also often defined in terms of components of theories such as information, knowledge, or facts (e.g., Bauer, 2007; McCabe, 2004; Pollio, 2003). For example, in a case study of an evidence-based approach to the treatment of social anxiety, Walsh & Hope (2010) on the one hand describe research evidence to support the treatment approach, but on the other hand also discuss evidence used in the treatment: "Jason learned to challenge these thoughts by evaluating evidence in support of and contrary to such thoughts (e.g., he had an excellent grade point average and he had been promoted to a supervisory position at work prior to beginning therapy)" (p. 59). In this particular instance, 'evidence' seems to be used to describe both research evidence but also a set of propositions about facts. More commonly, evidence is also conceptualized as knowledge or information, sometimes both within one article. For example, Pollio and MacGowan (2010) define evidence as knowledge: "Our definition of evidence includes not just clinical interventions, but also a knowledge of the impact of dynamics such as group process and structures, group leadership, member roles, and other factors" (p. 198). They also define 'evidence' as knowledge that leads to a certain outcome: "However, knowledge itself is not evidence, but becomes so when there is a reasonable probability that the information applied will be effective," p. 201). In addition, they define 'evidence' as knowledge that is part of theory (e.g., "On one level, the group worker has the knowledge of evidence for the choices he or she makes that is brought to the situation. In this sense, evidence functions in a similar manner to theory and group model," p. 203) and information (e.g., "to summarize, evidence, as used here, is any systematic collection of information," p. 198). Despite the variations in their definitions,

Pollio and MacGowan are consistent in their definition of 'evidence' in non-physical propositional terms.

Whether evidence is object-like or propositional is a longstanding philosophical question that will be addressed more fully in the next chapter. To provide a brief preview, however, it can be argued that whether evidence is conceptualized as physical or propositional in nature has important implications for the roles that evidence can play. Physical entities and propositions differ in terms of the attributes and roles that can meaningfully be ascribed to them if they are employed in a conventional manner. This has implications for the kinds of roles that evidence can play in EBP, such as the extent to which evidence is amenable to quantification via probabilities, a topic that will be taken up in much more detail in Chapter 7.

Objective versus Socially Constructed Evidence

A related divide in the EBP literature that partially maps onto the issue of whether evidence is physical or propositional has to do with whether evidence is taken to be objective, empirical, and publically observable, as opposed to based on social structures, processes, or consensus agreements. Despite the fact that the former position is certainly more prominent among proponents of evidence, the latter view is also well represented in the literature.

Empirical/objective evidence. **'**Empirical' is likely the word that is most commonly used in conjunction with the word 'evidence' in the EBP literature. There appears to be at least some consensus among proponents of EBP that, whatever evidence may be, it is most certainly empirical in nature. According to Task Force member, David Barlow (personal communication, August 6, 2011), "There's no real disagreement. I think almost everyone agreed that evidence was controlled observation, carefully controlled observation" and "explicitly everybody agreed that you know the evidence is basically empirically controlled observation" which he defined as "the empirical approach of relying on observations instead of theory, using theory to organize the observations…and collecting the data in reliable and valid ways that we all could agree on what we see." It would be fair to say that references to empirical evidence occur in the majority of articles on EBP (e.g., Goldenberg, 2006; Jacobs et al., 2010,

Silverman et al., 2008).[27] A closer examination of the use of the word 'empirical' in the context of evidence, however, indicates that it is also employed in a multitude of other ways. Most consistently with its conventional meaning in ordinary language and epistemology, 'empirical' is often defined as 'observable through the senses': "The 'evidence' in EBM refers to 'empirical evidence,' i.e. something that can be externally observed" (Morstyn, 2010, p. 222). In a move that may be either strategic and rhetorical or borne from a misunderstanding a much narrower conception of 'empirical' is also manifest in the literature and is exemplified in the following statement:

> Another misleading impression that may be created is that unless a treatment has been tested in an RCT, it has no empirical basis whatsoever. Although there is no doubt that the RCT provides a strong form of evidence that is difficult to refute, this does not mean that other types of evidence are not available or that those that are available are worthless. After all, it would be wasteful to commit the considerable expense and effort involved in running RCTs on treatments for which there was not already a strong body of evidence regarding effectiveness. Related to this is a common misuse of the term *empirical* in which it is implied that studies are only empirical if they use a group comparison method involving multivariate statistics. (Edwards et al., 2004, p. 590)

A traditional definition of the term 'empirical,' however, suggests that evidence must be observable through the senses and likely, observable reliably by multiple observers, which maps onto the notion that evidence is to serve as an objective adjudicator of fact.

The notion that evidence provides reliable and objective arbitration of facts and decisions is very commonly expressed in the EBP literature. Consistent with the notion that evidence must be observable to the senses, it is also often argued that evidence has to be such that it is the same irrespective of the person who is using it; that is, it must be objective (e.g., Edwards, 2007; Falkum, 2008; Misak, 2010). Such objectivity is thought to eliminate the impact of variations in personal opinion and bias: "The generally embraced view is that with the appropriate evidence at hand, decision and policy making will be optimal, legitimate and publicly accountable; that with the appropriate evidence, bias and arbitrary decisions will be eliminated, or at least monitored and kept at bay"

[27] In the article sample used for this chapter, a total of 135 references to "empirical evidence" (out of a total of 2728 references to evidence) was found.

(Montuschi, 2009, p. 426). As alluded to in the above section on data as evidence, there is also an assumption that standardization of procedures used to generate and utilize evidence is required in order for such evidence to be objective (e.g., Falkum, 2008; Hunsley et al., 2005; Jensen-Doss & Hawley, 2010). For example, it is argued that "Rather than adhering to traditional knowledge and authority, the clinician should rely on evidence produced by systematic, unbiased, and reproducible observations. The controlled, randomized clinical trial is considered the ideal model for establishing such evidence" (Falkum, 2008, p. 143). Similarly, when utilizing evidence it is argued that:

> Even if each measure used in an assessment is evidence based, because of limitations in human judgment, there is no guarantee that the resulting synthesis of information and conclusions are themselves truly evidence based. Although there can be considerable similarity in the case formulations developed for a patient by psychologists sharing the same theoretical orientation, the mean interrater reliability of such formulations is moderate at best. (Hunsley et al., 2005, p. 254)

The assumption thus is that evidence must be objective in the sense that it is consistent across people both in how it is generated and how it is applied, and that this is best accomplished through standardization procedures.

Evidence as a process/social construction. Contrary to objective accounts of evidence, some literature on EBP emphasizes the process in which evidence is employed as central to its role as evidence. In such accounts, it appears that whatever evidence might be (facts, data, findings, etc.) it only becomes evidence when it is applied in a certain process, sometimes referred to as the "evidence-based process" (e.g., Falzon et al., 2010; Jacobs et al., 2010; Pollio & MacGowan, 2010). Such accounts usually emphasize the integration of all three legs of the three-legged stool and appear to imply that only once evidence is integrated with clinical expertise and patient preferences does it become evidence. For example:

> The evidence-based process has been well documented and involves the following steps (this article concentrates on the first two): (1) Formulate a clear question about patient or research issue; (2) search the literature to find the best available evidence; (3) critically appraise the evidence for its validity, accuracy, and usefulness; (4) apply useful findings, integrating them with clinical expertise and patient's characteristics, culture, and

> preferences; and (5) evaluate the outcomes and, if needed, initiate a refined search. (Falzon et al., 2010, p.551)

In addition, some writers advance the notion that evidence changes in the process of its employment:

> Given the more fluid definition advocated in this article, evidence takes a more dynamic role in the practice situation. On one level, the group worker has the knowledge of evidence for the choices he or she makes that is brought to the situation. In this sense, evidence functions in a similar manner to theory and group model. However, given a critical thinking lens, the group worker makes his or her response based on critical thinking, namely a hypothesis on the results of the intervention. The response by the group or group member provides evidence as to whether the intervention has had the anticipated impact. This "single-system" approach can then be used as evidence to inform future practice situations—thus, the element of evidence represents an ongoing process. (Pollio & MacGowan, 2010, p.203)

According to such views, evidence is therefore neither empirical nor objective and in fact may not even be static but rather is the product of a dynamic process.

A small proportion of the EBP literature further refutes the objectivity of research evidence from a social constructionist perspective and instead conceptualizes evidence as a social process or social construction that is shaped by various social contextual factors. Some authors argue that consensus judgments by clinicians constitute a form of evidence: "Several important aspects of clinical practice in psychiatry and psychotherapy are not likely to be examined with controlled designs. Collegial consensus and systematically evaluated clinical experience may be the crucial sources of evidence concerning these aspects" (Falkum, 2008, p. 148). Messer (2004) similarly advocates for an approach to EBP where research is evaluated by a jury of experts who reach consensus agreement on a treatment, which then constitutes the evidence for that treatment. Many who write from this perspective emphasize the various social influences on the manner in which knowledge and evidence are conceptualized. For example:

> Because knowers are understood to be collaborative agents, whose epistemic projects are shaped by, and evaluated within, the communities where their knowledge-producing practices occur, standards of evidence

> are by no means ‘‘self-announcing’’, but rather historically relative, dynamic, and of our own making. (Goldenberg, 2006, p. 2626)

Most postmodern conceptions of evidence assume that all knowledge is shaped by social and contextual factors and that true objective empirical knowledge independent from the given social context of the knower is impossible:

> …rather than empirical evidence increasing certainty by factoring out the subjective features of everydayness that bias our understanding of things, the constructs of ‘‘objectivity’’, ‘‘universality’’, and ‘‘value-free’’ instead obscure the subjective elements that inescapably enter all forms of human inquiry. Since the evidence is by no means objective or neutral, but rather part of a social system of knowledge production, many feminist epistemologists recommend social models of scientific practice. (Goldenberg, 2006, p. 2626)

Arguments have also been put forth that evidence itself is quintessentially a political concept or an “instrument of power” (Goldenberg, 2006, p. 2630) used to advance certain interests in the context of a particular social and political system:

> The appeal to “science” or “scientific principles” can therefore have different meanings, depending on which pole is being used as a reference. The adversarial pole is of particular value when seeking to dismiss the arguments or claims of opponents. There is thus a politics of discourse and a politics of evidence (Edwards, 2007, p. 8)

Constructionist theorists and some other commentators on EBP highlight the extent to which evidence is inseparable from the process by which it is produced and the social context in which it is produced. Whether psychologists view ‘evidence’ as something that exists in a decontextualized fashion and is “discovered” through research, as opposed to something that is generated or created in a research context and gains its value through the process whereby it is created, may have some important implications for the manner in which they conduct their research and practice.

Where Does Evidence Come From?

An additional issue regarding the nature of evidence concerns its origin. The language that researchers use to describe the origins of evidence differs substantially

and implies particular views on the nature of evidence. Some researchers imply that evidence exists and is discovered or found through research or reviews of research, while other researchers imply that evidence is instead created through the research process. Again, this divide to a certain extent maps onto the divide between those researchers who view evidence as physical, empirical, and objective and those who view it as propositional, emergent from a process or socially constructed in a manner that will now be laid out.

Evidence exists and is found. Through the language they use when discussing evidence, a number of authors imply that evidence exists in some sort of independent, objective way (e.g., Hunter, 2009) and is merely "found" (i.e., discovered) through the process of research (e.g., Falzon et al., 2010; Hunsley, 2007a; Hunter, 2009). For example, Bernal, Jimenez-Chafey, and Rodriguez (2009) conceptualize research as a process of "gathering" evidence, implying that it exists (i.e., objectively) to be gathered: "gathering a critical mass of evidence for the use of adapted treatments for ECG's (ethnocultural groups) via these trials is impractical" (p. 363). Similarly, Treasure and colleagues (1998) conceptualize evidence as something that can be "gotten" through research: "It is possible that the randomized controlled trial is not the best way to get this sort of evidence" (p. 87). Other language that is often used to describe evidence involves the notion that evidence "accumulates" (e.g., Ayers et al., 2007; Katon et al., 2009), which at the very least implies a relatively passive manner in which evidence becomes knowable, or that evidence "emerges." For example, "Evidence also emerged that DC (drug court) was more effective than FC (family court with community service) at decreasing youth alcohol and polydrug use" (Henggeler et al., 2006, p. 51). The implication in this statement is that the evidence "emerged" through the trial conducted in the study reported, again implying that it came from somewhere and, thus, existed a priori.

Evidence is created through research. Much of the language in the EBP discourse, however, implies that evidence is created through research. The relationship between evidence and research is described using a variety of verbs. Evidence is discussed as being "developed" (e.g., Wild, 2006, p. 40), "generated" (e.g., Barrett et al., 2008, p. 131; Falkum, 2008, p. 143; Kazak et al., 2010, p. 91), "offered" (e.g., Tucker & Roth, 2006, p. 921), "created" (e.g., Treasure et al., 1998, p. 86), "produced" (e.g., Singh

& Oswald, 2004, p. 133), “provided” (e.g., Elliot et al, 2003, p. 219); and “delivered” (e.g., Barkham et al., 2003, p. 320) by research. All these terms imply that research plays a generative role with respect to evidence. For example, Kazak and colleagues state that “Other meta-analyses have generated evidence of the effectiveness of school-based interventions in preventing substance use and school dropout” (2010, p. 91). According to Treasure and colleagues (1998), “the randomized controlled trial is considered to be the gold standard method to creative evidence” (p. 86).

If evidence is created in some generative process in research, again, this has implications for the kinds of things that can be conceptualized as evidence, since research can meaningfully be said to create or generate certain things (e.g., facts, data, findings, literature) but not others (e.g., clinical expertise, patient preferences, physical objects). In addition, whether evidence is conceptualized as being created or “discovered” through research reveals certain ontological and epistemological assumptions with respect to the nature of the scientific process and the scientist’s access to reality, which will be discussed in more detail in the following chapter.

Conclusions – Evidence in EBP

A wide range of conceptualizations of evidence was apparent in the EBP literature. Many different conceptualizations co-exist in that literature. Logically incompatible conceptualizations at times co-exist within one and the same article and ambiguous conceptualizations of evidence make it difficult to ascertain what authors mean when they refer to evidence. Nonetheless, the EBP discourse appears to assume a much more narrow and technical definition of ‘evidence’ than its definition in general parlance. ‘Evidence’ in EBP is generally conceptualized as a particular kind of evidence produced or discovered in a particular manner using certain scientific methods. The EBP discourse may benefit from some ambiguity between the technical definition of ‘evidence’ that it employs and the more general meaning of the term in everyday language in the sense that reliance on evidence in its common sense may have more intuitive appeal than reliance on a more narrowly technical kind of evidence (e.g., evidence produced through a randomized-controlled trial).

The Object of Evidence in EBP

What exactly counts as evidence in EBP is far from obvious based on a review of the EBP literature. Unfortunately, the literature does not become any more unanimous or clear when examining that which the evidence is thought to be evidence for in EBP. Throughout the EBP literature, it is suggested that the objects of evidence are everything from various attributes of treatments and treatment approaches (including their efficacy, effectiveness, etc.), assessment (including the validity of tools employed or the assessment process itself), disorders (including definitions and causes of disorders), to treatment plans for particular patients, clinical expertise, therapeutic relationships, standards/policies, knowledge and truth, ethics, and, finally, even EBP itself.

Treatments as the Object of Evidence

Without a doubt, the most commonly espoused view within the EBP literature is that the objects of evidence are attributes (e.g., the efficacy, effectiveness, etc.) of particular treatments or treatment approaches. In fact, a large number of articles referred to evidence-based treatments, rather than EBP, in an apparent assumption that evidence is to be used to legitimize a particular treatment (e.g., Ruzek & Rosen, 2009; Waller, 2009; Yon et al., 2007).[28] What is generally meant by treatments in such articles is not the treatment of a particular patient, but rather more generally the effectiveness of a particular treatment approach such as cognitive-behavioural therapy or interpersonal psychotherapy. Claims such as the following are very common: "Cognitive-behavioural therapy (CBT) has a wide-ranging empirical base, supporting its place as the evidence-based treatment of choice for the majority of psychological disorders" (Waller, 2009, p. 119) or "As the availability of manualized evidence-based treatments continues to grow, the need for training will also continue to develop and intensify" (Ruzek & Rosen, 2009, p. 987). When reading such references while keeping in mind the Task Force's definition of EBP, one cannot help but wonder what it means for a treatment to be

[28] In the sample of articles used in this chapter, there were 627 references to evidence-based treatments and 1,389 references to EBT, the abbreviation commonly used for the term evidence-based treatment

evidence-based given that the Task Force's criteria for EBP seem to imply the integration of considerations that must occur at the level of the individual client and clinician. In fact, most of the Task Force members interviewed explicitly expressed disagreement with the use of the term 'evidence-based treatment':

> If you read the evidence-based literature, there is no, it's all a discussion about evidence-based treatments, right, as if the only research that's important is research about treatment and Drew was very persuasive about well there's research evidence on disorders – you need to know what the research on depression or social anxiety is – and that's basic, more basic science or experimental psychopathology than is treatment. But now when you say evidence-based it always is followed by the word treatment. And there's no official descriptor of what evidence-based treatment is and there's nothing in the policy that ever says evidence-based treatment. (D. Barlow, personal communication, August 6, 2011)

Or, very similarly:

> NL: Ok. Does the phrase evidence-based treatment even make sense to you? In your opinion, do you think there are evidence-based treatments or…?
>
> JN: No I prefer the term research-supported therapies. Because if it is evidence-based then you need to go through the three legs. And how do you begin, except as the treating psychologist, to begin to define that in a list, because it's by definition individualized and personalized. (J. Norcross, personal communication, April 28, 2011)

Many of the articles on evidence-based treatments implicitly confound EBP and empirically-supported treatment, a point which will be discussed in more detail towards the end of this chapter.

Assuming that the objects of evidence are particular treatment approaches, however, it is still far from clear what exactly evidence ought to establish about such treatments. One possibility is that evidence establishes that treatments are efficacious; that is, that they work in a controlled trial setting (e.g., Roth et al., 2008; Silverman et al., 2008; Veerman et al., 2007). Another possibility is that the evidence for a treatment approach is to show that such a treatment works in a clinical setting, defined as the treatment approaches' effectiveness (e.g., Longabaugh et al., 2005; McCabe, 2004; Rivett, Howarth & Harold, 2006). Yet another possibility is that evidence shows that

treatments work to ameliorate the symptoms of a particular disorder (e.g., Silverman & Hinshaw, 2008; Lee, 2007; Keel & Haedt, 2008). Still other articles imply that evidence can demonstrate a treatment's utility or cost effectiveness (e.g., Gray, 2002; Hunsley, 2007b; Reynolds, 2000) or its superiority to other treatment approaches (e.g., Parker & Fletcher, 2007). Finally, evidence is also thought to demonstrate that a treatment is delivered as it was intended, in an adherent manner (e.g., Miller et al., 2006) or in possibly direct opposition, that a treatment can be modified to work for a particular client (e.g., Lau, 2006). Therefore, even if evidence is primarily employed to support the use of particular treatment approaches, there are many possibilities in terms of what such evidence may be thought to establish about said treatment approaches. And, there appears to be no consensus on this.

Assessment as the Object of Evidence

Much less frequently, the EBP literature conceptualizes assessment as the object of evidence.[29] Nonetheless, a substantial number of articles do refer to assessment as the object of evidence and imply that assessments are to be based on certain kinds of evidence (e.g., Brown, 2006; Fago 2009; Lewis, 2009). Oftentimes, evidence-based assessment (EBA) is conceptualized as assessment that utilizes standardized instruments with certain psychometric properties established in prior research: "Despite the importance of EBA, much available evidence suggests that clinicians are not engaged in assessment practices consistent with EBA, including what is arguably the core component of EBA: use of standardized assessment tools with research support for their reliability and validity" (Jensen-Doss & Hawley, 2010, p. 886). Here then it seems that the evidence is thought to bear on the validity and reliability of the measurement tools employed. Discerning the meaning of evidence-based assessment is almost more complicated than doing so for evidence-based treatments, however. Clinicians commonly distinguish between assessment and testing. Testing involves the administration of certain measurement tools, while assessment involves the contextualized interpretation of the measurement tools for a particular patient. It would

[29] Only 167 references to evidence-based assessment and 475 references to EBA, its abbreviation, were found in the present sample of articles.

thus appear that evidence would have to support not only the tool employed but also the interpretation arrived at by the clinician. In fact, some psychologists have advocated such a distinction:

> …although at present little evidence bears on the issue, it is critical that the entire process of assessment (i.e., selection, use, and interpretation of an instrument, and integration of multiple sources of assessment data) be empirically evaluated. In other words, a critical distinction must be made between evidence-based assessment methods and tools, on the one hand, and evidence-based assessment processes, on the other. (Hunsley & Mash, 2007, p. 33)

In the context of evidence-based assessment, therefore, it is unclear whether the tools employed in the assessment must be supported by psychometric evidence, or whether the interpretation of the data derived from the application of these tools must also be based on empirical evidence, as advocated by Hunsley and Mash. Depending on which element of the assessment process one is concerned with, different kinds of evidence would be relevant (e.g., psychometric evidence in the former case, facts about the patient and the testing context along with research on such individual factors and their impact on testing in the latter case). In fact, it has also been suggested by some that one's appraisal of individual patient factors and their bearing on clinical judgments, as well as one's clinical judgments themselves ought to be based on evidence. These claims will be explored in the next section.

The 'Three-legged Stool' as the Object of Evidence

Some of the discourse on EBP discusses the relationship between evidence and both clinical expertise and patient preferences. In particular, some articles argue that both clinical expertise or judgment and one's appraisal of the impact of individual patient differences and preferences must be based on evidence. For example:

> With deserving respect for all who try to develop and disseminate explicit conceptualizations of EBP, definitions that (rightfully) honor clinician and patient variables can inadvertently confuse students of EBP by blurring the distinction between the broader notion of best clinical practices and the specific construct of EBPs. Such confusion will assuredly occur unless EBP definitions emphasize the adoption of *evidence on* clinical

> expertise and *evidence on* the influence of specific patient characteristics on therapeutic process and outcome. (McCabe, 2006, p. 53).

Some have suggested that even the manner in which one seeks to establish a relationship with patients must be based on principles of behaviour supported by evidence (Bohart, 2005; J. Norcross, personal communication, April 28, 2011). In addition, it has been proposed that the background knowledge clinicians have about treatments in general and particular clinical principles ought to be examined and must be based on evidence rather than clinical lore (e.g., Pollio & McGowan, 2010). According to the Task Force's policy statement (2005), treatments are also to be adjusted to individual patients based on monitoring of progress and outcomes. Therefore, patient progress can be conceptualized as another object of evidence gathered during treatment. Finally, a relatively large number of articles on EBP have focused on the applicability of treatment approaches and particular clinical practices to patients from diverse backgrounds (e.g., Campbell et al., 2008). According to these articles, EBP must establish whether practices can be applied to members of various cultures, and thus the object of evidence can also be cultural equivalence of a certain approach (e.g., Bernal, Jimenez, & Rodriguez, 2009; Campbell et al., 2008). According to some of the EBP literature, then, evidence can have as its object elements of both clinical expertise and patient preferences and characteristics.

Disorders as the Object of Evidence

At times, the EBP literature implies that the nature, etiology, and definition of mental disorders themselves are the objects of evidence. Most commonly, the literature refers to evidence for the treatment of the symptoms of a particular disorder (e.g., Hunsley, 2007a; Keel & Haedt, 2008; Lee, 2007). Other times, however, it is argued that the manner in which a disorder is defined and classified can be established or modified through evidence. For example:

> However, a growing body of evidence from the phenomenological, neurobiological, genetic, and treatment literatures has raised questions about this nosological classification and, as noted next, provides some justification for the consideration of OCD separate from the other anxiety disorders. (Barret et al., 2008, p. 132)

In addition, models of disorder are also thought to be subject to modification based on evidence: "Psychological models of OCD have been radically challenged by the relatively recent emergence of evidence for a biological basis for the disorder" (Frederick, 2007, p. 191). Finally, views of the nature of psychopathology are thought to be based on evidence:

> These strategies are based on our best evidence at this time of the nature of psychopathology, and there is the strong assumption (not yet fully validated but supported by evidence from other areas of health care) that this approach will facilitate an interactive process that will improve treatment outcome. (Barlow, 2005, p. 310)

According to these accounts, then, evidence can be used to establish what a mental disorder is, how it is to be defined or diagnosed, and what may have caused it. Some philosophical issues related to such accounts and the assumptions about the meaning of a mental disorder will be discussed in the following chapter.

Truth and Knowledge as the Object of Evidence

At times in the EBP discourse, it appears that evidence is employed in order to establish knowledge in the form of answers to questions, conclusions, theories, or justification for beliefs. Frequently, evidence is discussed as supporting or refuting a certain theory, as one would expect in any empirical science; for example, "When results are inconsistent with expectations, this may decrease the stature of the overall theory, but negative evidence almost never is sufficiently definitive to warrant tossing out the entire theory" (McFall, 2005, p. 319). Evidence is also thought to establish and support certain conclusions: "It is generally understood that actuarial risk assessment is more accurate than unaided clinical judgment for predicting critical target events like child maltreatment or juvenile delinquency. Scholars have amassed substantial empirical evidence in support of this conclusion" (Schwalbe, 2008, p. 1458). Finally, some scholars advocate the role of evidence in establishing the legitimacy of psychological practices and persuading the public of their value (e.g., Kelvin, 2005; Milton, 2002; Reynolds; 2000). In all these instances, the object of evidence is knowledge, truth, or the establishment of fact, an aim that appears to be much broader than that of establishing the efficacy of a particular treatment or assessment method.

The evidence that can be employed towards such ends appears to be much more varied than the evidence that can be used to establish the efficacy of a particular treatment approach.

Other Objects of Evidence

Beyond the summary provided above, objects of evidence have also been posited to include the establishment of standards and policies (e.g., Pawson et al., 2010; Yamada & Brekke, 2008) that are to be based on evidence, for example, "Absence of evidence-based standards for culturally responsive clinical assessment in PSR prevents the provision of uniform training programs across agencies, and efforts to train providers in needed skills are not always sufficient to change their assessment skills" (Yamada & Brekke, 2008, p. 1389). It has also been suggested that ethical policies and principles in research ought to be based on research: "One way to determine IRB best practices is through empirical research. Just as evidence-based medicine may improve patient outcomes, evidence-based research ethics may enhance the ethical conduct of research" (Anderson et al., 2007, p. 96). Such suggestions imply that normative judgments, such as ethical standards, can and, ought to be based on evidence, which raises questions about the kinds of evidence that can be employed towards such ends. It is difficult to see how some kinds of evidence (e.g., effect size data from randomized-controlled trials or meta-analyses of randomized-controlled trials) could be employed as the basis for the establishment of ethical standards. Of course, such evidence may be used to establish whether a given practice is harmful or helpful in a certain group of people, but defining harm itself requires normative presuppositions that may be difficult to base on such evidence and rather are contingent on particular value systems.

Lastly, throughout the EBP literature, it is commonly suggested that EBP itself is an approach to practice that can or ought to be supported by evidence (e.g., Nathan, 2004; Spielmans, Gaitlin & McFall, 2010; Westen et al., 2005). Oftentimes, it is suggested that evidence must be collected to demonstrate that EBP produced superior outcomes compared to other approaches to psychological practice (e.g. "Although far from complete, the mounting evidence indicates that, not only is EBP in psychology feasible, it also is likely to improve upon many of the behavioural health care services

currently available to the public" -- Hunsley, 2007b, p. 39). Of course, ironically, establishing the legitimacy of EBP based on evidence is a practice that itself assumes the value of the tenets of EBP; that is, it assumes that practices are to be based on or established through evidence and as such leads to an infinite regress.

Summary

Based on the above review, it is apparent that the EBP literature suggests a great number of possibilities with respect to the components of clinical practice that ought to be adjudicated by means of evidence. Most commonly, it is suggested that treatment must be based on evidence, although it is unclear whether such evidence is to establish efficacy, effectiveness, symptom reduction, cost effectiveness, or adherent delivery of a treatment approach or the treatment of a particular individual. At times, assessment is also put forth as the object of evidence; here it is unclear whether the validity of assessment tools must be based on evidence or the interpretation of the data derived from the application of the tool to a particular individual. In addition, some have also suggested that the impact of patient differences on clinical practice and aspects of clinical expertise such as ways to establish or maintain the therapeutic relationship must also be based on evidence. Based on yet other proposals, evidence should inform clinician's understanding of the definition, diagnosis, and causes of mental disorders. It is also sometimes suggested that guidelines and ethical standards must be based on evidence. Finally, some have proposed that even the legitimacy of EBP itself must be established based on evidence, in a manner that appears quite logically circular. Although evidence may well be brought to bear on a variety of different claims, the diversity with respect to the objects of evidence in EBP necessitates similar diversity in the kinds of evidence that may be required to adjudicate such claims. It is unlikely, for example, that the same standards of evidence can be applied to evaluate evidence for the efficacy of a treatment approach as can be applied to evaluate the evidence supporting particular ethical guidelines for treatment. Therefore, at the very least, if many components of clinical practice can be adjudicated via evidence, the evidence employed must be quite varied, as must the manner in which the evidence can be thought to bear on these varied claims.

The Relationship between Evidence and Its Object in EBP

Given the tremendous variations in perspectives on what constitutes evidence and what constitutes the object of evidence in EBP, questions as to the way in which diverse types of evidence might be related to such varied objects arise. The relationship between evidence and its object in EBP appears to be particularly important to the clinician engaged in EBP, who will be tasked with considering such relationships between evidence and practice in a manner that permits the integration of evidence into practice. Of the three dimensions of the conceptualization of evidence (evidence, its object, and their relationship), the relationship appears to receive the least attention in the EBP literature. Few of the articles in the present sample dealt with the relationship between evidence and its object explicitly. In the APA Task Force documents, the integration of evidence and practice is also not elaborated much. Nonetheless, a few main perspectives on the relationship between evidence and its object can be identified in the EBP literature. Most commonly, the literature deals with the evaluation of evidence and particularly the question of what constitutes "best evidence." Most articles imply an ordinal weighting scheme for evidence in the form of a hierarchy of evidence based on types of evidence or the manner in which the evidence was generated. Dichotomous evaluations of evidence whereby a practice either is or is not supported by evidence have also been suggested. Some literature also posits a quantitative statistical relationship between evidence and its object based on probability or the quantity of evidence that exists. In addition, a few authors posit that one cannot consider parts of the evidence in isolation but must consider the total evidence available in order to arrive at sound conclusions. These various perspectives on the relationship between evidence and its objects will now be explored.

What is "Best Evidence" in EBP

The EBP discourse reflects some varied perspectives on how evidence should be evaluated or how one should go about determining the "best evidence." In the EBP literature, hierarchies of evidence based on research designs employed, the quality of the study through which the evidence was produced, or the evidence of patient preferences and clinical expertise relative to evidence are proposed. All these

perspectives suggest that the relationship between evidence and its object can be based on some hierarchy of evidence such that forms of evidence placed higher on the hierarchy provide stronger support for their objects than do other forms. In such an approach, then, one evaluates evidence based on certain qualitative criteria and determines the strength of the relationship between that evidence and its object based on these criteria. Whether best evidence is determined based on the methods used to generate that evidence or other factors may be important if psychologists are to select "best evidence" to support their practices.

The hierarchy controversy. In its official documents, the APA Task Force (2006) provides a list of research methods that may contribute to EBP, with the caveat that "multiple research designs contribute to EBP, and different research designs are better suited to address different types of questions" (p. 274). The list of research methods is mostly adapted straight from the Institute of Medicine statement upon which the Task Force's document was based. In this document, it is presented as a hierarchy of evidence. As well, in its policy statement (2005), the Task Force included a statement that "The validity of conclusions from research on interventions is based on a general progression from clinical observation through systematic reviews of randomized clinical trials, while also recognizing gaps and limitations in the existing literature and its applicability to the specific case at hand" (p. 1). It is therefore not entirely clear whether the list of methods provided is intended to be interpreted in a hierarchical fashion or not.

In my interviews with Task Force members, it became very clear that some Task Force members viewed the methods listed as a mere list of research methods which are each useful for a different purpose, while other Task Force members viewed it as a hierarchy from least to most rigorous sources of evidence. For example, S. Hollon (personal communication, April 26, 2011) voiced his support for a hierarchy of evidence as follows:

> If the question is what kind of designs do we use to rule out rival alternatives then I think there's a bit of a hierarchy and I think the public report ended up talking about a hierarchy moving from simple clinical observation up through consensus up through uncontrolled trials up through randomized controlled trials.

D. Barlow (personal communication, August 6, 2011) agreed with Hollon's interpretation: "Yeah, you find some who favor the more clinical prediction and say well they're all equal, that sort of thing. So that would be a disagreement. But the APA policy clearly states that there's a hierarchy of evidence." Other Task Force members advocated for the value of all methods on the list and expressed some ambivalence about whether the list is to be interpreted as a hierarchy. For example, J. Norcross (personal communication, April 28, 2011) argued that whether the list is to be interpreted as a hierarchy "depends upon the question." Finally, some Task Force members (e.g., B. Wampold, R. Levant) strongly disagreed with a hierarchical interpretation of the various research methods:

> I personally conceptualize, and I think that the Task Force conceptualized that list not as hierarchy and I not only think that, I know that. A close reading of the document will reflect that and some of the background materials that Lynn has given you, that these were, it wasn't that the top of the list is kind of unimportant and when you get down to the RCT's that's really important, not at all. You know, it really is different research for different purposes. (R. Levant, personal communication, June 27, 2011)

The issue of whether the relationship between evidence and its object can be interpreted based on a hierarchy of research methods through which evidence is generated is certainly a controversial one, even among Task Force members.

In the evidence-based practice literature, hierarchies of evidence are widely endorsed (e.g., Patterson et al., 2004; Reynolds, 2000; Montuschi, 2009). Many times such hierarchies are suggested based on research methods used e.g. "In the evidence-based discourse, evidence is graded according to the methods used to collect it" (Broekart et al., 2010, p. 229). Often such hierarchies are described as integral to evidence-based practice:

> To be considered "evidence-based" a practice must have been established as effective through scientific research according to set criteria (Mullen, 2002) involving a hierarchical ranking of evidence, with randomized controlled trials ranked highest, typically followed by well-designed non-randomized trials, cohort or case-control trials, multiple time series trials, and finally descriptive studies and case reports. (Scheyett, 2006, p.72)

Many of the articles on EBP also interpret APA's list of research methods as a hierarchy:

> Consistent with approaches to evidence-based practice in other professions, the APA Presidential Task Force on Evidence-Based Practice recommended the use of an evidence hierarchy when psychologists seek to determine how to use the scientific literature to inform their professional practices. (Hunsley, 2007a, p. 114)

Nonetheless, the conception of a hierarchy has been a target of many critics of EBP (e.g., Chwalisz, 2003; Ramey & Grubb, 2009; Tannenbaum, 2003) who often oppose the notion that some research methods are more useful than others: "All types of evidence could make equally valuable contributions to understanding, although quality variation will obviously exist within a given type of evidence. For instance, findings of a rigorous qualitative study could be of greater evidentiary value than findings of a poorly conducted clinical trial" (Chwalisz, 2003, p. 502).

Additional hierarchies and classification schemes for evidence are proposed in the EBP literature. As argued by Chwalisz (2003), above, some argue that qualitative criteria should be employed to assess strength of evidence (e.g., Hennsey et al., 2006; Hunsley, 2007a; Kratochwill, 2007). Hierarchies to rank the three components of the three-legged stool (i.e., best evidence, clinical expertise, and patient characteristics and preferences) also have been suggested. Some argue that all three legs are of equal importance and serve equally in their respective roles as evidence and thus oppose any hierarchical view of the three factors (e.g., Adams & Drake, 2006). Others have advocated for a hierarchy that privileges best evidence over the remaining two legs (e.g., Schlonsky et al., 2004). All hierarchies and classification schemes, however, share a sense that the strength of the relationship between evidence and its object or the extent to which evidence supports its object is based on certain qualitative criteria, such as how the evidence was generated or what it pertains to. According to such accounts, the strength of different types of evidence varies as a matter of degree. Hierarchies provide an ordinal way to evaluate evidence and emphasize how strongly a given claim is supported by evidence, based on the evidence's location in the hierarchy, rather than simply whether it is supported or not.

Dichotomous relationships between evidence and its object. In contrast to the hierarchies of evidence often proposed, a large number of articles in the EBP literature also treat the relationship between evidence and its object as binary or dichotomous. According to such perspectives, evidence either does or does not demonstrate or prove its object (e.g., Logsdon et al., 2007; McCurry et al., 2007; Scogin, 2007). In such accounts, criteria for the evaluation of evidence are usually suggested as a means to make a simple decision as to whether the evidence does or does not support a given practice or decision (e.g., have a certain number of randomized-controlled trials shown consistent effects; Silverman et al., 2008). Based on such criteria, it is then assumed that certain treatments can be "considered" evidence-based (e.g., Scheyett, 2006), "designated" evidence-based (e.g., McCabe, 2006), or "rated" as evidence-based (e.g., Gorman & Huber, 2009).

These perspectives for the most part confound evidence-based practice with empirically supported treatments. They employ some variant of the criteria proposed by the task force on empirically-supported treatments as a means of deciding whether a treatment or practice is evidence-based or not. Confusions between evidence-based practice and empirically supported treatments are disturbingly common in the literature (e.g., Spielmans, Gaitlin & McFall, 2010; Wachtel, 2010; Yon et al., 2007). Many authors simply treat these two concepts as synonymous and imply that EBP is merely a change in terminology from empirically supported treatment. For example, "Over approximately the past 15 years, the terms used to reference EBT have changed, from empirically validated therapies to empirically supported therapies to evidence-based therapies" (Spielmans, Gaitman & McFall, 2010, p. 235). Often they imply that there exist objective criteria whereby it can be determined whether a treatment is evidence-based or not:

> In accordance with the EBT criteria, we searched for interventions with at least two supporting peer- reviewed, published studies that had prospective designs and at least 30 participants randomly assigned to the same treatment condition, either within the same study or across studies using the same or a similar treatment approach. (Logsdon et al., 2007, p. 29)

Notably, the criteria listed to make dichotomous judgments about the value of evidence in establishing whether a treatment is evidence-based or not are usually the criteria used to identify empirically supported treatments. Such criteria are not suggested anywhere in the documents of the Task Force on Evidence-based Practice and in fact are quite inconsistent with the conceptions of EBP endorsed by Task Force members.

In addition, some of the proponents of a dichotomous approach to the evaluation of evidence also imply that such dichotomous judgments can be based on the existence of evidence compared to the absence of evidence to support the use of a certain treatment approach or practice (e.g., De Los Reyes et al., 2008; Fonagy, Roth & Higgitt, 2005). They tend to take a strong stance against the use of any treatment, which cannot be supported by evidence, but usually constrain evidence to be only evidence generated by randomized-controlled trials:

> In the practice of clinical psychology, non-EBTs for adults and youths continue to be used when EBTs that target the same behaviors are available. Although a given study might reveal inconsistent outcomes—and this raises significant issues—this ought to be presented in the context of a key reality: Hundreds of‘ ‘evidenceless’ treatments are being administered to patients. (De Los Reyes & Kazdin, 2008, p. 50)

In most of the literature that treats the relationship between evidence and its object as dichotomous, it is implied that treatment approaches not supported by evidence that meets a certain criteria should not be employed by clinicians.

Many authors, of course, correctly distinguish between empirically supported treatments and EBP, and view EBP as primarily an idiographic, individualized approach based on a consideration of all three of the components highlighted in the Task Force statement (e.g., Wifstad, 2008; Williams & Fulford, 2007). However, references to dichotomous standards of evidence are very common in the EBP literature and seem to suggest that treatments themselves (e.g., cognitive-behavioural therapy, interpersonal psychotherapy) can be either evidence-based or not (e.g., Pelham & Fabiano, 2008; Redmond et al., 2009; Yon & Scogins, 2007). While in the former view the evidence in EBP pertains to a particular treatment context (provider and setting) and client in that context, in the latter view it pertains merely to the treatment approach itself. If EBP is, as APA (2005) defines it, “the integration of the best available research with clinical

expertise in the context of patient characteristics, culture and preferences" (p.1), it becomes difficult to see how one can speak of "evidence-based treatments" in general nomothetic terms in a meaningful way. Such a definition seems to imply that the treatment for each individual will vary based on the individual, the judgment of the treating clinician and the available evidence and thus the relevant evidence in each treatment context will differ and subsequently, the specific treatment approach suggested by the evidence and remaining factors may also differ from one context to the next.

A quantitative relationship between evidence and its object. Aside from proponents of ordinal and dichotomous relationships between evidence and its object, many authors also suggest that the former relationship can be characterized in a quantitative, continuous manner based on certain statistical relationships. Similar to ordinal accounts discussed above, in such quantitative accounts, support for a given claim or practice is a matter of degree. In addition, the strength of the evidence for a given claim, in such accounts, it is thought can be quantified in some manner. For example, it has been suggested that more evidence for a certain object provides stronger evidence and thus strength of evidence is a matter of the quantity of available evidence. In very simplistic terms, in such an account then, five studies supporting a practice provide stronger evidence for said practice than two studies supporting a competing practice. The notion that more studies in support of a certain practice constitute stronger evidence for such practice appears to be an intuitive one (and one that has been advanced by many philosophers as discussed in the next chapter). This notion is certainly well represented in the EBP literature (e.g., Falzon et al., 2010; Salyers & Tsemberis, 2007; Sturmey, 2009;). More complex systems to weight the quantity of available evidence based on some qualitative criteria have also been suggested however. Some authors suggest that a sufficiently large number of studies of a certain kind (e.g., single-case studies) can trump studies of another kind (e.g., RCT's) in terms of their evidentiary value: "When carefully designed, repeated single case studies could provide evidence that is just as convincing as the evidence from RCTs" (Veerman et al., 2007, p. 218). Many authors suggest that a greater number of studies provide stronger evidence (e.g., Salyer et al., 2007).

In addition, some contributors to the EBP literature also conceptualize the relationship between evidence and its object based on statistics such as correlations or effect sizes. Some authors suggest that statistics such as reliability and validity coefficients[30] provide indices of the strength of the evidence such that larger coefficients indicate stronger evidence. For example, Campbell and colleagues (2008) argue:

> For reliability evidence, we determined that internal consistency reliability .80 was ‘‘good’’ and temporal stability reliability .70 was ‘‘good.’’ For validity evidence, we summarized different types of data and in the case of evaluating validity coefficients (e.g., correlation between the test and similar measures) no specific thresholds were established. (p. 1002)

Other authors interpret effect sizes, especially in the context of meta-analyses, as an index of the strength of evidence such that larger effect sizes provide stronger evidence (e.g., Thorn, Cross & Walker, 2007; Veerman et al., 2007). Finally, the relationship between evidence and its object is often characterized as a probabilistic one where greater probabilities indicate stronger evidence (e.g., Montuschi, 2009; Pollio & MacGowan, 2010). Usually, evidence is thought to increase the probability of a certain outcome (its object), such as a treatment success. For example, “evidence is incorporated into the intervention process to increase the likelihood of effectiveness relative to unsystematic clinical intuition” (Pollio & MacGowan, 2010, p. 200). In quantitative accounts of the relationship between evidence and its object, the relationship is thus characterized in terms of the quantity of available evidence, reliability and validity, effect sizes, or probability and is thought to be adjudicated by simply looking at a number posited to represent the strength of the evidence. Of course, such an interpretation of statistics itself disregards various limitations of statistics as estimates of relationships, as well as the extent to which other factors such as study designs and samples may bear on the interpretation of findings and its relevance as evidence for a given claim.

Comprehensive evidence. Some of the EBP literature also emphasizes the importance of considering all evidence for clinical decisions (e.g., Hunsley, 2007a;

[30] It has to be noted here that the consensus agreement among statisticians is that no statistic can provide a “validity coefficient.”

Schlonsky et al., 2004; Wachtel, 2010). In such accounts, often the danger of ignoring pieces of evidence and focusing only on some evidence that meets certain criteria (e.g., randomized-controlled trials) is highlighted: "…social phenomena can (and ought to) be approached by different methods, each of which displays different degrees of effectiveness…ignoring evidence coming from different sources might be not only wasteful, but harmful" (Montuschi, 2009, p. 431). Some authors express concerns that important aspects of clinical practice, such as findings with respect to the therapeutic relationship, will be overlooked if only some types of evidence (i.e., randomized-controlled trials) are considered (e.g., Hunsley, 2007a). The relationships between different pieces of evidence can be complex. Some pieces of evidence can contradict and defeat others or at least attenuate the strength of the relationship between other pieces of evidence and their objects. Therefore, considering only individual pieces of evidence or basing a practice on a certain portion of the evidence risks leading to erroneous conclusions (e.g., Shlonsky & Gibbs, 2004).

Describing the Relationship between Evidence and Its Object in EBP

Different language is used throughout the EBP discourse to describe the relationship between evidence and its object. This language also provides some clues as to how the relationship between evidence and its object is conceptualized. At times, a relationship of possession is implied whereby treatments "have evidence" (e.g., Ginsburg, 2006; Roth et al., 2008; Willenbring et al., 2004). The treatments themselves seem to be attached to the evidence. For example "This reflects the fact that CBT has the most substantial evidence base supporting its effectiveness in the treatment of depression and anxiety" (Roth et al., 2008, p. 130). Different terminology is also used to describe the way in which evidence supports a claim: evidence "proves" (e.g., Tibbits et al., 2010, p. 252), "demonstrates" (e.g., Spring, 2007, p. 617), "supports" (e.g., Schwalbe, 2008, p. 1458), "suggests" (e.g., Joinier et al., 2005, p. 272), and "guides" (e.g., Shlonsky & Gibbs, 2004, p. 138). Some of these terms imply a fairly deterministic relationship (e.g., proves, demonstrates), while others imply a more probabilistic relationship (e.g., supports, suggests, guides).

Summary

Based on the conceptualizations described above, the relationship between evidence and its object has been characterized in a variety of ways. When the relationship is characterized in more qualitative terms, hierarchies of evidence or dichotomous judgments of the evidence based on certain criteria are suggested. At times, dichotomous perspectives confound EBP and the empirically supported treatment movement, suggesting that simple criteria exist to decide whether a treatment approach is evidence-based or not. In more quantitative accounts of the relationship, the quantity of evidence in support of a practice is suggested as one way to gauge the strength of the relationship between the evidence and the practice. In addition, statistics such as reliability and validity coefficients, effect sizes, and claims invoking probability in one way or another also have been conceptualized as measures of the strength of the relationship between evidence and its object. Again, what can constitute evidence and its object in each of these relationships will vary depending on how the relationship is conceptualized.

Conclusion: Evidence in the EBP Literature

If only one thing is revealed in the above review, it is that evidence, its object, and the relationship between evidence and its object are characterized in diverse, complex, and, most problematically, inconsistent ways, both by Task Force members and within the EBP literature (sometimes within the same article). With respect to the nature of evidence itself, several questions emerge from a review of this literature. Some argue that only research evidence is to be conceptualized as evidence and that clinical expertise and patient characteristics and preferences are additional but separate components of EBP. Others, however, argue that all three of these components are conceptualized as evidence. Even if research evidence is the only kind of evidence in EBP, it is unclear what part of research can serve the role of evidence: the data, or the research literature? Different perspectives have also been put forth in terms of whether evidence is physical or propositional in nature. It is also unclear whether evidence is objective and empirical or instead part of a dynamic process and socially constructed. Finally, different language is used to describe the origin of evidence and implies that

evidence is either given and merely discovered through research, or instead created or produced through research. All these perspectives on the nature of evidence come with different ontological and epistemological implications that will be discussed in the next chapter.

The EBP discourse also proposed numerous possibilities with respect to that for which evidence serves as evidence. Potential candidates as objects of evidence include treatment approaches, assessments, all components of the three-legged stool, and disorders, as well as truth and knowledge itself. It also seems that different kinds of evidence can potentially serve as evidence for each of these objects. However, how one conceptualizes the object of evidence will have implications for the kinds of things that one can consider to be evidence, a topic to which I shall return in chapter 7.

The relationship between evidence and its object is less focal in the EBP discourse. Nonetheless, the characterization given of it is diverse. At times, it is conceptualized as a qualitative relationship based on hierarchies of evidence (ranked according to certain qualitative criteria); at other times it is understood in terms of a dichotomous evaluation of evidence. In such accounts, often EBP is treated as synonymous with empirically supported treatments, despite the fact that these two approaches are historically, conceptually, and methodologically distinct. Quantitative perspectives on the relationship between evidence and its object focus on the quantity of available evidence or statistics such as reliability and validity statistics, effect sizes, or references to probability. Again, each of these characterizations of the relationship between evidence and its object can sensibly be applied only to certain kinds of evidence and certain kinds of objects.

The goal of the present chapter was to highlight some of the conceptual dimensions of the EBP discourse that would benefit from additional clarification and analysis, which will be the main focus of the following chapter. What ought to be clear from what has been presented in the current chapter is that the EBP discourse presently is diverse at best and opaque at worst with respect to the nature of ‘evidence’, its object, and its relation to its objects.

Chapter 7:
Philosophy of Evidence

Philosophers of science and those interested in epistemology have long been concerned with the concept of 'evidence' and its role in both the scientific process and the establishment or justification of knowledge. In the third chapter of this work, I argued that philosophical treatments of predecessors to the concept of 'evidence' (i.e., concepts employed to serve the function that evidence serves in modern discourse) could be traced back to Antiquity. The concept of 'evidence' itself, however, has come to play a more prominent role in modern epistemology and philosophy of science. In fact, Kelly (2008) argues that evidence is one of the most central issues of modern epistemology and Ayer (1982) goes so far as to state that modern philosophy can best be described as 'the study of evidence.'

Some (e.g., Achinstein, 2000), however, have argued that philosophical accounts of evidence, for the most part, are of little value to modern science. Achinstein contends that 'evidence' in philosophical accounts does not serve the functions that 'evidence' is called upon to serve in science and therefore philosophical accounts have been and should continue to be ignored by scientists. In some realms of science, however, particularly as highlighted in the previous chapter in the EBP discourse, a great degree of conceptual and theoretical confusion reigns. In addition, as laid out in the introduction to this thesis, defining or clarifying the concept of 'evidence' is a conceptual not an empirical matter and cannot be accomplished on the basis of evidence without presupposing that which it is seeking to establish. Contrary to Achinstein, in this Chapter I will argue and demonstrate that major perspectives in philosophical treatments of 'evidence' map well onto the conceptual and theoretical issues that arise in a review of the evidence-based practice discourse. Most of the competing perspectives on 'evidence' in evidence-based practice have been addressed previously in some manner in debates of philosophers. Here, I turn to these philosophical accounts of evidence in

an attempt to clarify the implications of various conceptions of evidence adopted in the evidence-based practice discourse. I briefly outline philosophical perspectives on the nature of 'evidence,' the object of evidence, and the relationship between evidence and its object and apply them to the evidence-based practice discourse. Throughout this Chapter, I argue that philosophy can provide clarification of assumptions and implications of various perspectives of 'evidence' employed in evidence-based practice in psychology.

What is Evidence?

In the context of philosophy, various perspectives on what can constitute or serve the role of 'evidence' exist. No account has emerged victoriously in evidence debates. Consensus on what constitutes evidence in science or epistemology remains out of reach. Nonetheless, philosophical accounts of evidence greatly elucidate the implications that various accounts of evidence have for knowledge and science; particularly, they provide a much clearer discussion of the kinds of uses and claims that can legitimately be attributed to different kinds of evidence. If psychologists were to choose an ideal candidate for the role of evidence in EBP from the many competing candidates that currently exist, it would behove them to be aware of the implications of their choice. If, on the other hand, psychologists continue to employ various conceptualizations of evidence with some flexibility, it would seem even more important for them to be aware of the roles that different kinds of evidence can play, in order to make an informed choice among the multiple conceptualizations of 'evidence' that may be available to them and in order to employ 'evidence' appropriately. Philosophy can certainly shed some light on the debate between competing views of evidence as object-like or propositional, objective and publically observable or subjective and internal, the role of background knowledge (or in the case of psychology – clinical expertise) in the application of evidence, and finally the kind of evidence that can adjudicate between competing claims and evidence that cannot. Therefore, when it comes to conceptualizing evidence, philosophical accounts potentially can contribute valuable insights to the EBP discourse.

Is Evidence Propositional or Physical?

According to Kelly (2008), most philosophical accounts of the ontology of evidence can be divided into those that view evidence as exclusively propositional and those that propose that at least some evidence can be non-propositional or more object-like.

Classical foundationalism. In most ordinary language contexts, the concept 'evidence' is primarily employed to describe physical objects. For example, in criminal trials evidence often consists of objects from the crime scene, such as bullets, guns, photographs, and fingerprints. In medical contexts, diagnoses are based on evidence such as blood and tissue samples or physically visible signs (e.g., lumps, marks, changes in color). When a car mechanic tries to provide evidence to the car's owner that brakes need to be replaced, he will point to worn brake pads as his evidence. In each of these cases, evidence is a physical object that can be directly apprehended by the senses. The notion that evidence consists of physical objects is therefore a very intuitive one for most people.

Some philosophers who could be characterized as classical foundationalists (e.g., Russell, 1910-1911) provide arguments to justify the adoption of such an account of evidence, beyond its intuitive appeal. Classical foundationalism is based on the premise that if all knowledge is justified by inference, an infinite regress of such justification arises (Fumerton, 2010). For example, if a proposition is inferred from a premise, that premise again needs to be inferred from another premise, etc. Classical foundationalists argue that the only way to terminate such an infinite epistemological regress is to posit some foundational or basic knowledge upon which the justification of other inferences can be based. According to foundationalists then, there is some basic non-inferential knowledge that provides justification for other inferential knowledge and thus forms the basis from which all knowledge is derived. Descartes' meditations provide one example of classic foundationalism. Descartes relied on rationalist assumptions in his account. In contrast, some foundationalists argue that basic non-inferential justification is provided by sense data gained from physical objects. For example, according to Russell (1910-1911), knowledge by direct acquaintance through sense data forms the foundation of all knowledge or, in other words, the basic evidence

on which all knowledge can be based consists of physical objects known through direct acquaintance via the senses. In recent years, many foundationalists have taken a more moderate position, arguing that evidence can be non-propositional, rather than that all evidence is non-propositional (e.g., Conee & Feldman, 2004; Plantinga, 1993)

Propositional evidence. Contrary to accounts that conceptualize physical objects as evidence, many recent philosophers argue that all evidence is propositional (e.g., Williamson, 1997). Propositions, in the context of philosophy "are the sharable objects of the attitudes and the primary bearers of truth and falsity. This stipulation rules out certain candidates for propositions, including thought- and utterance-tokens, which presumably are not sharable, and concrete events or facts, which presumably cannot be false" (McGrath, 2012, n.p.). Williamson (1997), a main proponent of a propositional account of evidence, equates evidence with knowledge. He argues that evidence is what justifies belief and that evidence must be known, concluding that evidence is equivalent to knowledge. He further posits that knowledge, which constitutes evidence, must be propositional. To support this point, he examines the way in which the concept of 'evidence' is employed in ordinary language. Evidence, as per Williamson, is explained by hypotheses (in fact the best hypothesis is chosen based on the fact that it best explains the evidence). Hypotheses explain why something is the case. In order to explain why something is the case, however, one must first grasp that it is the case. In Williamson's words, "One can use an hypothesis to explain why --- only if one grasps the proposition that ---. Thus only propositions which one grasps can function as evidence in one's inferences to the best explanation. By this standard, only propositions one grasps count as part of one's evidence" (p. 726). In addition, evidence is often described as being inconsistent. According to Williamson, the adjective "inconsistent" can be applied meaningfully only to propositions. There is no meaningful way, for example, to speak of physical objects as being inconsistent (e.g., the tire tracks and footprints were inconsistent), but propositions about physical objects (e.g., the propositions that the tire tracks implicated Paul in the crime, and the proposition that the footprints implicated John) or any other propositions can be inconsistent with each other. Finally, Williamson points out that a central function often attributed to evidence is that of probabilistic confirmation. Only propositions, however, can have probabilities attached to them in the sense that a probability is a probability that something is or has occurred.

In response to accounts that equate sensory experiences of physical objects to evidence, Williamson (1997) argues that of course physical objects and sensory experiences are non-propositional. However, he argues that they provide propositions and that the propositions they provide are in fact the evidence. For example, a person could describe the shape of a mountain, and from the shape infer that it is one mountain or another. The evidence in this case is the proposition that the mountain is a given shape, rather than the mountain itself or one's perceptual experience of it (which after all includes much more than just its shape). Physical objects and perceptual experiences thus are conceived of as sources of propositional evidence, according to Williamson, but are not in and of themselves evidence.

Relevance to evidence-based practice. The debate between propositional and non-propositional conceptions of evidence of course maps well onto the EBP discourse, particularly the various views regarding what component of research can be conceptualized as evidence, but also the various conceptualizations of evidence in physical terms and in terms of knowledge/beliefs and facts. To recap from the previous chapter, even among those participants in the EBP discourse who appear to agree that evidence consists of research, there is no clear agreement with respect to what component of the research constitutes the evidence. Some argue that research designs provide evidence in the form of research data or that evidence is constituted by findings contained in research literature. Research designs themselves, of course, are not propositional since they are neither shareable nor bearers of truth and falsity. The notion that the implementation of a particular research design in a given setting potentially provides evidence, however, is consistent with a propositional account of evidence, and in fact quite reminiscent of Williamson's argument that perceptual experiences provide evidence in the form of propositions. Data are also usually propositional, in that propositions about a data set are shareable bearers of truth and falsity. The findings or claims in research literature are also best conceptualized as propositional and thus propositional accounts of evidence are consistent with views that findings in research literature constitute evidence rather than the literature (e.g., the article as a document) itself.

Given the fact that only propositions can be assigned probabilities, only evidence conceptualized as propositional can be thought of as being related to its object in a

probabilistic relationship. If one conceptualizes 'evidence' as the research design itself, as some proponents of randomized-controlled trials do in a manner that may betray careless use of language, or conceptualizes it as the documents or physical research literature, again in what may well result from careless use of language, there is no meaningful way that one can speak of its probabilistic relationship to a certain object (e.g., a treatment approach). Only those perspectives that conceptualize 'evidence' as beliefs, knowledge, or facts are consistent with propositional and thus probabilistic accounts of evidence. Those accounts that imply that evidence is a physical entity cannot also plausibly conceptualize the relationship between evidence and its object as probabilistic.

Internalistic versus Externalistic Evidence and Phenomenal versus Realist Evidence

A second major divide among philosophers of evidence separates the so-called 'evidential internalists' (e.g., Feldman & Conee, 2001) from the so-called 'evidential externalists' (e.g., Goldman, 1979; Williamson, 2000). At a most basic level, internalist accounts posit that all justification for belief is a function of internal states (Cruz & Pollock, 2004). In other words, a person's evidence includes that and only that of which the person is aware. Externalist accounts on the other hand insist that there are external constraints on justifications for beliefs. In externalist accounts, then, there can be evidence that a person is not aware of, which nonetheless constitutes evidence. In addition, erroneous or unreliable beliefs that do not match external reality also cannot serve as evidence in externalist accounts. For example, given an internalist account, smoke at a distance would be evidence for fire for a person who sees the smoke but not for a person who is blind and thus cannot see the smoke. In an externalist account, however, such smoke is evidence for both people. In addition, if a person mistakenly perceived steam at a distance as smoke and inferred the existence of a fire, the "smoke" would constitute a basis for justified belief and thus evidence from an internalist perspective, but not from an externalist perspective.

Internalist and externalist accounts of evidence map to some extent onto phenomenalist versus realist accounts of evidence (e.g., Kelly, 2008). Phenomenalist

accounts assume that evidence must be, epistemologically, easily accessible to the subject in order to serve a function of underwriting knowledge, since one can know only that which one can experience in some manner. Realist accounts, however, point out that much of the subject matter of science (e.g., electrons, atoms, etc.) is not part of a person's phenomenological experience. Nonetheless, knowledge about these subjects is presumably based on some version of evidence. Realist accounts of evidence therefore posit some theoretical connection between phenomenological experiences and the external world and treat only those phenomenological experiences that are related to or brought about by the external world as evidence.[31]

The distinctions between internalist versus externalist and phenomenal versus realist views of evidence have important implications for evidence-based practice, particularly the more objectivist compared to the more social constructionist accounts of evidence espoused throughout the evidence-based practice discourse. In the previous chapter, accounts that posit that evidence is objective and knowable independently or is empirically consistent across time and social contexts were contrasted with other accounts that tend to emphasize the extent to which evidence is employed in a process that may well be unique to a given social context, gain its meaning in that social context and is therefore socially constructed rather than objectively given.[32] If one interprets internalist and phenomenalist accounts of evidence as individualist or subjective, one could argue that both kinds of accounts are externalist and realist. The external constraint on evidence in objectivist accounts is an objective reality that is separate from social contexts, while the external constraint on evidence in social constructionist accounts is the social reality given by the context in which the evidence is employed. A

[31] This distinction is further articulated in the context of broad movements in philosophy of science, namely the shift from logical positivism to empirical realism, which will be discussed in more detail later in this chapter.

[32] It is important to note here that social constructionist accounts are not necessarily anti-realist or subjectivist in the sense that evidence in such accounts does not differ for each individual. Rather, the constraint placed upon evidence in social constructionist accounts is social in nature, such that individuals in different social contexts might be justified in holding different views. Individuals in the same social context in such accounts however would not be justified in holding different views and individuals are not at liberty to choose and define their evidence at their own discretion. The constraint placed on evidence in objectivist accounts however is taken to be more objectively given independently from social contexts or the current state of knowledge.

more accurate reading of externalist and realist accounts of evidence however, emphasizes that evidence exists a priori and need not be known by anybody in order to count as evidence, while in internalist and phenomenalist accounts it becomes evidence once it is known and applied towards a certain end as evidence. To provide a brief example, DNA evidence, according to strictly realist, objectivist, and externalist accounts, was evidence before the scientific discoveries related to it prompted its utilization in various social contexts. Discoveries related to genetics and DNA permitted the use of DNA as evidence, but prior to this point it was simply evidence that was not yet known. More social constructionist, internalist, and/or phenomenalist accounts of evidence, however, would posit that DNA evidence only became evidence once the necessary context of knowledge and social practices had emerged that permitted its employment as evidence of a particular kind. Prior to these discoveries and the emergence of these social contexts, DNA was meaningless and therefore it could not play the role of evidence. To recap, those accounts that emphasize that evidence objectively exists and is discovered through the activities of science are more consistent with strict externalist, objectivist, and realist perspectives on evidence, and those that emphasize that evidence becomes evidence when it is employed in a given social context towards a given social end in the context of a social practice are more consistent with internalist and phenomenalist perspectives on evidence.

Whether one conceptualizes 'evidence' in an externalist and objectivist manner or an internalist, phenomenalist, or social constructionist manner will have implications for the roles evidence can serve in EBP. Most of the EBP discourse acknowledges the importance of considering a variety of social and contextual factors in the application of evidence, as is clear from the inclusion of the third leg (patient values and preferences) of evidence in the EBP discourse and the numerous published articles on the impact of cultural diversity on the interpretation of evidence in EBP. At the same time, however, one of the functions that evidence is at times called upon to serve is that of adjudicating between competing claims, theories, or hypotheses. An externalist, objectivist account of evidence may allow for evidence to provide such adjudication independent of social contextual factors. Based on an internalist, phenomenalist, and/or social constructionist account of evidence, however, evidence can only serve such an adjudicative function within a given social context. A relatively high degree of complexity may therefore be

necessary in EBP in the sense that what may constitute evidence in one particular social context may not constitute evidence in a different context or different conclusions may be justified in different contexts. It is also often argued that EBP can standardize treatment delivery. Such standardization can only occur at the level of a given social context based on a social constructionist view. Indeed, given the process of EBP that is advocated by the Task Force, it is difficult to see how evidence could serve such a unifying and objective role, especially given the integration of evidence with clinical expertise and patient characteristics, which appears to make EBP unique to each and every clinical situation. It therefore seems that the EBP discourse could benefit from greater clarity with respect to whether 'evidence' is to be conceptualized as objective and independent from the social context in which it is applied or as contingent on the social context in which it is employed, all of which has implications for the functions that can meaningfully be attributed to evidence.

Bias versus Evidence

A philosophical distinction that is closely related to the contention between those who conceptualize evidence as objective and independent from social context and those who view it as more socially constructed, is that of 'evidence' versus 'bias.' This distinction is reflected in Bayesian accounts of probability. Bayesians, in very broad terms, conceptualize the probability of a proposition as the credibility of the proposition (Strevens, 2006). Objectivist Bayesian accounts conceptualize probability as a means to quantify belief warranted by standards of logic and consistency while subjectivist Bayesians conceptualize probability as a quantity representing personal belief.[33] Evidence, according to Bayesians, increases the probability of the truth of a proposition compared to the probability of its truth prior to the consideration of relevant evidence. The probability prior to the consideration of a piece of evidence is called the prior probability in Bayesian language, while the probability conditional on that piece of evidence is called the posterior probability. Posterior probability is a function of both the prior likelihood of a proposition and its likelihood given the evidence. For Bayesians, the

[33] It is important to note that personal belief for Bayesians is best conceptualized as the belief warranted for a person in a given context rather than any belief held by an individual.

prior probability of a proposition, before it becomes conditional on any piece of evidence, is bias. That which alters its probability from the prior probability, on the other hand, is evidence (Joyce, 2005). Persons' inherent or pre-conceived beliefs about the probability of a proposition prior to encountering a piece of evidence, or their bias, may be the result of various factors, including general knowledge, past evidence, or prior experiences. Bayesians therefore quite explicitly assign a very central role to the prior judgment of the probability of a proposition in conceptualizing evidence.

One of the implications of the Bayesian interpretation of 'evidence' is that, because posterior probability is a function of bias and evidence, in cases of extremely low or extremely high prior probability, evidence will have much less impact on the posterior probability than in cases where prior probability is more moderate. For example, a weather forecast that predicts a very high likelihood of snow for the following day will have a greater impact on judgment of the likelihood of snow for the following day in December than it will in the middle of August, for a person residing in southwestern Ontario. The prior probability of snow in August is so low, given prior knowledge of weather patterns in the region and experiences of past summers that even a high probability prediction by the weather forecast will have little impact on one's expectations. Bayesians thus accommodate prior knowledge and experiences, general and deductively derived principles, or even possibly personal beliefs and superstitions held in particular circumstances, in the impact of evidence on a given hypothesis or proposition.

In the context of the EBP movement, Bayesian conceptions may shed some light on the debate about the 'three-legged stool' of evidence, or more specifically, the integration of best evidence with clinical expertise and patient characteristics and preferences. Debates with respect to these components and their role in EBP summarized in the previous chapter relate to whether all three components ought to be conceptualized as evidence, or how evidence is to be integrated with the remaining two components of EBP (i.e., clinical expertise and patient characteristics and preferences) if these two components are not taken to constitute a form of evidence. From a Bayesian perspective, best evidence (or research evidence) could be conceptualized as bona fide evidence. Clinical expertise and patient characteristics and preferences on the other hand may be two factors that contribute to the prior probability distribution and may thus

be conceptualized as bias. In more concrete terms, a clinician who encounters a given patient, based on current clinical experience and knowledge, her knowledge and understanding of the patient's individual differences, and the patient's expressed preferences, may have a notion of a favored treatment approach. Additional or new research evidence that is relevant to the situation may alter the clinician's belief concerning the probability of success of a previously favored treatment, either positively or negatively. Bayesian perspectives are therefore consistent with those accounts in EBP that separate clinical expertise and patient characteristics and preferences from the evidence, but still value their role in EBP. Bayesian perspectives are inconsistent with accounts that conceptualize all three components as bona-fide evidence. Only Bayesian conceptions can accommodate a relationship between evidence and its object that is characterized by an increase or decrease in probability of a hypothesis conditional on the evidence relative to its unconditional or prior probability. Therefore, only accounts that separate evidence from clinical expertise and patient characteristics and preferences are strictly consistent with a probabilistic conception of the relationship between evidence and its object. Hence again, the way in which 'evidence' is conceptualized has implications for the kind of relationship in which evidence can stand relative to its object.

Indicative versus Normative Evidence

Lastly, philosophy also distinguishes between indicative and normative accounts of evidence (Kelly, 2008). Indicative accounts emphasize the probabilistic relationship between evidence and its object (e.g., Joyce, 2005). They are based on the notion that evidence is reliably indicative of that for which it is evidence (e.g., Hacking, 1975). Indicative accounts of evidence can be traced back to Antiquity (see Chapter 2) when evidence was often conceptualized as a sign that is more easily accessible than that for which it serves as a sign, and one that reliably predicts that for which it serves as a sign. For example, smoke provides indicator evidence for fire in a distance when the smoke is more easily visible than the fire and consistently predicts fire. This account of evidence is intuitively consistent with the manner in which evidence is conceptualized in much of ordinary language and many social practices. For example, the lipstick on a man's shirt collar and the smell of a woman's perfume on his shirt are said to provide evidence of

infidelity (it is more easily observable than the infidelity itself and increases the probability of judging an infidelity to have occurred relative to the probability of doing so in the absence of such evidence); the increased temperature of a person's forehead is said to provide evidence of infection (the infection may not be observable, but fever reliably indicates infection). Such accounts, however, have been criticized by some (e.g., Achinstein, 1978) for setting a very low threshold for evidence, or making the relationship between evidence and its object quite weak. For instance, buying a lottery ticket certainly increases the probability that one will win, but hardly constitutes evidence that one will win (see Achinstein, 1978). In more reliabilist terms, as the famously overused example from introductory statistics textbook reminds us, ice cream sales are reliably correlated with increases in drowning incidents (due to their shared relationship with warm weather), yet it seems farfetched to treat ice cream sales as evidence for drowning incidents.

Normative accounts of evidence on the other hand emphasize that evidence provides reasonable justification or a good reason for a conclusion or belief (e.g., Harman, 1986). According to such accounts, there must be some logically adequate explanatory connection between the evidence and its object. Only that which grounds justified belief is evidence. The standards of adequacy here are based on norms of logic or reason. Normative accounts can make sense of the ice cream example used above. An increase in ice cream sales does not appear to be a good reason to believe that there will be incidents of drowning. The two are not logically connected in a manner that makes sense to most people. There is a connection between normative accounts and the concept of bias described above (which is typically associated with Bayesian reliabilist accounts of evidence) in the sense that what one is justified to believe depends on one's prior knowledge (Kelly, 2006). For example, persons who witness the optical illusion of sitting in a train and watching another train take off in the opposite direction for the first time may well be justified in believing that their own train is moving, since the evidence of their perception supports such a belief. However, a person who knows about this optical illusion and takes trains routinely would not be justified in this belief. In the same sense then, normative accounts of evidence also appear to be more consistent with internalist perspectives of evidence. What one is justified to believe varies across individuals based on their background knowledge, which in turn is contingent on their

sociocultural context. In addition, one may well be justified in holding a certain belief that may be incorrect, as illustrated in the above example. Whether incorrect or false evidence (e.g., the perception of one's train moving in the above example) should still be considered evidence is a contested issue among philosophers (Kelly, 2006). Some argue that evidence must be veridical in order for it to play any meaningful role in the establishment of knowledge, and therefore reject a strictly normative conception of evidence.

The debate between indicative and normative evidence appears to have some relevant implications for the EBP discourse. First, a normative account of evidence, which allows for the possibility that evidence may justify beliefs that are incorrect or that misleading or false evidence can be considered evidence, appears to be highly problematic for EBP. According to much of the EBP literature, the evidence to be employed in EBP carries the promise of ensuring a certain standard of quality for treatments, facilitating the use of the most effective treatments, and discriminating among practices that work and do not work (e.g., Cape & Barkham, 2002; Stricker, 2006). The notion that evidence can be misleading or can justify incorrect beliefs therefore runs counter to the very purpose and promise of the EBP movement. The second implication of a normative conception of evidence, that background knowledge may lead to differences in the beliefs that a person is justified in holding based on evidence also creates serious challenges for EBP and particularly its promise to standardize treatments across providers and reduce variations in treatments provided (e.g., Stricker, 2007). If clinicians in different social contexts are justified in different conclusions based on the same evidence, evidence is unlikely to standardize practice.

Nonetheless, much of the EBP literature appears to assume that only beliefs (or treatments, or clinical practices, or assessment practices) based on evidence are justifiable. Some go as far as to argue that the justification of practices using evidence is an ethical mandate for clinicians (e.g., Veerman et al., 2007). Such accounts therefore do imply a certain normative conception of evidence. However, the notion that evidence can be anything that justifies belief also runs counter to the very narrow definitions of evidence (e.g., results from randomized controlled trials) that are often advanced in the EBP discourse. A normative account of evidence suggests a much broader notion of evidence that includes anything that can justify a belief (e.g., casual perception,

testimony by others, etc.). EBP therefore seems to be premised on normative conceptions of evidence that may be quite incompatible with the aims of EBP.

In fact, many of the conceptions of evidence employed in the EBP discourse, including the privileged role assigned to certain methods that are thought to yield reliable and reproducible evidence (e.g., randomized-controlled trials), also imply an indicative account of evidence. The use of data and statistics as a source of evidence, the suggestion that evidence can standardize treatment across providers, and the notion that evidence can get at truth all imply an indicative conception of knowledge. The notions that various research designs provide more or less reliable means of examining clinical practices and that evidence can therefore be ranked in strength based on the manner in which it is generated, as suggested by "hierarchies of evidence," also seem more consistent with an indicative account of evidence. Instead of exclusively adopting an indicative conception of evidence, however, it appears that the EBP discourse flexibly moves between indicative and normative accounts of evidence and may actually confound them. According to much of the EBP literature, only evidence that reliably, consistently, and specifically supports a practice can be considered evidence and can justify that practice. Evidence here is then taken to be both indicative and normative. Treating normative and indicative conceptions of evidence as one and the same may allow EBP to benefit from the very intuitive appeal of a normative conception of evidence, without having to broaden evidence to include all justifications of belief, or conceding to internalist accounts of evidence. In line with a normative account of evidence, EBP can argue that practices that are not based on evidence are haphazardly based on arbitrary choice or personal preference, and therefore not justifiable, while also defining that which can be considered evidence much more narrowly and much more consistently with indicative accounts. In addition, only indicative evidence can reasonably be used to adjudicate between competing claims in different social contexts, since based on normative evidence two people may be justified in holding contradictory and inconsistent beliefs if they exist in different sociocultural contexts. For example, people may have been justified in their belief that the earth was flat prior to the Copernican revolution, based on the evidence available at the time, but would be justified in their belief that it is round following the change in available evidence during the Copernican revolution. Consequently, if evidence is to adjudicate between various

practices in EBP, without being relativized to a particular social context, it must be conceptualized as indicative evidence.[34]

What is the Relationship Between Evidence and Its Object?

Philosophers have characterized the relationship between evidence and its object in a variety of ways. Philosophical treatments of the relationship between evidence and its object recognize that the manner in which this relationship has been characterized restricts the ways in which evidence can be conceptualized. Given that many of the philosophical conceptualizations of the relationship between evidence and its object map quite well onto the various ways in which this relationship is conceptualized in the EBP discourse, philosophy can provide some conceptual clarification of the different accounts of this relationship that are advanced in the EBP discourse. More importantly, however, philosophy may be able to clarify the implications that characterizing the evidential relationship in EBP in various ways may have for the kinds of roles evidence can play and the kinds of things that evidence can be.

Quantitative versus Qualitative Evidential Relationships

Philosophical perspectives on the relationship between evidence and its objects can be broadly divided into qualitative and quantitative perspectives (Carnap, 1950; Hempel, 1945, cited in Kelly, 2006). Quantitative perspectives on evidential relationships characterize such relationships in terms of probability and reliability. Qualitative accounts on the other hand posit explanatory or deductive relationships. Some mixed perspectives on the relationship between evidence and its object have also been proposed (Achinstein, 1978).

Quantitative relationships. Quantitative perspectives on the relationship between evidence and its object characterize the relationship as graded. They tend to

[34] It must be noted here that the philosophical binaries explored in this work may also be unnecessarily restrictive and that rather when it comes to psychological phenomena, such restrictive and exclusionary alternatives may not be necessary. A detailed exploration of this issue is beyond the scope of this paper but may warrant future attention.

focus on the extent to which evidence supports its object as a matter of degree (Kelly, 2006). Reliability and probability have both been suggested as means to quantify the relationship between evidence and its object.

Reliability relationships. Accounts of evidence often ascribe two properties to evidence. Evidence must be more easily accessible than that for which it is evidence (epistemological accessibility) and evidence must consistently be related to that for which it is evidence (reliability). The latter attribute stipulates a relationship between evidence and its object based on reliability. Early perspectives focused on reliability (e.g., Dretske, 1971) and posited that justified beliefs are based on reasons (evidence) that reliably predict those beliefs. The previous discussion of indicative evidence already highlighted some of the limitations or potential criticisms of such accounts of evidence. More recent philosophical treatments that focus on reliability as a relationship between evidence and its object turn their attention, instead, to the process whereby evidence is generated. According to a set of theories known as process reliabilism (e.g., Goldman, 2006), the process used to arrive at a belief from particular experiences must be 'generally reliable' or 'truth-conducive.' These accounts usually do not quantify or stipulate cut-offs for such 'general reliability' but rather emphasize the fact that some cognitive processes of arriving at conclusions from evidence are more 'truth conducive' than others and that such processes lead to more justified beliefs. In such accounts, reliability pertains to the process through which evidence is generated or used as opposed to the relationship itself. The process itself must be generally reliable, rather than reliable in a particular case. However, it has been pointed out that such accounts require some means of quantifying how truth conducive certain processes are, which in turn requires knowledge of truth independently from evidence. It has been argued that if one were in a position to have such knowledge, one would also no longer require evidence. In other words, if it is clear how truth-conducive a process is, one likely would not require evidence for it, since one would already have certain knowledge.

Probability relationships. Most commonly, the relationship between evidence and its object has been characterized in terms of probabilities. Two variations of such accounts have been put forth. The earliest probabilistic accounts (e.g., Carnap, 1950) are premised on the notion that evidence must make a conclusion highly probable in order to serve as evidence for it. In such accounts, a cut-off for the probability of a

conclusion can be stipulated (e.g., in order for e to serve as evidence for h, the probability of h given e must be greater than k) (DiFate, 2007). Such accounts have been criticized for their lack of specificity. Achinstein (1978), for example, points out that the probability that a man who has been eating a certain breakfast cereal will not get pregnant is extremely high. Yet, to argue that his eating this breakfast cereal is evidence for the fact that he will not get pregnant is unsound.

A more complex probabilistic account has been proposed by Bayesians, and has already been described in the above discussion of evidence and bias. To recap, according to Bayesian accounts, evidence changes the probability of a conclusion relative to its probability in the absence of the evidence. Evidence is therefore characterized by an increase in probability relative to a previous probability. This solution appears to resolve the breakfast cereal dilemma: the probability of the cereal-eating man avoiding pregnancy was high irrespective of his cereal eating habits and remains unchanged by his choice of cereal. Therefore, his cereal choice is not evidence for his avoidance of pregnancy. However, Bayesian accounts are also not without problems. Achinstein (1978), for example, points out that buying a lottery ticket changes the probability that one will win the lottery and yet hardly constitutes evidence that one will win.

Qualitative relationships. Qualitative or categorical perspectives on the relationship between evidence and its object tend to characterize this relationship in binary term. According to such perspectives, evidence either does or does not support a conclusion (Kelly, 2006). Two prominent versions of these perspectives include deductive and explanatory perspectives.

Deductive relationships. The most popular qualitative accounts of the relationship between evidence and its object posit that evidence that supports a claim can be deduced from that claim. According to Popper's (1963) famous hypothetico-deductive model, predictions can be deduced from a hypothesis and then tested. If the predictions hold true, they constitute evidence for the hypothesis and if they do not, they provide evidence against the hypothesis. This model is often used as a description of the scientific method and much scientific discourse implies the use of this model. Most research reports in psychology are structured based on this model: hypotheses are

stated, tested, and either supported or rejected. One criticism of this perspective that is often advanced is that the same evidence can be deduced from multiple hypotheses and that evidence could therefore in theory support multiple contradictory hypotheses (DiFate, 2007). In this model then, evidence cannot necessarily adjudicate among multiple hypotheses.

Hempel (1965) advanced a more narrow deductivist model. According to Hempel's deductivist-nomological model, evidence must be a positive instance of a universal hypothesis. For example, a white swan is an instance of the hypothesis that all swans are white. A problem with Hempel's model, however, according to Glymour (1975) is that if evidence is to be observable, it cannot be a positive instance of a hypothesis that contains unobservable theoretical terms.[35] Glymour argues that it is precisely hypotheses containing such theoretical terms that are commonly tested in science, and therefore Hempel's account cannot be applied to scientific studies.

Explanatory relationships. Explanatory models of the relationship between evidence and its object have already (at least in part) been outlined in the description of normative evidence provided above. According to explanatory models, evidence provides a good reason or a justification for holding a belief. Such models provide a solution to some of the limitations of both reliabilistic and probabilistic quantitative models described previously. Neither the purchase of a lottery ticket nor the eating of a certain breakfast cereal provides justification or a good reason for the belief that one will win the lottery or avoid pregnancy, respectively. Therefore, explanatory accounts disqualify anything that cannot provide a good reason from serving as evidence. The determination of what constitutes a good reason, however, is necessarily dependent on certain background knowledge (e.g., factors that lead to pregnancy, the odds of winning the lottery) and therefore, as previously stated, such accounts are limited in that they may permit variations in what can constitute evidence in different circumstances.

Mixed relationships. Upon recognizing the limitations of both quantitative and qualitative accounts of evidential relationships, some philosophers propose conceptions

[35] Note that in the language used by Glymour, what is observable is the positive instance of a theoretical term or, in other words, an entity denoted by a theoretical term.

that combine these two approaches. Achinstein (1978), for example, proposes an account that relies on both objective probabilities (i.e., evidence must make a conclusion highly probable) and explanatory relationships (i.e., it must also provide a good explanation of the conclusion). In doing so, however, his account becomes limited by many of the same constraints that apply to explanatory accounts in that it relies on a clear conceptualization of a standard for what constitutes a good explanation, which itself raises problematic issues with respect to the relevance of background knowledge. In fact, since an internalist account of evidence that permits for contextual variations in justifiable evidence is inconsistent with an indicative account implied by Achinstein's first requirement of reliability, it appears that he must be implying externalist standards for adequate explanations. The nature and origin of such standards is, however, quite elusive, and if they are thought to be given by social norms or contexts, again evidence and conclusions warranted become contingent on these social contexts, limiting the kinds of universal standardized conclusions that proponents of EBP often seem to wish to draw from evidence.

Implications for EBP. Both quantitative and qualitative conceptualizations of the evidential relationship are represented in the EBP discourse, as described in the previous chapter. Accounts of EBP that emphasize the extent to which a consideration of evidence can standardize clinical practice appear to rely on quantitative and particularly reliabilist conceptions of the relationship between evidence and its object. Process reliabilism appears to be particularly aligned with perspectives on EBP that emphasize the use and ranking of different methods of generating evidence according to their ability to support practices in a consistent manner. Philosophical commentaries of reliabilist accounts, however, also serve as a good reminder that evidence or a process of generating evidence may well be reliably related to an outcome without providing a good explanation of this outcome. In fact, EBP proponents are well aware of this limitation and therefore more commonly advocate the use of evidence to support a specific practice. The central role of randomized-controlled trials as a source of evidence in EBP therefore seems much more consistent with Bayesian probabilistic accounts of evidence in that control conditions are set up to establish that a practice increases the probability of a desired outcome over and above its probability in the absence of the practice. Unfortunately, the statistics usually employed in randomized-

controlled trials rely on a significance testing approach that is more consistent with a stipulated cut-off for adequate probability (as set by adopting a particular alpha level in most statistical tests). Probabilistic conceptions of evidence, as previously stated, also assume a propositional account of evidence. All quantitative accounts, as they are typically employed in psychological inquiry, are also most consistent with indicative and externalist conceptions of evidence.

Some of the EBP literature on the other hand, conceptualizes the relationship between evidence and its object as dichotomous in a manner that appears to be more consistent with explanatory or deductivist conceptions of evidence. Explanatory accounts of evidence allow for a much broader conception of evidence. At the same time, however, in the absence of objective standards for what constitutes good evidence, such a conception is much more consistent with an internalist account of evidence and limits the extent to which evidence can be used to adjudicate between competing practices, a role that it is called upon to serve in EBP. Although at times objective standards to adjudicate what constitutes good evidence are suggested by those who view EBP as equivalent to empirically supported treatments, such standards themselves are open to debate. Deductivist accounts on the other hand are also limited in that they cannot provide definitive support for any particular hypothesis or practice. Finally, although the representation of both qualitative and quantitative conceptualizations of evidence may suggest that a combined account such as Achinstein's is most consistent with EBP, such an account raises theoretical inconsistencies and further questions in terms of how to define good explanations that have yet to be answered in the context of EBP.

Requirement of Total Evidence

Philosophers have also been concerned with the relationship between evidence and its object when the evidence is inconsistent and supports multiple contradictory conclusions. Some philosophers (e.g., Ayer, 1957) have also questioned how one determines the probability of a conclusion, given that the addition of new evidence can increase or decrease the probability of said conclusion when it is unclear if earlier or later supported probabilities are superior. To address these issues, according to some

philosophers, degree of confirmation must either be based on all available evidence or a subset of the evidence that provides the same degree of confirmation as the total evidence, in which case the remaining evidence is redundant (Hempel, 1960). This requirement resolves the difficulties laid out above in the sense that when the addition of new evidence alters the degree of confirmation of a conclusion, the degree of confirmation that is based on more evidence is always superior. In addition, if one set of evidence supports one conclusion and another set of evidence supports a contradictory conclusion, based on the requirement of total evidence, the degree of support for both conclusions has to be based on all evidence and is thus less than the degree of confirmation offered by one or the other set of evidence (since the other set disconfirms each respective conclusion). This requirement of total evidence highlights the difficulties inherent in considering only a subset of evidence that supports a conclusion: there may well be other evidence that disconfirms the same conclusion or the evidence in support of the conclusion. Therefore a multitude of contradictory conclusions can be supported if one allows for considerations of only some of the evidence.

The requirement of total evidence has some implications for EBP. First of all, some proponents of EBP directly stress the importance of considering all relevant evidence in EBP. In addition, meta-analysis is often considered to be privileged evidence in EBP with the assumption that meta-analysis usually summarizes all available relevant evidence.[36] Nonetheless, the requirement of total evidence does raise serious theoretical and practical problems with respect to certain parts of the EBP discourse. First, on a theoretical level, certain dichotomous conceptions of EBP that closely mirror requirements laid out in the empirically supported treatment movement become quite problematic in light of this requirement. For example, it is often suggested that a certain number of studies of a certain type (e.g., two randomized-controlled trials) are required in order to make a particular clinical practice evidence-based. Based on this requirement, contradictory and inconsistent clinical practices could all be considered supported by a subset of the evidence. Based on the requirement of total evidence,

[36] Of course, the well-known file drawer problem in meta-analysis (i.e., the fact that positive findings are more likely to be published than null findings and therefore also more likely to be included in meta-analyses) is problematic precisely because of the requirement of total evidence. Data collection and statistical methods have been devised in meta-analysis that are thought to address this issue to some extent.

however, the existence of studies supporting multiple inconsistent practices in fact weakens the evidence for each individual practice. Therefore, any approach to EBP that is based on a consideration of only a portion of the evidence in inconsistent with the requirement of total evidence.

In addition, it may be difficult to determine what constitutes relevant evidence for any particular clinical practice. Does relevant evidence only include research evidence of a particular type, all the types of evidence listed by the Task Force, or all parts of the three-legged stool? Even if only treatment research provides relevant evidence, is only research conducted with a particular disorder, population, or in an identical setting relevant? Is research pertaining to the particular practice one is considering relevant, or should one also consider the evidence pertaining to all possible alternate practices? Are researchers obligated to make attempts to locate unpublished research in order to ensure they are not only considering research with positive findings? It is easy to see how many questions are raised by the requirement of total evidence and how quickly a consideration of total evidence may require the consideration of overwhelmingly large amounts of evidence. Here, practical concerns with respect to evidence become glaringly obvious if one endorses the requirement of total evidence. In the absence of constantly updated and thorough research summaries, the requirement of total evidence seems to place demands of a Herculean magnitude on the clinician, who is to ensure that practices are supported by the totality of evidence. Nonetheless, in light of the impact that considering only a subset of the evidence can have on conclusions reached, it is difficult to justify disposing of the requirement of total evidence for the purpose of EBP.

Balance, Weight, and Specificity

Bayesian accounts of evidence provide a further set of distinctions with respect to the relationship between evidence and its objects and expand on the requirement of total evidence. Bayesians argue that conceptions according to which the total evidence is simply "the sum of all those considerations that tell in favor of its truth" are misleading (Joyce, 2005, p. 158). Bayesians instead argue that the total evidence consists of three components: balance, weight, and specificity. They also argue that only one of these

three components is directly relevant to the credence of a proposition. The distinction between balance and weight was first advanced by Keynes (1921). In very simple terms, the balance of evidence reflects the degree to which the evidence supports a proposition. The weight of evidence on the other hand is a matter of the amount of evidence that is available. The specificity of evidence reflects the extent to which the evidence discriminates between incompatible alternative propositions or the clarity with which it supports a specific proposition, rather than being subject to ambiguity that may affect its support for a proposition (Joyce, 2005). According to Joyce, only the balance of evidence is directly relevant to the credence of a proposition. Knowing the likelihood of a proposition, however, provides absolutely no information about the volume of evidence on which this likelihood is based, or its weight. The weight of evidence, on the other hand, merely determines the extent to which support for a proposition is resilient to change through additional evidence (Skyrms, 1980). Finally, specificity affects the degree of certainty one can have with respect to the balance of a proposition. To provide a very practical example, borrowed from Kelly (2008), upon seeing a certain numbers of coin tosses with the same coin, I may decide that the balance of the proposition that the coin tossed will land heads up is 0.5. Tossing the same coin an additional 50, 100, 1000, or 10000 times may not change the balance of my evidence (assuming it's a fair coin) but will substantially increase the weight of my evidence. If after two coin tosses (one of which is heads and one tails) the third toss yields heads, I may alter my belief with respect to the likelihood of a result of heads on the next toss to 0.66. If, after 1000 coin tosses yielding equal heads and tails, the 1001st toss yields heads, however, I am unlikely to alter my beliefs about the next coin toss. Further, with respect to the specificity of evidence, after 1000 fair coin tosses, my evidence neither specifically supports the proposition that the next toss will be heads nor that it will be tails. Rather, it equally supports both propositions.

The distinctions between balance, weight, and specificity of evidence again have some implications for EBP. At times, the EBP discourse implies that a greater number of findings in support of a practice constitute "stronger evidence" for such a practice. In addition, practices supported by more studies are sometimes treated as more strongly supported by the evidence. Based on the distinctions laid out by Bayesian philosophers, both such claims warrant caution. In the case of a small number of studies that support

a very large effect size for a given practice for example, the evidence more decisively supports that practice over another practice that may be supported by a larger number of studies that provide evidence for a smaller effect size. The greater weight of the evidence for the latter practice merely makes it less subject to revision as additional evidence emerges. As well, if two contradictory practices appear to have the same effect sizes based on the available evidence, the evidence does not specifically support one practice over the other. The only sense in which larger quantities of studies can therefore be described as stronger evidence, is in the sense that they may be less subject to revision as additional evidence emerges. They do not, however, show that the particular practice they support is more efficacious. Rather, they show that the estimate of its efficacy based on the available research is more certain than the estimate of the efficacy of other practices. In the EBP movement, the confounding of balance, weight, and specificity may lead to some confusion with respect to the manner in which clinicians ought to choose between competing practices based on the evidence. In that sense, this distinction can therefore be very valuable in creating greater clarity with respect to the relationship between evidence based on the volume of available evidence, the support lent by the evidence, and the ability of the evidence to discriminate between incompatible practices.

The Object of Evidence

With respect to the object of evidence, philosophy of science can offer two conceptual frameworks that may be of use to psychologists concerned with EBP. First, distinct historical movements in philosophy of science, such as logical positivism, logical empiricism, and post-modernism, can help to clarify what inferences can be drawn from evidence and the extent to which scientists can go beyond evidence in conclusions they draw. Secondly, the philosophical distinction between empirical and conceptual issues, which was touched on in the introduction to this thesis, can aid in clarifying matters that can be adjudicated by evidence versus matters that cannot be subject to adjudication by evidence and thus may not be legitimate objects of evidence.

Inference Beyond the Evidence

Evidence serves as evidence for something else or in the words of Hacking (1975) "points beyond itself." Different philosophical views exist, however, with respect to that to which evidence can point. The different conceptions of evidence and its relationship to its objects summarized above and represented in the EBP discourse rely on commitments to different and incompatible conceptions of philosophy of science that have been advanced in the past century.[37]

Logical positivism. Logical positivist accounts of science adopt one of the most restrictive positions with respect to the extent to which one can generalize beyond observed evidence. Logical positivism was a movement in philosophy of science that was proposed by members of the Vienna Circle in the 1920's. Logical positivists aimed to rid science of all discussion of metaphysics. Science, according to the logical positivists, was to limit itself to the examination of only those statements that can be directly verified by observation (synthetic statements) and to leave the examination of statements that are true by definition (analytic statements) to philosophers and logicians. The meaning of a term or statement, for logical positivists, was its method of verification, and thus all scientific terms could be defined in terms of observation language. The logical positivists therefore also distinguished between propositions containing only observation and logical terms (which can be verified directly through observation) and propositions containing theoretical terms (which cannot be verified directly through observation) and eschewed the use of the latter.

The previous outline is rather cursory but suffices as a basis to delineate the implications that a logical positivist philosophy of science will have for 'evidence.' For logical positivists, evidence has to consist of observations. Observations furnish evidence for particular analytic claims (e.g., all swans are white) but cannot provide evidence for anything beyond that which is observable. In other words, no theoretical claims beyond the observable can meaningfully be made in science. For example, in a logical positivist framework, the observation of an object dropping from height may

[37] Only a very cursory sketch of each of these movements is possible here. Countless detailed discussions of these movements exist in the philosophy of science literature – e.g., Friedman (1999); Lyotard (1984), and also in the theoretical psychology literature (e.g., Slaney, 2012).

provide evidence for the claim that objects drop from heights. Such a drop can provide evidence for gravity only if one takes gravity to mean only that objects will drop from heights but not if one wants to make theoretical claims about gravity's nature (e.g., as a force). No additional realist or existential meaning can be assigned to the concept of gravity and the object dropping cannot be taken as evidence for greater theoretical claims about gravity in a logical positivist framework. Logical positivism therefore takes one of the most restrictive perspectives on the extent to which one can go beyond the evidence in one's conclusions.

The evidence in logical positivism is evidence for observational statements, the latter of which must be strictly and directly observable and testable in light of publically observable fact (i.e., evidence, from this perspective). The only theoretical terms admitted by logical positivists are those that are used to heuristically summarize classes of observations. Evidence must be directly deducible from the conclusion and testable through observation. A logical positivist account of evidence is therefore quite consistent with more normative, propositional but also more deductive accounts of evidence. To logical positivists, the limitations of deductivist accounts of evidence laid out above, in terms of the ability to draw broad theoretical conclusions, are not limitations at all, but practices that are quite consistent with their view of the nature of scientific inquiry. According to logical positivists, science should only concern itself with that which is directly testable and observable.

Logical empiricism/empirical realism. Logical empiricism and what later become known as empirical realism, was a movement that emerged in the 1930's and 1940's in response to criticisms of the stringent requirements of logical positivism (Slaney, 2012). Logical empiricists rejected the requirement that all content of science be reducible to observation terms and that the only admissible theoretical terms in science were those that merely summarized classes of observables. Logical empiricists extended the content of science to theoretical terms that refer to unobservable, but in principal observable entities (e.g., microbes), which they separated from those entities that were in principle unobservable (e.g., God's will). They argued that real, currently unobservable, in principal-observable entities can be indirectly and causally related to observables through networks of propositions that connect unobservable and observable entities to each other. Feigl (1950) called such networks that connect observables with

each other, unobservables, and unobservables to other unobservables, nomological networks. Theories in science, according to Feigl, consist of such networks of propositions.

In logical empiricism then, evidence consists of observables, which are causally related to unobservable but real theoretical or hypothetical entities. Evidence provides indicators of theoretical entities and the theoretical entities can serve to provide explanations of the observable evidence. As science progresses, logical empiricists hoped that nomological networks would become increasingly clearly articulated, more and more densely populated, and thus the meaning of the theoretical terms that cause observables in the networks would be increasingly well-established. To empirical realists then, the meanings of scientific statements were not reducible to observation terms but rather based on real entities, which they denote.

Logical empiricist accounts of evidence are therefore much more consistent with indicative conceptions of evidence and also reliabilist conceptions of the relationship between evidence and its object. Evidence, in a logical empiricist framework is causally related to theoretical, unobservable entities and in turn, theoretical entities explain the evidence. The nomological network constitutes a theory from which propositions may be deduced and tested, providing additional evidence, which in turn feeds back into the nomological network. At any given point, what is known about a phenomenon based on the evidence is therefore expressed by the nomological network. With increasing evidence, increasing connections within the nomological network between evidence, unobservable theoretical propositions, and theoretical propositions are made. As the relationship between evidence and its theoretical explanations and causes becomes better understood, the very meaning of the theoretical terms that are the focus of scientific inquiry becomes increasingly clear. Logical empiricists therefore allow for evidence that points to unobservable but physically real and existential entities.

Postmodernism/Social constructionism. Postmodernism is a more recent movement in the philosophy of science (it was also a much broader social movement in architecture, art, literature, and cultural studies) that started in the 1980's (Lyotard, 1984). Social constructionism is a particular strand of postmodernism that emphasizes the socially constructed nature of social phenomena. Work in philosophy of science in

the 1960's and 1970's increasingly called into question the notion of the continuity of scientific progress (Kuhn, 1962), as well as the objectivity of empirical observation (Feyerabend, 1970). Social constructionist and postmodern conceptions of science emerged in response to these concerns. Although many versions of postmodernism and social constructionism have been advanced over the years,[38] both are centrally based around a rejection of the notion of objective knowledge. Postmodernists reject that view that science can arrive at objective knowledge of reality and instead endorse a more relativistic and socially constructed and contextualized view of truth and reality. Knowledge outside the constraints of social context and language is impossible, according to postmodernists, and therefore science is quintessentially socially constituted (e.g., Rorty, 1989). Social phenomena, according to social constructionists, are constructed by society based on the constraints and requirements of a given social context. They arise in response to social requirements, shaped by the particular contexts in which they emerge. According to social constructionists then, when it comes to studying social phenomena, there is no reality independent from the social context that is to be studied (in the same way that, when studying the rules of a game or sport for example, independently from that game there are no rules to be studied). Science allows for an articulation of a narrative that society agrees on, yet the narrative of science is in no way privileged over other possible narratives. In a different time, place, or society, with different demands and contexts, the language of science could be entirely different.

In terms of its relevance to knowledge then, postmodernists reject the central claim of logical empiricism that evidence affords increased knowledge of an objective reality in isolation from the context in which the evidence arises. The very process of generating evidence and studying a phenomenon in fact often changes and shapes that phenomenon, according to social constructionists. At the same time, many postmodernists and social constructionists also would not agree with the logical positivists that observables are any less epistemologically problematic than

[38] It is important to note that postmodernism and social constructionism are not unified movements in the same way that logical positivism and logical empiricism were and therefore there does not exist a unified strand of either of these two philosophies. A general summary of common claims of many strands of thought that might be classified as postmodern or social constructionist will be provided here.

unobservables. All observations according to postmodernists are shaped by the social contexts in which they occur and the frameworks that are available to those who are doing the observing. The only thing that evidence therefore permits access to is a particular socially constructed narrative. Evidence in science, when examined from a postmodern perspective, allows for an understanding of the social context in which it emerged, and the social ends toward which it is applied, but cannot be understood independently from the social context and does not provide an understanding of anything that can be separated from its social context.

Implications for EBP. The three philosophies of science outlined ever so briefly here have some very clear implications for EBP and what EBP can hope to accomplish. First, logical positivism, at least on first sight, appears quite consistent with EBP in the sense that empirical evidence is often assigned a privileged role in the EBP discourse. In fact, some of the EBP discourse appears to be shaped by certain positivist commitments. Quite commonly, especially when EBP is treated as synonymous with empirically supported treatment, it is conceptualized as a process where empirical (observable) evidence is used as a means to establish the utility of a treatment, which must also be operationally defined (e.g., administered in a standardized manner based on a manual) for a disorder that has to be diagnosed based on observable symptoms. Yet, of course, a logical positivist interpretation of EBP would limit the clinician to a rather narrow and specific interpretation of 'evidence.' Treatments, in a logical positivist framework, need to be understood as merely heuristic summaries of specific techniques used. Disorders also are understood as nothing more than a collection of observable behaviours. As soon as clinicians want to make any claim about the processes whereby a treatment becomes effective, or the causes of a disorder beyond that which is observable, their understanding enters the realm of theoretical language and thus becomes inconsistent with logical positivism.

A logical empiricist account of EBP may therefore be more tenable for most clinicians. In a logical empiricist framework, evidence would provide increasingly dense nomological networks related to a given disorder or treatment approach that constitute increased understanding of the treatment or disorder. As evidence accumulates, our understanding of the theoretical explanations of the evidence, including the active mechanisms of treatments and therapeutic techniques employed, and the manner in

which they interact with what is theorized to be the nature and cause of various disorders, will become increasingly clear. The push for the increased use and generation of evidence becomes most easily defensible in the context of a logical empiricist understanding of science. According to such an understanding, as evidence accumulates, our understanding of the nature of psychopathology and the manner in which psychological practices can reduce such psychopathology will become increasingly complete.

Postmodernists on the other hand reject the notion that evidence can provide cumulative and increasing knowledge of phenomena independently from the social contexts in which the evidence is generated. In a postmodernist understanding, both psychopathology and psychological practice are quintessentially socially embedded practices that can only be fully understood in light of the social contexts in which they occur. What is, and is not, considered pathological is based on the requirements and demands of particular social structures, according to postmodernists. Therefore, there is no reality to psychopathology or psychological practice beyond its social reality that must be recognized and understood.[39] In fact, postmodern critiques of evidence have been advanced in the EBP literature along these very lines (e.g., Edwards, 2007; Goldenberg, 2006).

It seems therefore that the mainstream EBP discourse relies mostly on logical empiricist commitments, which is not surprising given that such commitments have shaped most scientific discourse and practice in the latter part of the 20th century. Nonetheless, some of the criticisms of these perspectives advanced by postmodernists have been troubling to some clinicians, who do not want to subscribe to a postmodern rejection of the notion that phenomena can be understood independently from the social context in which they occur via evidence. A more moderate position on this topic will be suggested in the following section.

[39] I concede here that some social constructionist accounts (e.g., Hacking, 2000) allow for a more nuanced perspective whereby, for example, certain aspects of psychopathology may be socially constructed, but not all psychopathologies are. Within the scope of this work, only a very general treatment of social constructionist and postmodern perspectives was possible. In fact, there are major distinctions between postmodernism and social constructionism that will have implications for the manner in which 'evidence' is conceptualized that are also beyond the scope of this project.

When do we need Evidence?

A middle ground between logical empiricist perspectives of science, which appear to neglect the social nature of psychological practice, and postmodern perspectives, which appear to force the psychologist to abandon strict objectivist realist commitments, may be found in the philosophy that has emerged out of Wittgenstein's later writings. Specifically, Baker and Hacker (e.g., 1980) have advanced a philosophical distinction based on Wittgenstein's post-Tractatus work. They distinguish between conceptual and empirical issues in a manner that was touched on in the first Chapter of this work, but will now be expanded. The distinction between conceptual and empirical issues may allow for the separation of those issues that are in fact socially constructed, and those issues that may be conceived of as more objective, and therefore may benefit from examination through evidence.

To recap from the first Chapter of this thesis, Baker and Hacker (1980) reject traditional accounts of language that posit ostensive definitions of concepts (i.e., that the meaning of a concept is the entity denoted by the concept) and instead argue that the meaning of a concept is given by the rules of its correct employment. Baker and Hacker also contrast conceptual issues against empirical issues. Conceptual issues bear solely or mostly on logical relations between concepts and concepts and their referents. Empirical issues, on the other hand, concern the existential properties (e.g., origin, quantity, material composition) of a concept's referent or other empirical relations between referents of concepts. Conceptual issues, such as any issues related to the determination of the meaning of a concept, can only be adjudicated in reference to normative standards regarding the sanctioned application of a term or expression. Empirical issues, according to Baker and Hacker (1980), can be resolved through empirical means. However, empirical investigation presupposes and requires previous conceptual clarification of the concepts at play. Conceptual clarification precedes empirical investigation and is in-principle non-empirical.

Based on this very brief summary of the difference between conceptual and empirical issues, proposals with respect to the object of evidence in the EBP discourse described in the previous chapter can be examined. Some of the issues that proponents of EBP suggest may be adjudicated through empirical means do in fact appear well

suited to such adjudication based on Baker and Hacker's (1980) framework. For example, whether the use of a certain treatment results in a reduction in a particular symptom is very definitely an empirical issue. Issues related to the efficacy or effectiveness of a certain clinical practice most certainly are well-suited to empirical investigation. Accounts that conceptualize the efficacy or effectiveness of a practice as the object of evidence therefore are quite sensible within this framework. Questions with respect to the causes of particular disorders or the factors that increase the likelihood of the occurrence of certain symptoms again are empirical matters, and require empirical evidence for their adjudication. In each of these cases, however, in order to investigate these empirical matters, conceptual clarity is required with respect to the meanings of the concepts employed. For example, a definition of a particular clinical practice or a definition of a disorder must be understood in advance of investigating the effects of the practice on the disorder. It is impossible, for example, for a researcher to study the causes of an anxiety disorder, without first knowing the meaning or definition of an anxiety disorder. The linguistic meaning of a concept such as 'anxiety disorder' cannot be determined through empirical investigation. Rather it is laid down in the linguistic conventions that sanction particular ascriptions of the concept in particular contexts (e.g., diagnostic criteria 'anxiety disorder').

Other perspectives on EBP on the other hand appear to confuse empirical and conceptual issues in a manner that may lead to confusion about the possible applications of evidence to clinical practice. For example, any suggestion that evidence can clarify the meaning of a disorder (i.e., its diagnostic criteria) or the nature of a treatment (i.e., the techniques commonly employed in treatments designated by a certain concept) suggests that conceptual issues can be clarified using empirical means. What it means to have a diagnosis of an anxiety disorder, for example, is defined by the psychological and psychiatric community in a normative manner based on clinical practices (e.g., In phrases such as: …by an anxiety disorder we mean a person who meets the following criteria…). It is impossible to study anxiety disorders in order to understand how they ought to be defined, because their very investigation presupposes a definition of them. In the absence of a definition of the norms of the employment of a particular diagnosis or treatment, the diagnosis or treatment cannot be identified and therefore cannot be investigated. This, however, does not rule out the possibility that

empirical observations can sometimes lead to modifications in conceptual definitions (i.e., in this case diagnostic criteria).

Even more confusingly, some proponents of EBP argue that the meaning of EBP itself ought to be investigated empirically. In such proposals, it is sometimes suggested that through empirical investigation, scientists can determine the meaning of the concept of EBP. For example, Fonagy and colleagues (2005) state that "the criteria that are used to determine what counts as evidence-based practice must themselves be empirically tested" (p. 2) in a manner that confuses empirical and conceptual issues. Again, any investigation of EBP by empirical means implies that it first can be identified. In order for EBP to be identifiable, standards for the application of the concept 'EBP,' that is, what constitutes EBP, are necessary. Given the multiple conflicting conceptualizations of EBP implied in the EBP discourse, it appears that some additional conceptual clarity with respect to the concept of EBP may be a good starting point for any investigations that rely on this concept or any practitioners who aim to utilize EBP.

Conclusion

Philosophical discussions that examine whether evidence is propositional or physical, internal or external, and indicative or normative, shed some light on the implications that adopting a particular conception of evidence has for the relationship that evidence can have with its object. In addition, philosophical distinctions between bias and evidence may clarify the relationship between evidence, clinical expertise, and patient characteristics and preferences in EBP. Philosophical accounts of the relationship between evidence and its object propose different types of quantitative and qualitative relationships that also have implications with respect to the kinds of things that can serve as evidence in EBP. In particular, the requirement of total evidence poses important theoretical and practical challenges for EBP. Philosophical work on the balance, weight, and specificity of evidence can also illuminate the relationship between the amount of evidence in support of a proposition and the degree of support for it. Finally, philosophical accounts that focus on the object of evidence can provide perspectives on the kinds of generalizations that can be made based on evidence, the

kinds of issues that are suitable for adjudication by means of evidence, and matters that may require other forms of clarification.

The body of philosophical works on evidence is voluminous. The current chapter provides merely a cursory overview of some of its main ideas and debates. An exhaustive review of philosophical works on evidence and all of their implications for EBP is impossible within the scope of this project. I hope, however, that the brief presentation of some of this literature provided in the present chapter is sufficient to justify my disagreement with Achinstein's (2000) claim that the philosophical literature on evidence is of little value to the scientist. Instead, given the conceptual confusions inherent in the EBP discourse, philosophy may well provide one means for psychologists to resolve conceptual issues related to EBP. If such conceptual clarification occurs, eventually, the empirical investigations of EBP that proponents of this approach call for may at least become possible. The conceptual confusion inherent in the EBP discourse is well explained by Lamiell's (2012) comments on the impact of the separation of psychology and philosophy on psychological knowledge:

> Therein lies the problem for us: long ago, and very much against Wundt's advice, our discipline, in its collective wisdom, divorced itself from philosophy. As a result, it is today no longer populated with, or especially nurturant of, young scholars who are attuned to philosophical questions and appreciative of their importance for the long-range intellectual health of the field. Our discipline has thus made itself largely blind to such questions, and, as Wundt foresaw would be the case, conceptual confusion has been the result. Worse yet, that conceptual confusion is often obscured behind the façade of scientific progress.

Lamiell's comments, although directed at psychological practices in general, appear particularly well-suited to the current status of EBP in psychology.

Chapter 8:
Conclusions

A reading of the chapters of this work will hopefully support the following claim: 'evidence' is not a simple or self-explanatory concept. Throughout human history, evidence has consisted in different things, and has served a variety of different roles in a number of social contexts and practices. Very recently, 'evidence' has come to play a central role in a variety of scientific and professional contexts in which evidence-based practice has been lauded as an approach that promises to standardize and improve the delivery of many services, particularly health care services. In this way, 'evidence' has entered the discourse of psychological practitioners, who have been increasingly asked to employ evidence-based practices. The purpose of the present project was to examine the concept of 'evidence' more closely, at first in its broader employment in a wide variety of social contexts across history, then in psychology more specifically, and finally in the EBP movement in clinical psychology.

This thesis provided a general overview of the history of the concept of 'evidence' in society, psychology, and psychotherapy and examined conceptions of evidence explicit and implicit in the EBP movement in contemporary psychotherapy research. A main goal was to examine the various conceptualizations of 'evidence' that practitioners and researchers in psychology have endorsed in the relevant psychological literature, and to uncover and analyze conceptual and epistemological assumptions inherent in these endorsements. A wide variety of historical, archival, and qualitative empirical methods were employed towards this end. Detailed attention was given to the interview-generated statements of members of the APA Task Force on EBP, relevant archival records of Task Force deliberations, and articles published that relate to the EBP movement. A review of philosophical treatments of evidence was also added to aid a clarification of various conceptual issues that emerged.

Summary of Conclusions

My historical review uncovered a surprising variety of definitions and conceptualizations of 'evidence' that were at times complimentary but frequently widely disparate and mutually incompatible. A review of dictionary definitions of 'evidence' yielded definitions that often emphasize the relationship between evidence and sensory data but also define 'evidence' as accessible through the senses, an indicator of something that may not be accessible through the senses, or as ground for belief. Occasionally, 'evidence' is also defined as consisting of testimony or documents presented in the manner of legal proceedings. Clearly, varying definitions of 'evidence' have co-existed historically and continue to co-exist today.

'Evidence' has played an important role in legal contexts throughout history. Testimony provided the most accepted form of evidence in early legal proceedings, which were mainly intended to strengthen or re-establish existing social orders. As the purpose of legal proceedings shifted towards obtaining divine justice, ordeals, wagers of law, and battles came to play an increasingly central role in the Middle Ages. Testimony and objects became important forms of evidence during the Enlightenment, when the purpose of legal proceedings became to ascertain truth and fact. The nature and role of evidence in legal contexts has shifted along with social changes in understandings of the purpose of legal proceedings and the contexts in which they occurred.

In science, the role of 'evidence' and its conceptual predecessors has also shifted significantly from Antiquity to the Modern Age. In Antiquity, signs played the role of what would now be considered evidence. Signs, however, did not figure significantly in the sciences of Antiquity, which were more focused on logically certain knowledge and eshewed rhetorical and conjectural practices that were considered to be outside the realm of science. Significant theoretical diversity exists with respect to the role of signs in such practices and debates in Antiquity anticipated theoretical issues that continue into modern philosophy. In the Middle Ages, science was shaped by a Christian worldview and evidence came to be understood as a means of accessing divine knowledge. Humans, it was thought, can receive access to some of God's complete knowledge through a process of divine illumination via the senses. Experiences and

sensory data therefore began to play a more important role in the scientific practices of the Middle Ages. In the Renaissance, several epistemological shifts occurred that paved the way for a more contemporary understanding of evidence in science. Empirical data emerged as a legitimate source of evidence, natural laws and regularities became the goals of scientific inquiry, and the formulation of hypotheses became a central part of the scientific process. Mathematics also began to be viewed as a way to quantify uncertainty in the pursuit of scientific knowledge. Finally, during the Enlightenment, a mechanical worldview emerged and evidence of the mechanical principles that govern the world was sought through empirical measurement and publically observable experiments. The concept of 'evidence' emerged as a means to inductively support scientific practices.

A historical review of social practices related to evidence (including language, law, and science) therefore highlights both the diverse conceptualizations of 'evidence' that have existed throughout history and the widely disparate roles that evidence has played in different social contexts. 'Evidence' is far from a unified concept. Rather, what 'evidence' is taken to be very much depends on the importance attributed to it, as well as the purposes and contexts of its employment.

When turning one's attention to psychology, despite the relatively short history of the discipline, 'evidence' and its role in psychological practices also have shifted across time. Psychology came to be viewed as a scientific discipline relatively early in its history, and experiments were conceptualized as a central source of psychological evidence. Experiments, however, did not play an important role in the development of psychological treatments, which were mostly based on case studies and theories about the causes of psychopathological symptoms. An understanding of the role of the clinician as a scientist emerged only in the middle of the 20th century and only then did experimental research become an important source of evidence for clinical practice.

Early psychotherapy research focused on the efficacy of psychotherapy in general. Disappointing early research findings prompted clinicians to develop specific methods to demonstrate the efficacy of psychotherapy and eventually randomized-controlled trials and meta-analysis were developed and adopted. These methods increasingly were employed to compare the effectiveness of various approaches to

treatment. Gradually, the amenability of a treatment to experimental investigation became a necessary criterion for its consideration as a bona fide scientifically based psychological treatment. A review of the history of evidence in psychotherapy indicates that experiments have played an important role in clinical research only in the past three to four decades. Previously, a much wider variety of sources of evidence were employed in the development of psychological treatments.

The ascendance of evidence, and experimental evidence in particular, to a central position in the discourse of clinical psychology has recently been marked by the EBP movement. In examining its history, it becomes clear that EBP emerged in response to various social and professional pressures in the late 20th century. Amidst concerns about rising costs of healthcare, and the application of principles of managerial science to medical practice, evidence based medicine emerged as a means to efficiently and effectively manage the delivery of healthcare. Almost inevitably, psychology and psychological services began to be drawn into the evidence-based practice movement.

Originally, psychologists focused their attention on the developments of standards and guidelines for treatment to ensure that they would not be left behind in the rush to demonstrate accountability and effectiveness. The clinical psychology division of the American Psychological Association eventually formed a task force that generated specific recommendations and standards to be used in the evaluation of psychological treatments. Based on these recommendations and standards, a list was generated of what were considered to be "empirically-validated treatments." Healthcare Management Organizations in the United States began to use these lists as a basis for reimbursement decisions for psychological services. Amidst growing concerns about such developments, the APA responded by forming a Task Force on EBP in psychology. The Task Force generated a document, recommendations, and an APA policy statement on EBP, which received mixed reviews from the psychological community. Many of the responses to the Task Force document focused on the practical feasibility of EBP. However, comparatively few discussions of the theoretical assumptions inherent in EBP in psychology were advanced. It is clear from this review of the history of EBP in psychology that this movement arose in response to particular social and economic pressures and aimed to establish the legitimacy of psychological treatments as a means to treating mental health problems.

Despite the unanimous approval of an APA policy on EBP, a review of the EBP literature reveals that conceptualizations of 'evidence,' its object, and its relationship to its object vary widely both within and between published articles. The manner in which evidence is characterized throughout the EBP discourse raises numerous questions in terms of what 'evidence' is, how it can be found, how it relates to that which it is intended to support, and what exactly evidence is to be employed in support of. When the EBP literature in psychology is considered carefully, it contains such diversity in conceptualizations and assumptions about evidence that it becomes doubtful whether 'evidence' can possibly serve all of the purposes attributed to it in this literature. Most notably, despite the various functions assigned to it, it is still unclear what evidence is assumed to be. Conceptual confusion about 'evidence' and evidence-based practice raises questions about the extent to which such practice can yield the results it is being promoted to achieve. At the same time, 'evidence' as it is employed in the EBP discourse is often defined in a more narrowly technical manner than its definition in general parlance.

An examination of philosophical treatments of 'evidence' was then employed with the hope of clarifying some of the conceptual confusions inherent in the EBP literature in psychology. Philosophical treatments of evidence, its object, and the relationship between evidence and its object provided some clarification about the impact that various conceptualizations of 'evidence' might have on how evidence might function in EBP. The philosophical literature also illuminated the implications of adopting particular conceptions and assumptions of the relationship between evidence and its object. Philosophical treatments of 'evidence,' particularly the extent to which generalization beyond evidence is feasible and the kinds of issues that are amenable to adjudication via evidence, also provided some additional clarification of the way in which EBP might be developed in a more conceptually and theoretically sound and consistent manner.

It seems fair to conclude that throughout human history and across various social contexts, there have been marked shifts in conceptualizations of 'evidence' and in the roles evidence has played in the generation of knowledge. These shifts carried with them implications for and led to changes in the practice of science, both in the scientific methods used to generate evidence and in conventions regarding what constitutes proper interpretation of scientific data or evidence. Many of the various conceptions of

'evidence' that have emerged throughout the history of science co-exist in the context of evidence-based practice in clinical psychology. Different understandings of evidence and its role in scientific practice are apparent throughout the EBP discourse, often within the same document or account. Accounts of evidence-based practice shift flexibly between these various conceptualizations and often equivocate between them. Frequently, use of the term 'evidence' appears to lend an illusion of conceptual consensus in the absence of definitional, conceptual, or theoretical clarity.

A philosophical examination of various theories of evidence shows that each theory has distinctive implications for scientific practice and that the assumptions inherent in conceptualizing evidence in each theoretical framework are at times conceptually, logically, and practically incompatible. As a result, the definitional plurality of evidence in the EBP literature and discourse, albeit possibly useful for the purpose of social and professional consensus, is problematic for coherent scientific and clinical practice in psychology.

Implications

I hope to have highlighted the need to understand and clarify the conception of evidence central to the EBP movement in clinical psychology. A full clarification of all the conceptual issues raised herein is, of course, well beyond the scope of the present work. My primary aim here was to establish the necessity of such clarification. Future clarification of some of the conceptual confusions and challenges in evidence-based practice will allow for a better understanding of the scope and limitations of evidence-based practice. If clinicians are to employ EBP and if the employment of such practice is in fact an ethical mandate for clinicians, as some have suggested, it is essential that conceptual clarity about the nature of EBP is sufficient to allow clinicians to identify EBP and discuss it clearly and sensibly.

Given the current state of the EBP literature, it seems fair to conclude that caution may be warranted in the use, acceptance, and promotion of EBP by clinicians. The community of clinical practitioners has for the most part accepted the EBP paradigm relatively uncritically. Many clinicians now are quick to insist that they employ evidence-

based practices, in a manner that appears to have become akin to ascertaining that one is practicing responsibly and in an informed manner. As the chapters of this work illustrate however, in the absence of clarity with respect to what constitutes evidence-based practice, it makes little sense to insist that one is using it. It therefore seems that clinical psychologists may benefit from some consideration of the boundaries of the term 'evidence-based practice.' Greater clarity about the distinctions between evidence-based practice and empirically supported treatments is certainly needed. Given the numerous and incompatible conceptualizations of EBP, however, it also may not be appropriate to treat this term as denoting a unified and clearly defined movement, but rather a term that has been flexibly employed to describe a large number of practices that employ and conceptualize evidence in many different ways. It may be useful to generate a classification scheme for various types of evidence-based practice, such that when clinicians employ the term, it would be more obvious what specific definition of the term they are referring to.

Based on the discussions in the previous chapter, it also seems that the psychological community might do well to abandon the use of the term 'evidence-based treatment,' which may be perpetuating the confusion between the evidence-based practice movement and empirically supported treatments. Most Task Force members denied the utility of this term and acknowledged that its definition is unclear. Unless the term 'treatment' is used to describe the treatment of a particular individual by a specific clinician, there is no meaningful way to define evidence-based treatment that is consistent with the three-legged approach endorsed by the APA Task Force. Since clinicians often interpret 'treatment' to mean a particular treatment approach, this term leads to confusion and is interpreted in a manner that suggests that particular treatment approaches can be either evidence-based or not in a dichotomous manner. In the absence of clear criteria to justify the application of this term to a particular treatment approach then, it may be wise to abandon this term and speak either of evidence-based practice or empirically supported treatments, depending on the approach one intends.

More importantly, if clinicians are to employ EBP responsibly, it appears that they must consider their views on various theoretical issues related to EBP, many of which have been highlighted in the previous chapters of this work. First, it will be important for clinicians to examine their views on the nature of evidence and its object. In particular, it

is important for clinicians to recognize the philosophical commitments inherent in adapting a particular conception of evidence and its implications for the roles that evidence can serve. Clinicians would also do well to become clearer in their understandings of the relationship between evidence and its objects. It will be important for practitioners to reach some consensus on the extent to which more evidence constitutes stronger evidence for a practice, total evidence must be considered, and evidence can be weighted or ranked based on certain standards. All of these matters will affect the manner in which clinicians will utilize evidence in their practice. Finally, it will be important to also become clear on what parts of clinical practice are to be adjudicated by means of evidence (e.g., treatment approaches, the therapeutic relationship, patient preferences). It may well be that different standards of evidence will be required for different objects of evidence. For example, the evidence that may serve as a basis for a consideration of patient preferences (i.e., patients' statements of their preferences) will differ from the evidence that provides an ideal basis for the consideration of the efficacy of a treatment approach in a given population (e.g., a randomized-controlled trial).

Evidence-based practice has also become a political and economic issue in psychology. Increasingly, treatment facilities are requiring EBP, familiarity with EBP is becoming part of the hiring criteria for psychologists in various settings, and knowledge of EBP is becoming part of the licensure process for psychologists in several jurisdictions. Consumers of psychological services may now be advised to ensure they select treatment providers based on their familiarity with EBP and community mental health settings are attempting to move towards greater employment of EBP. In all of this, it would be good to know that advocates and those concerned about EBP know what they are talking about. Clarification of the concept of evidence in EBP can contribute to the establishment and refinement of useful and sound policies and practices that ensure appropriate and beneficial delivery of psychological services.

Limitations and Future Directions

A complete history of the concept of evidence in all its various historical and sociocultural contexts would be voluminous and would far exceed what is possible within

a single thesis. Even a focus on evidence in psychology as an entire discipline would be extensive. The present project focused primarily on conceptions of evidence in EBP in clinical psychology. EBP also plays a central role in many other sub-disciplines of psychology (e.g., educational psychology, forensic psychology, and industrial/organizational Psychology) and other, related disciplines (medical practice, nursing, occupational therapy, social work etc.). The role of evidence and conceptualizations of evidence in these sub-disciplinary and disciplinary contexts may well differ from those extant in psychotherapy practice and research. Consequently, the findings in this project may not be directly transferrable to these other contexts.

My intention was not to provide an exhaustive account of all the conceptual issues that can be identified in the EPB discourse in clinical psychology, but to highlight some of the major and most obvious issues in an effort to support the claim that further conceptual clarification is needed. The use of a different sample of articles and documents or the integration of additional works (e.g., books and monographs) perhaps might have changed the findings of this project. Nonetheless, since academic journal articles play a very central role in psychological discourse, and a rather large sample of such articles was reviewed for this study, it is likely that at least some of the issues identified would also be found in other parts of the EBP discourse in clinical psychology. The convergence between findings from journal articles and issues that emerged in discussions with Task Force members further supports this conclusion.

A major limitation of this work relates to the practical requirements imposed by data analytic choices, specifically the decision to employ qualitative data analysis software and search functions within the software. Doing so permitted the analysis of a very large sample of articles that would not have been feasible without the use of such electronic aids. At the same time however, a large number of the articles in the sample for this work were not available in a format that permitted their analysis using the software employed. It is possible that the sample of articles used for this study was biased towards articles from more mainstream journals, which may be more likely to make their articles accessible in electronic format. In fact, a review of the articles that were not electronically available and those that were supports this possibility. Ideally, a future analysis would include a manual review of those articles that could not be

analyzed electronically. For the present project, however, this was not possible, given time constraints and the already large scope of this project.

Conclusions about conceptions of evidence were based on published articulations of such conceptualizations within the psychotherapy research and practice literature, and on interviews with some of those who participated on the APA Task Force. Only some of the members of the APA Task Force were available for interviews. Conceptualizations therefore may or may not fully represent the views of the Task Force or the psychology community as a whole. Nonetheless, such views can be taken to represent many of the important perspectives and conceptions at play in the EBP movement in psychotherapy research and practice.

Finally, a very brief and cursory discussion of philosophical perspectives on evidence was provided in this paper based on the interest, knowledge, and views of the author, and the issues that emerged from the conceptual review of the EBP discourse. This review could not possibly do justice to the depth and complexity of these philosophical accounts, which themselves warrant the attention of many focused projects. The review understandably fell short of the full breadth of relevant philosophical issues that might pertain to EBP. Its goal, rather than finality, was to establish that philosophy does indeed have something to contribute to the psychological discourse on EBP and indeed may be the only feasible tool for the conceptual clarification of the nature of evidence in EBP.

Many more projects could highlight countless additional and alternate philosophical and theoretical perspectives that would stimulate discussion and debate on the EBP movement in clinical psychology. Additional empirical research using interviews and surveys of practicing clinical psychologists and researchers could provide a broader overview of the various conceptualizations of evidence endorsed, and actually applied within practice. Finally, similar historical, empirical, and theoretical examinations of the nature of evidence in EBP in other sub-disciplines of psychology and other disciplines could also be fruitful, in order to clarify commonalities and differences across these various contexts.

Concluding Thoughts

If there is one major goal that was hopefully attained through the present project, it was to increase the reader's awareness of the complexity of the issues that surround the notion of evidence in EBP. What may at first glance appear to be a perfectly straightforward proposal, upon further examination contains a number of unexamined philosophical issues that do not lend themselves to empirical investigation. EBP may very well have the potential to advance the discipline beyond its current point, and to lead to more rigorous and socially beneficial practice. At the same time, given psychology's reliance on complex, social concepts, and the numerous conceptual issues that precede any empirical examination, it is unlikely that EBP will provide the answer to all of the questions raised herein. In the midst of calls for increased EBP and statements that the use of EBP is an ethical mandate for clinicians, I propose that the ultimate measure of the rigorous or responsible clinician may not be the extent to which he or she employs EBP or derives evidence that can be translated into practice, but rather the extent to which he or she is aware of the types of issues that do and do not lend themselves to empirical adjudication (empirical issues versus issues that are in principle non-empirical), the context within which EBP arose, and thus is able to responsibly and appropriately apply or perhaps choose not to apply EBP, in careful consideration of the particular circumstances within which clinical psychological interventions are embedded

References

Achinstein, P. (1978). Concepts of evidence. *Mind, 87*(345), 22-45.

Achinstein, P. (2000). Why philosophical theories of evidence are (and ought to be) ignored by scientists. *Philosophy of Science, 67*(3), S180-S192.

Adams, J. R., & Drake, R. E. (2006). Shared decision-making and evidence-based practice. *Community Mental Health Journal, 42*(1), 87-105.

Addis, M. (2002). Methods for disseminating research products and increasing evidence-based practice: Promises, obstacles, and future directions. *Clinical Psychology: Science and Practice*, *9*(4), 367-378.

Addis, M., & Waltz, J. (2002). Implicit and untested assumptions about the role of psychotherapy treatment manuals in evidence-based mental health practice: Commentary. *Clinical Psychology: Science and Practice*, *9*(4), 421-424.

Addis, M., Wade, W., & Hatgis, C. (1999). Barriers to dissemination of evidence-based practices: Addressing practitioners' concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice*, *6*(4), 430-441.

Allen, J. (2001). *Inference from signs: Ancient debates about the nature of evidence*. Oxford: Oxford University Press.

American Psychological Association. (1947). Committee on training in clinical psychology. recommended graduate training program in clinical psychology. *American Psychologist, 2*(12), 539-558.

American Psychological Association Division 12. (n.d.). *Dissemination subcommittee of the committee on science and practice*. Retrieved from http://www.apa.org/divisions/div12/rev_est/committee.html

American Psychological Association, Policy statement on evidence-based practice in psychology. (2006). *American Psychologist, 61,* 271-285.

American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *The American Psychologist, 57*(12), 1052-1059.

American Psychological Association. (2004). Presidential task force on evidence based practice task force charge. Unpublished document.

American Psychological Association. (2005). Council item: Policy recommendation and position paper of the 2005 presidential task force on evidence-based practice. Unpublished document.

Anderson, E. E., & DuBois, J. M. (2007). The need for evidence-based research ethics: A review of the substance abuse literature. *Drug and Alcohol Dependence*, *86*(2-3), 95- 105.

Andersson, G., Cuijpers, P., Avdelningen för klinisk och socialpsykologi (CS), Institutionen för beteendevetenskap och lärande, Filosofiska fakulteten, & Linköpings universitet. (2009). 'Psychological treatment' as an umbrella term for evidence-based psychotherapies? *Nordic Psychology, 61*(2), 4-15.

Andrews, G., & Titov, N. (2009). Hit and miss: Innovation and the dissemination of evidence based psychological treatments. *Behaviour Research and Therapy*, *47*(11), 974-979.

Aquinas, S.T. (1473/1920). *Summa theologica: Part One* (trans. Fathers of the English Dominican Province). Retrieved from http://www.newadvent.org/summa/1.htm.

Araujo, S. (2012). Why did Wundt abandon his early theory of the unconscious? Towards a new interpretation of Wundt's psychological project. *History Of Psychology*, *15*(1), 33-49.

Aristotle. (350BCE). *Physics (*trans. Hardie, R.P. & Gaye, R.K.), 184a, 10-15. Retrieved from http://classics.mit.edu/Aristotle/physics.html.

Ayer, A. J. (Alfred Jules). (1956). *The problem of knowledge*. New York: Macmillan.

Ayer, A. J. (Alfred Jules). (1982). *Philosophy in the twentieth century*. London: Weidenfeld.

Ayers, C. R., Sorrell, J. T., Thorp, S. R., & Wetherell, J. L. (2007). Evidence-based psychological treatments for late-life anxiety. *Psychology and Aging, 22*(1), 8-17.

Bacon, F. (1620/2005). *The new organon or: true directions concerning the interpretation of nature* (trans. J. Bennett) Retrieved from http://www.earlymoderntexts.com/pdfbits/bacno1.pdf

Bacon, R., Bridges, J. H., British Library, & Adam Matthew Digital (Firm). (1900). *The 'opus majus' of roger bacon*. London: Williams and Norgate.

Baker, D. B., & Benjamin, J., L T. (2000). The affirmation of the scientist-practitioner. A look back at boulder. *The American Psychologist, 55*(2), 241-247.

Baker, G.P., & Hacker, P.M.S. (1980). *Wittgenstein meaning and understanding*. Chicago: The University of Chicago Press.

Barlow, D. H. (2004). Psychological treatments. *American Psychologist*, *59*, 869-879.

Barlow, D. H. (2005). What's new about evidence-based assessment?. *Psychological Assessment*, *17*(3), 308-311.

Barrett, P. M., Farrell, L., Pina, A. A., Peris, T. S., & Piacentini, J. (2008). Evidence-based psychosocial treatments for child and adolescent obsessive-compulsive disorder. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 37*(1), 131-155.

Bauer, R. M. (2007). Evidence-based practice in psychology: Implications for research and research training. *Journal of Clinical Psychology, 63*(7), 685-694.

Beilby, J. (2002). The relationship between faith and evidence in St. Augustine. *Sophia, 41*(1), 19-32.

Benjamin, J., L.T. (2007). *A brief history of modern psychology*. Oxford: Blackwell Publishing.

Bentham, J. (1802/2000). *The correspondence of Jeremy Bentham*. Charlottesville, Va: InteLex Corporation.

Bernal, G., Jiménez-Chafey, M. I., & Domenech Rodríguez, M. M. (2009). Cultural adaptation of treatments: A resource for considering culture in evidence-based practice. *Professional Psychology: Research and Practice, 40*(4), 361-368.

Blunt, H. W. (1904). Bacon's method of science. *Aristotelian Society, Proceedings, 4*, 16-31.

Bohart, A. (2005). Evidence-Based psychotherapy means evidence-informed, not evidence-driven. *Journal of Contemporary Psychotherapy*, *35*(1), 39-53.

Bower, P. (2003). Efficacy in evidence-based practice. *Clinical Psychology & Psychotherapy*, *10*(6), 328-336.

Bowler, P. J., & Morus, I. R. (2005). *Making modern science: A historical survey*. Chicago: University of Chicago Press.

Broekaert, E., Autrique, M., Vanderplasschen, W., & Colpaert, K. (2010). 'The human prerogative': A critical analysis of evidence-based and other paradigms of care in substance abuse treatment. *The Psychiatric Quarterly, 81*(3), 227-238.

Brooke, R. (Summer 2006). Controversial discussions: II. The return of expertise in evidence based practice. *Psychologist-Psychoanalyst*, 23–26.

Brown, L. S. (2006). Still subversive after all these years: The relevance of feminist therapy in the age of evidence-based practice. *Psychology of Women Quarterly, 30*(1), 15-24.

Brysbaert, M., & Rastle, K. (2009). *Historical and conceptual issues in psychology.* Essex: Pearson Education Ltd.

Butterfield, H., Sir. (1957). *The origins of modern science: 1300-1800*. New York: Macmillan.

Byrne, E. F. (1968). Probability and opinion: A study in the medieval pre-suppositions of post-medieval theories of probability. The Hague: Martinus Nijhoff.

Campbell Black, H. (1892). Antiquities of the law of Eevidence: The competency of witnesses. *The American Law Review (1866-1906), 26*(6), 821.

Campbell, J. M., Brown, R. T., Cavanagh, S. E., Vess, S. F., & Segall, M. J. (2008). Evidence-based assessment of cognitive functioning in pediatric psychology. *Journal of Pediatric Psychology, 33*(9), 999-1014.

Cape, J., & Barkham, M. (2002). Practice improvement methods: Conceptual base, evidence-based research, and practice-based recommendations. *The British Journal of Clinical Psychology / the British Psychological Society, 41*(Pt 3), 285-307.

Carnap, R. (1950). *The logical doundations of probability*. Chicago: University of Chicago Press.

Carpenter, W. S. (1958). *Foundations of modern jurisprudence*. New York: Appleton-Century Crofts, Inc.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685- 716.

Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49,* 5-18.

Chwalisz, K. (2003). Evidence-based practice: A framework for twenty-first-century scientist-practitioner training. *The Counseling Psychologist*, *31*(5), 497-528.

Cicero, M. T. (1851). The orations of marcus tullius cicero, literally tr. by C. D. Yonge (p. 241-380). London: George Bell & Sons.

Collins, A. F. (1999). The enduring appeal of physiognomy: Physical appearance as a sign of temperament, character, and intelligence. *History Of Psychology*, *2*(4), 251-276.

Conee, E., & Feldman, R. (2004). *Evidentialism*. Oxford: Oxford University Press.

Cooper, S. E., Benton, S. A., Benton, S. L., & Phillips, J. C. (2008). Evidence-based practice in psychology among college counseling center clinicians. *Journal of College Student Psychotherapy, 22*(4), 28-50.

Copernicus, N. (1959). Three copernican treatises: The commentariolus of copernicus, the letter against werner, the narratio prima of rheticus (trans. Rosen, E.). New York: Dover Publications.

Crombie, A. C. (1958). Robert Grosseteste and the origins of experimental science, 1100-1700. Oxford: Clarendon Press.

Cruz, J., & Pollock, J. (2004). The chimerical appeal of epistemic externalism. In Schantz, R. (Eds.). *The externalist challenge*. New York: Walter de Gruyter.

Danziger, K. (1979). The positivist repudiation of Wundt. *Journal of the History of the Behavioral Sciences, 15*, 205–230.

Danziger, K. (1985). The origins of the psychological experiment as a social institution. *American Psychologist,40*(2), 133-140.

Danziger, K. (1990). Constructing the subject: Historical origins of psychological research. New York, NY US: Cambridge University Press.

Danziger, K. (1993). Psychological objects, practice, and history. In H. V. Rappard, et al. (Eds.), *Annals of theoretical psychology, Vol. 8.* (pp. 15-47). New York, NY US: Plenum Press.

Danziger, K. (1997). Naming the mind: How psychology found its language. Thousand Oaks, CA: Sage.

Danziger, K., (2008). *Marking the mind: A history of memory*. Cambridge, UK: Cambridge University Press.

Daston, L. (1991). Marvelous facts and miraculous evidence in early modern europe. *Critical Inquiry, 18*(1), 93-124.

Daston, L., & Galison, P. (2007). *Objectivity*. New York: Zone Books.

Daston, L., Stolleis, M. (2008). Natural law and laws of nature in early modern europe: Jurisprudence, theology, moral and natural philosophy. Farnham, England: Ashgate Pub. Company.

Pollio, D.E., & Macgowan, M.J. (2010). The andragogy of evidence-based group work: An integrated educational model.*Social Work with Groups, 33*(2/3), 195-209.

Davison, G. C. (2006). An invitation to auseindersetzen about the Evidence-Based Practices in psychology task force report. *The Clinical Psychologist/American Psychological Association*, *59*, 1 - 4.

De Los Reyes, A., & Kazdin, A. E. (2008). When the evidence says, "yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions.*Current Directions in Psychological Science, 17*(1), 47-51.

Dear, P. (2006). The meanings of experience. In. K. Park & L. Daston (Eds.), *The Cambridge History of Science – Volume 3: Early Modern Science* (pp. 106-131). Cambridge: Cambridge University Press.

Dehue, T. (1997). Deception, efficiency, and random groups: Psychology and the gradual origination of the random group design. *Isis, 88*(4), 653-673.

Descartes, R. (1641/1998). Meditations on first philosophy. In R. Ariew & E. Watkins (Eds.), *Modern Philosophy: An Anthology of Primary Sources* (pp. 22-55). Indianapolis, IN: Hackett Publishing Company.

DiFate, V. (2007). *Evidence – Internet encyclopedia of psychology*. Retrieved from http://www.iep.utm.edu/evidence.

Dretske, F. (1971). Conclusive Reasons. *Australasian Journal of Philosophy*, *49*, 1–22.

Eddy, D. M. (2005). Evidence-based medicine: A unified approach. *Health Affairs (Project Hope), 24*(1), 9-17.

Eddy, D.M. (1990). Practice Policies: Where Do They Come From? *Journal of the American Medical Association, 263*, 1265- 1275.

Ede, A. & Cormack, L. B. (2004). *A history of science in society: From philosophy to utility*. Peterborough, Ont: Broadview Press.

Edwards, D. A., Dattilio, F. M., & Bromley, D. B. (2004). Developing evidence-based practice: The role of case-based research. *Professional Psychology: Research and Practice*, *35*(6), 589-597.

Edwards, D., Dattilio, F., & Bromley, D. (2004). Developing Evidence-Based Practice: The Role of Case-Based Research. *Professional Psychology: Research and Practice*, *35*(6), 589-597.

El Baz, N., Middel, B., Van Dijk, J. P., Oosterhof, A., Boonstra, P. W., & Reijneveld, S. A. (2007). Are the outcomes of clinical pathways evidence-based? A critical appraisal of clinical pathway evaluation research. *Journal of Evaluation in Clinical Practice, 13*(6), 920-929.

Elliot, S. A., Ashton, C. C., Gerrard, J. J., & Cox, J. L. (2003). Is trainer training an effective method for disseminating evidence-based practice for postnatal depression?. *Journal of Reproductive and Infant Psychology*, *21*(3), 219-228.

Evidence-based Medicine Working Group. (1992). Evidence-based Medicine: A New Approach to Teaching the Practice of Medicine. *Journal of the American Medical Association, 268*, 2420–2425.

Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. Journal of Consulting Psychology, 16(5), 319-324.

Eysenck, M. W. (1978). Levels of processing: A critique. *British Journal of Psychology, 69*(2), 157-169.

Fago, D. P. (2009). Comment: The evidence-based treatment debate: Toward a dialectical rapprochement. *Psychotherapy (Chicago, Ill.), 46*(1), 15-18.

Falkum, E. (2008). Phronesis and techne: The debate on evidence-based medicine in psychiatry and psychotherapy.*Philosophy, Psychiatry, & Psychology, 15*(2), 141-149.

Falzon, L., Davidson, K. W., & Bruns, D. (2010). Evidence searching for evidence-based psychology practice. *Professional Psychology: Research and Practice, 41*(6), 550-557.

Feigl, H. (1950). Existential hypotheses: Realistic versus phenomenalistic interpretations.

Feldman, R., & Conee, E. (2001). Internalism defended. *American Philosophical Quarterly, 38*(1), 1-18.

Feltham, C. (2005). Evidence-based psychotherapy and counselling in the UK: Critique and alternatives. *Journal of Contemporary Psychotherapy, 35*(1), 131-143.

Ferguson, M. C. (1960). A day in court in justinian's rome: Problems of evidence, proof, and justice in roman law. *Iowa Law Review, 46*, 732-772.

Feyerabend, P. (1976). *Against method: Outline of an anarchistic theory of knowledge*. Atlantic Highlands: NLB.

Findlen, P., & Bianchi, M. L. (1990). Empty signs: Reading the book of nature in renaissance science. *Studies in History and Philosophy of Science, 21*(3), 511-518.

Fischer, N. W. & Unguru, S (1971). Experimental science and mathematics in Roger Bacon's thought. *Traditio, 27*, 353-378.

Fonagy, P. (1999). Achieving evidence-based psychotherapy practice: A psychodynamic perspective on the general acceptance of treatment manuals. *Clinical Psychology: Science and Practice*, *6*(4), 442-444.

Fonagy, P., Roth, A., & Higgitt, A. (2005). Psychodynamic psychotherapies: Evidence-based practice and clinical wisdom. *Bulletin of the Menninger Clinic, 69*(1), 1-58.

Franklin, J. (2001). *The science of conjecture: Evidence and probability before pascal*. Baltimore, MD: Johns Hopkins University Press.

Frederick, C. (2007). Hypnotically facilitated treatment of obsessive-compulsive disorder: Can it be evidence-based?.*The International Journal of Clinical and Experimental Hypnosis, 55*(2), 189-206.

Friedman, M. (1999). *Reconsidering logical positivism*. New York: Cambridge University Press.

Fumerton, R. (2010). Foundationalist theories of epistemic justification. *The Stanford Encyclopedia of Philosophy.* Retrieved from http://plato.stanford.edu/archives/sum2010/entries/justep-foundational

Furukawa, T. A., Shear, M., Barlow, D. H., Gorman, J. M., Woods, S. W., Money, R., & Leucht, S. (2009). Evidence-based guidelines for interpretation of the panic disorder severity scale. *Depression and Anxiety, 26*(10), 922-929.

Gagarin, M. (1986). *Early Greek Law*. Berkley, CA: University of California Press Ltd.

Gagarin, M., & Woodruff, P. (1995). *Early greek political thought from homer to the sophists*. Cambridge: Cambridge University Press.

Gallo, P. S. (1978). Meta-analysis: A mixed metaphor?. *American Psychologist, 33*(5), 515-517.

Galton, F.R.S. (1884). Anthropometric laboratory for the determination of weight, span, breathing power, strength of pull and squeeze, quickness of blow, hearing, seeing, color-sense, and other personal data. London: William Clowes and Sons, Ltd.

Gilbert, G., Sir. (1756). The law of evidence: By a late learned judge. in this edition, the errors in the irish edition have been corrected, the original references carefully compared, and many new references added, particularly to the late sir john strange's reports. with a complete table to the whole. London: printed by Henry Lintot; for W. Owen.

Ginsburg, G. S. (2006). Evidence-based treatments for children and adolescents. Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 35(3), 480-486.

Glaser, B. G. and Strauss, A. L. (1967) The discovery of grounded theory: strategies for qualitative research. Chicago.: Aldine.

Glymour, C. (1975).Relevant evidence. *Journal of Philosophy, 72*, 403-420.

Goddu, A. (1984). *The physics of william ockham.* Leiden: E.J. Brill.

Goldenberg, M. J. (2006). On evidence and evidence-based medicine: Lessons from the philosophy of science. *Social Science & Medicine, 62*(11), 2621-2632.

Goldman, A. (2008). Immediate Justification and Process Reliabilism. In Q. Smith, (Eds). *Epistemology: New Essays* (pp.63-82). Oxford: Oxford University Press.

Goldman, A. I. (1976). What is Justified Belief?. In G.S. Pappas (eds.) *Justification and Knowledge* (pp.1-23*)*. Dordrecht: D. Reidel.

Goodheart, C. D. (2005). Placing psychotherapy case studies within the framework of the APA evidence-based practice in psychology (EBPP) model. *Pragmatic Case Studies in Psychotherapy, 1*(3), 1-10.

Goodheart, C. D., Kazdin, A. E., & Sternberg, R. J. (2006). *Evidence-based psychotherapy: Where practice and research meet*. Washington, D.C: American Psychological Association.

Gorman, D. M., & Conde, E. (2010). The making of evidence-based practice: The case of project ALERT. *Children and Youth Services Review, 32*(2), 214-222.

Gorman, D. M., & Huber, J. C., Jr. (2009). The social construction of "evidence-based" drug prevention programs: A reanalysis of data from the drug abuse resistance education. *Evaluation Review, 33*(4), 396.

Grant, E., & Nelson, B. (1962). Hypotheses in late medieval and early modern science. *Daedalus, 91*(3), 599-616.

Gray, S.H. (2002). Evidence-based psychotherapeutics. *Journal of the American Academy of Psychoanalysis, 30*(1), 3-16.

Gregg, A. (Chairman.) *The place of psychology in an ideal university. The report of the University Commission to advise on the future of psychology at Harvard*. Cambridge: Harvard University Press, 1947. p. 42

Gruenbaum, A. (1984). *The Foundations of Psychoanalysis: A Philosophical Critique*. Berkeley, CA: University of California Press.

Guyatt, G. H. (1991) Evidence-based medicine (Editorial). American College of Physicians Journal Club. *Annals of Internal Medicine*, *114*(Suppl. 2), A16.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hacking, I. (2000). *The Social Construction of What*. Boston: Harvard University Press.

Hankinson, R. J. (1987). Causes and empiricism. *Phronesis: A Journal for Ancient Philosophy, 32*(1-3), 329-329.

Harman, G. (1986). *Change in view: Principles of reasoning*. Boston: MIT Press.

Harper, W. L (2011). Isaac newton's scientific method: Turning data into evidence about gravity and cosmology. Oxford: Oxford University Press.

Harrison, P. (2006). Miracles, early modern science, and rational religion. *Church History: Studies in Christianity and Culture, 75*(3), 493-510.

Haynes, S. N., Kaholokula, J. K., & Nelson, K. (1999). The idiographic application of nomothetic, empirically based treatments. *Clinical Psychology: Science and Practice, 6*(4), 456-461.

Hempel, C. (1945). Studies in the logic of confirmation, *Mind, 54*, 1-26.

Hempel, C. (1960). Inductive inconsistencies, *Synthese, 12*, 439-469.

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

Henggeler, S. W., Halliday-Boykins, C. A., Cunningham, P. B., Randall, J., Shapiro, S. B., & Chapman, J. E. (2006). Juvenile drug court: Enhancing outcomes by integrating evidence-based treatments. *Journal of Consulting and Clinical Psychology, 74*(1), 42-54.

Hennessy, K. D., Finkbiner, R., & Hill, G. (2006). The national registry of evidence-based programs and practices: A decision-support tool to advance the use of evidence-based services.*International Journal of Mental Health, 35*(2), 21-34.

Henry, J. (2001). The scientific revolution and the origins of modern science. New York, NY: Palgrave.

Hernandez Marcos, M. (2009). Conceptual Aspects of Legal Enlightenment in Europe. In E. Pattaro (Ed.), (2005). *A treatise of legal philosophy and general jurisprudence* (pp.69-134). Dordrecht: Springer.

Ho, H. L. (2004). The legitimacy of medieval proof. *The Journal of Law and Religion, 19*(2), 259-289.

Hoagwood, K. (2002). Making the translation from research to its application: The Je ne sais pas of evidence-based practices. *Clinical Psychology: Science and Practice, 9*, 210-213.

Home, A., & Biggs, T. (2005). Evidence-based practice in the real world: A group for mothers of children with invisible disabilities. *Groupwork an Interdisciplinary Journal for Working with Groups, 15*(2), 39-39.

Hunsberger, P. (2007). Reestablishing clinical psychology's subjective core. *American Psychologist*, *62*(6), 614-615.

Hunsley, J. (2007a). Addressing key challenges in evidence-based practice in psychology. *Professional Psychology: Research and Practice, 38*, 113-121.

Hunsley, J. (2007b). Training psychologists for evidence-based practice. *Canadian Psychology/Psychologie canadienne*, *48*(1), 32-42.

Hunsley, J., & Johnston, C. (2000). The role of empirically supported treatments in evidence-based psychological practice: A Canadian perspective. *Clinical Psychology: Science and Practice*, *7*(3), 269-272.

Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment, 17*(3), 251-255.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*(1), 29-51.

Hunter, D. J. (2009). Relationship between evidence and policy: A case of evidence-based policy or policy-based evidence? *Public Health, 123*(9), 583-586.

Imel, Z. E., & Wampold, B. E. (2008). The importance of treatment and the science of common factors in psychotherapy. In S. D. Brown & R. W. Lent (Eds.), Handbook of Counseling Psychology (4th ed, pp. 249-266). New York: John Wiley.

Ingraham, C. L., & Oka, E. R. (2006). Multicultural issues in evidence-based interventions. *Journal of Applied School Psychology, 22*(2), 127-149.

Institute of Medicine (U.S.). Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, D.C: National Academy Press.

Jacobs, S., Kissil, K., Scott, D., & Davey, M. (2010). Creating synergy in practice: Promoting complementarity between evidence-based and postmodern approaches. *Journal of Marital and Family Therapy, 36*(2), 185-196.

James, W. (1918). *The principles of psychology.* New York: Henry Holt & Co.

Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 39*(6), 885-896.

Johns, A. (2004). Print and Public Science. In. R. Porter (Ed.), *The Cambridge History of Science – Volume 4: Eighteenth Century Science* (pp. 536-560). Cambridge: Cambridge University Press.

Joiner, T. E., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment, 17*(3), 267-277.

Joyce, J. M. (2005). How probabilities reflect evidence. *Philosophical Perspectives, 19*(1), 153-178.

Katon, W. J., Zatzick, D., Bond, G., & Williams, J. (2006). Dissemination of evidence-based mental health interventions: Importance to the trauma field. *Journal of Traumatic Stress,19*(5), 611-623.

Kazak, A. E., Hoagwood, K., Weisz, J. R., Hood, K., Kratochwill, T. R., Vargas, L. A., & Banez, G. A. (2010). A meta-systems approach to evidence-based practice for children and adolescents. *The American Psychologist, 65*(2), 85-97.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *The American Psychologist, 63*(3), 146-159.

Keel, P. K., & Haedt, A. (2008). Evidence-based psychosocial treatments for eating problems and eating disorders. Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 37(1), 39-61.

Kelly, T. (2006). Evidence. In The Stanford Encyclopedia of Philosophy, E. N. Zalta (ed.). Retrieved from http://plato.stanford.edu/archives/fall2006/entries/evidence/.

Kelly, T. (2008). Evidence: Fundamental concepts and the phenomenal conception. *Philosophy Compass, 3*(5), 933-955.

Kelvin, R. G. (2005). Capacity of tier 2/3 CAMHS and service specification: A model to enable evidence based service development. *Child and Adolescent Mental Health, 10*(2), 63-73.

Keynes, J.M. (1921). *A treatise on probability*. London: Macmillan.

Kim, J. (1988). What is naturalized epistemology? In J. Tomberlin (Ed.) *Philosophical Perspectives 2, Epistemology (381-405)*. Atascadero, CA: Ridgeview Publishing Co).

King, B.D., Viney, W., & Woody, D.W. (2009). *A History of Psychology: Ideas and Context*. Boston: Pearson Education, Inc.

King, P., & Ballantyne, N. (2009). Augustine on testimony. *Canadian Journal of Philosophy, 39*(2), 195.

King, R. (1998). Evidence-based practice: Where is the evidence? The case of cognitive behaviour therapy and depression. *Australian Psychologist*, *33*(2), 83-88.

Kratochwill, T. R. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *The American Psychologist, 62*(8), 826.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

La Roche, M. J., & Christopher, M. S. (2009). Changing paradigms from empirically supported treatment to evidence-based practice: A cultural perspective. *Professional Psychology: Research and Practice, 40*(4), 396-402.

Laird, W. R. (1987). Robert Grosseteste on the subalternate sciences. *Traditio, 43*, 147.

Lambert, M. J. (1992). Psychotherapy research and its achievements. In D. K. Freedheim (Ed.), *History of psychotherapy: A century of change*. Washington, D.C: American Psychological Association.

Lamiell, J.T. (2012). Psychology's struggle for existence: Appraising wundt's pessimistic prediction a century on. Division 24 Presidential Address delivered at the 120th Annual Convention of the American Psychological Association, Orlando, FL.

Langbein, J. H. (1996). Historical foundations of the law of evidence: A view from the Ryder sources. *Columbia Law Review, 96*, 1168-1202.

Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of Evidence-Based treatments: Examples from parent training. *Clinical Psychology: Science and Practice, 13*(4), 295-310.

Lee, C. M. (2007). From clinical trials to professional training: A graduate course in evidence-based interventions for children, youth, and families. *Training and Education in Professional Psychology, 1*(3), 215-223.

Lehman, A. F. (2010). Adopting evidence-based practices--our hesitation waltz. *Schizophrenia Bulletin, 36*(1), 1-2.

Levant, R. F., & Hasan, N. T. (2008). Evidence-based practice in psychology. *Professional Psychology: Research and Practice, 39*(6), 658-662.

Lewis, J. I. (2009). The crossroads of countertransference and attribution theory: Reinventing clinical training within an evidence-based treatment world. *The American Journal of Psychoanalysis, 69*(2), 106-120.

Locke, J. (1690/1998). An essay concerning human understanding. In R. Ariew & E. Watkins (Eds.), *Modern Philosophy: An Anthology of Primary Sources* (pp. 270-373). Indianapolis, IN: Hackett Publishing Company.

Logsdon, R. G., McCurry, S. M., & Teri, L. (2007). Evidence-based psychological treatments for disruptive behaviors in individuals with dementia. *Psychology and Aging, 22*(1), 28-36.

Lovasz, N., & Slaney, K.L.(2012) What makes a hypothetical construct “hypothetical”? Tracing the origins and uses of the ‘hypothetical construct’ concept in psychological science. *New Ideas in Psychology, 31*(3), 22-31.

Luborsky, L. (1954). A note of Eysenck's article 'the effects of psychotherapy: An evaluation. *The British Journal of Medical Psychology, 45*(2), 129-131.

Lyotard, J. F. (1984). *The postmodern condition: A report on knowledge*. Minneapolis: University of Minnesota Press.

Mace, C. (2006). Setting the world on wheels: Some clinical challenges of evidence-based practice. *Group Analysis, 39*(3), 304-320.

Maclean, I. (2000). Evidence, logic, the rule and the exception in renaissance law and medicine. *Early Science and Medicine,5*(3), 227-227.

Manetti, G. (1993). *Theories of the sign in classical antiquity*. Bloomington: Indiana University Press.

Martin, J. (2010). Self-Concept as persons' understanding and evaluation of their own actions and experiences: Looking backward and forward from where we are. In Urdan, T. & Karabenick, S. (Eds.). *Advances in motivation and achievement, Vol. 16*. London: Pergamon/Elsevier.)

McCabe, O. L. (2004). Crossing the quality chasm in behavioral health care: The role of evidence-based practice. *Professional Psychology: Research and Practice, 35*(6), 571-579.

McCabe, O. L. (2006). Evidence-based practice in mental health accessing, appraising, and adopting research data.*International Journal of Mental Health, 35*(2), 50-69.

McCurry, S. M., Logsdon, R. G., Teri, L., & Vitiello, M. V. (2007). Evidence-based psychological treatments for insomnia in older adults. *Psychology and Aging, 22*(1), 18-27.

McDonald, P. W., & Viehbeck, S. (2007). From evidence-based practice making to practice-based evidence making: Creating communities of (research) and practice. *Health Promotion Practice, 8*(2), 140-144.

McFall, R. M. (2005). Theory and utility--key themes in evidence-based assessment: Comment on the special section. *Psychological Assessment, 17*(3), 312-323.

McGrath, M. (2012). Propositions. In *The Stanford Encyclopedia of Philosophy,* E. N. Zalta (ed.). Retrieved from http://plato.stanford.edu/archives/sum2012/entries/propositions.

McReynolds, P. (1997). *Lightner Witmer: His life and times*. Washington, D.C: American Psychological Association.

Messer, S. (2004). Evidence-based practice: Beyond empirically supported treatments. *Professional Psychology: Research and Practice*, *35*(6), 580-588.

Meehl, Paul E. (Paul Everett). (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Michelson, J. (2004). Critique of (im)pure reason: Evidence-based medicine and common sense. *Journal of Evaluation in Clinical Practice*, *10*(2), 157-161.

Michie, S., & Abraham, C. (2004). Interventions to change health behaviours: Evidence-based or evidence-inspired?*Psychology & Health, 19*(1), 29-49.

Miller, W. R., Sorensen, J. L., Selzer, J. A., & Brigham, G. S. (2006). Disseminating evidence-based practices in substance abuse treatment: A review with suggestions. *Journal of Substance Abuse Treatment, 31*(1), 25-39.

Milton, M. (2002). Evidence-based practice: Issues for psychotherapy. *Psychoanalytic Psychotherapy, 16*(2), 160-172.

Misak, C. J. (2010). Narrative evidence and evidence-based medicine. *Journal of Evaluation in Clinical Practice, 16*(2), 392-397.

Molewijk, A., Stiggelbout, A., Otten, W., Dupuis, H., & Kievit, J. (2003). Implicit normativity in evidence-based medicine: A plea for integrated empirical ethics eesearch. *Health Care Analysis*, *11*(1), 69-92.

Molloy, G. N., Murphy, G. C., & King, N. J. (2007). On the decline of n = 1 research in 'behaviour change': The rise of the evidence-based practice movement as one explanation for the trend. *Behaviour Change, 24*(2), 114-121.

Montuschi, E. (2009). Questions of evidence in evidence-based policy. *Axiomathes, 19*(4), 425-439.

Morawski, J. G. (1988). *The rise of experimentation in American psychology*. New Haven, CT US: Yale University Press.

Morrow, W.R. The development of psychological internship training. *Journal of Consulting Psychology, 1946*, 10, 165-183.

Morstyn, R. (2010). How the philosophy of merleau-ponty can help us understand the gulf between clinical experience and the doctrine of evidence-based psychotherapy. *Australasian Psychiatry : Bulletin of Royal Australian and New Zealand College of Psychiatrists, 18*(3), 221-225.

Muñoz, R. F., & Mendelson, T. (2005). Toward evidence-based interventions for diverse populations: The san francisco general hospital prevention and treatment manuals. *Journal of Consulting and Clinical Psychology, 73*(5), 790-799.

Muntersbjorn, M. M. (2003). Francis bacon's philosophy of science. *Philosophy of Science, 70*(5), 1137-1148.

Nathan, P. (2004). The evidence base for evidence-based mental health treatments: Four continuing controversies. *Brief Treatment and Crisis Intervention*, *4*(3), 243-254.

Nelson, T. D., & Nelson, J. M. (2010). Evidence-based practice and the culture of adolescence. *Professional Psychology: Research and Practice, 41*(4), 305-311.

Newnes, C. (2002). The rhetoric of evidence-based practice. *Ethical Human Sciences & Services*, *4*(2), 121-130.

Newton, I. (1678). *The Mathematical Principles of Natural Philosophy* (trans. A. Motte). Retrieved from http://books.google.ca/books/about/The_Mathematical_Principles_of_Natural_P.html?id=Tm0FAAAAQAAJ&redir_esc=y

Newton, I., Sir. (1730). *Opticks: Or, a treatise of the reflections, refractions, inflections and colours of light*. the fourth edition, corrected. London: printed for William Innys.

Norcross, J. C., Beutler, L. E., & Levant, R. F. (Eds.) (2005). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington, DC: APA Books.

Norcross, J. C. (Ed.). (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient needs*. New York: Oxford University Press.

Oldham, J. (1994). Truth-telling in the eighteenth-century english courtroom. *Law and History Review, 12*(1), 95-121.

O'Donnell, J. M. (1979). The clinical psychology of Lightner Witmer: A case study of institutional innovation and intellectual change. *Journal of the History of the Behavioral Sciences*

Pagoto, S. L., Spring, B., Coups, E. J., Mulvaney, S., Coutu, M., & Ozakinci, G. (2007). Barriers and facilitators of evidence-based practice perceived by behavioral science health professionals. *Journal of Clinical Psychology, 63*(7), 695-705.

Paris, J. (2000). Canadian psychiatry across 5 decades: From clinical inference to evidence-based practice. *The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie, 45*(1), 34-39.

Parker, G., & Fletcher, K. (2007). Treating depression with the evidence-based psychotherapies: A critique of the evidence.*Acta Psychiatrica Scandinavica, 115*(5), 352-359.

Patterson, J. E., Miller, R. B., Carnes, S., & Wilson, S. (2004). Evidence-based practice for marriage and family therapists.*Journal of Marital and Family Therapy, 30*(2), 183-195.

Pawson, R., Owen, L., & Wong, G. (2010). The today programme's contribution to evidence-based policy. *Evaluation: The International Journal of Theory, Research and Practice*, *16*(2), 211-213.

Pelham, J., William E, & Fabiano, G. A. (2008). Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 37*(1), 184-214.

Plantinga, A. (1993). *Warrant and proper function*. New York: Oxford University Press.

Pollio, D. E. (2003). The evidence-based group worker. *Social Work with Groups, 25*(4), 57-70.

Pollio, D. E., & MacGowan, M. J. (2010). The andragogy of evidence-based group work: An integrated educational model.*Social Work with Groups, 33*(2-3), 195-209.

Popper, K. (2004). Conjectures and refutations : The growth of scientific knowledge (Reprinted. ed.). London: Routledge.

Raimy, V.C. (Ed.). (1950). *Training in clinical psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Ramchandani, P., Joughin, C., & Zwi, M. (2001). Evidence-based child and adolescent mental health services: Oxymoron or brave new dawn? *Child Psychology and Psychiatry Review,6*(2), 59-64.

Ramey, H. L., & Grubb, S. (2009). Modernism, postmodernism and (evidence-based) practice. *Contemporary Family Therapy,31*(2), 75-86.

Redmond, C., Spoth, R. L., Shin, C., Schainker, L. M., Greenberg, M. T., & Feinberg, M. (2009). Long-term protective factor outcomes of evidence-based interventions implemented by community teams through a community-university partnership. *Journal of Primary Prevention, 30*(5), 513-530.

Reill, P.H. (2003). The Legacy of the "Scientific Revolution": Science and the Enlightenment. In. R. Porter (Ed.), *The Cambridge History of Science – Volume 4: Eighteenth Century Science* (pp. 23-43). Cambridge: Cambridge University Press.

Reyes, A.D., & Kazdin, A.E. (2008) When the evidence says, "yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science, 17*(1), 47-51.

Reynolds, S. (2000). Evidence based practice and psychotherapy research. *Journal of Mental Health*, *9*(3), 257-266.

Richards, I.A. (1932). *Mencius on the mind: Experiments in multiple definition*. London: Routledge & Kegan Paul.

Richards, P. S., & Worthington, E. L., Jr. (2010). The need for evidence-based, spiritually oriented psychotherapies. *Professional Psychology, Research and Practice, 41*(5), 363-370.

Rivett, M., Howarth, E., & Harold, G. (2006). Watching from the stairs': Towards an evidence-based practice in work with child witnesses of domestic Violence. *Clinical Child Psychology and Psychiatry, 11*(1), 103-125.

Rogers, C. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21* (2): 95-103

Rohricht, F. (2009). Body oriented psychotherapy. the state of the art in empirical research and evidence-based practice: A clinical perspective. *Body, Movement and Dance in Psychotherapy, 4*(2), 135-156.

Rorty, R. (1989). *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press.

Rose, N. S. (1998). *Inventing our selves: Psychology, power, and personhood*. Cambridge, England: Cambridge University Press.

Rosenberg, L. (2010). The reality of implementing evidence-based practices. *The Journal Of Behavioral Health Services & Research*, *37*(1), 1-3.

Rosner, R. I. (2005). Psychotherapy Research and the National Institute of Mental Health, 1948-1980. In W. R. Pickren & S. F. Schneider (Eds.), *Psychology and the National Institute of Mental Health: A historical analysis of science, practice, and policy* (pp. 113-150). Washington, DC US: American Psychological Association.

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural And Cognitive Psychotherapy*, *36*(2), 129-147.

Routh, D. K. (2000). Clinical psychology training. A history of ideas and practices prior to 1946. *The American Psychologist, 55*(2), 236-241.

Russell, B. (1910-1911). Knowledge by acquaintance and knowledge by description. *Aristotelian Society, Proceedings,11*, 209-232.

Ruzek, J. I., & Rosen, R. C. (2009). Disseminating evidence-based treatments for PTSD in organizational settings: A high priority focus area. *Behaviour Research and Therapy, 47*(11), 980-989.

Rychlak, J.F. (1959). Clinical psychology and the nature of evidence. *American Psychologist, 14,* 642-648.

Sackett D.L. (1986) Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest, 89*, 2s–3s.

Salyers, M. P., & Tsemberis, S. (2007). ACT and recovery: Integrating evidence-based practice and recovery orientation on assertive community treatment teams. *Community Mental Health Journal, 43*(6), 619-641.

Sanderson, W. (1998). The case for evidence-based psychotherapy treatment guidelines. *American Journal of Psychotherapy*, *52*(3), 382-387.

Scheyett, A. (2006). Silence and surveillance: Mental illness, evidence-based practice and a foucaultian lens. *Journal of Progressive Human Services, 17*(1), 71-92.

Schum, D. A. (2005). *Thoughts About a Science of Evidence.* Retrieved from http://128.40.111.250/evidence/content/Science.doc.

Schwalbe, C. S. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review, 30*(12), 1458-1464.

Scogin, F. (2007). Introduction to the special section on evidence-based psychological treatments for older adults. *Psychology and Aging, 22*(1), 1-3.

Serene, E. F. (1979). Robert Grosseteste on induction and demonstrative science. *Synthese, 40*(1), 97-115.

Serjeantson, R.W. (2006). Proof and persuasion. In. K. Park & L. Daston (Eds.), *The Cambridge History of Science – Volume 3: Early Modern Science* (pp. 132-175). Cambridge: Cambridge University Press.

Sexton, T. L., & Kelley, S. D. (2010). Finding the common core: Evidence-based practices, clinically relevant evidence, and core mechanisms of change. *Administration and Policy in Mental Health, 37*(1-2), 81-88.

Shakow, D. (1976). What is clinical psychology? *American Psychologist, 31*(8), 553-560.

Shapin, S., Schaffer, S., & Hobbes, T. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the Experimental Life* : Including a translation of Thomas Hobbes, Dialogus Physicus de Natura Aeris by Simon Schaffer. Princeton, N.J: Princeton University Press.

Shapin, S. (1996). *The Scientific Revolution.* Chicago: University of Chicago Press.

Shapiro, D. (1996). 'Validated' treatments and evidence-based psychological services. *Clinical Psychology: Science and Practice*, *3*(3), 256-259.

Sharpe, M., Gill, D., Strain, J., & Mayou, R. (1996). Psychosomatic medicine and evidence-based treatment. *Journal of Psychosomatic Research, 41*(2), 101-107.

Shedler, J. (2010) The efficacy of psychodynamic psychotherapy. *American Psychologist*, 65, 98-109.

Shlonsky, A., & Gibbs, L. (2004). Will the real evidence-based practice please stand up? teaching the process of evidence-based practice to the helping professions. *Brief Treatment and Crisis Intervention, 4*(2), 137-153.

Silverman, D. (2005). What works in psychotherapy and how do we know?: What evidence-based practice has to offer. *Psychoanalytic Psychology*, *22*(2), 306-312.

Silverman, W. K., & Hinshaw, S. P. (2008). The second special issue on evidence-based psychosocial treatments for children and adolescents: A 10-year update. *Journal of Clinical Child & Adolescent Psychology, 37*(1), 1-7.

Silverman, W. K., Ortiz, C. D., Viswesvaran, C., Burns, B. J., Kolko, D. J., Putnam, F. W., & Amaya-Jackson, L. (2008). Evidence-based psychosocial treatments for children and adolescents exposed to traumatic events. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 37*(1), 156-183.

Simpson, J. A. (Ed.). (1989). *Oxford English Dictionary Online*. Retrieved from http://dictionary.oed.com/

Singh, N. N., & Oswald, D. P. (2004). Evidence-based practice. part I: General methodology. *Journal of Child and Family Studies, 13*(2), 129-142.

Siraisi, N. G. (2004). The fielding H. garrison lecture : Medicine and the renaissance world of learning. *Bulletin of the History of Medicine, 78*(1), 1-36.

Skyrms, B. (1980). *Causal necessity*. London: Yale University Press

Slaney, K. L. (2012). Laying the Cornerstone of Construct Validity Theory: Herbert Feigl's influence on early specifications. Theory and Psychology, 22, 290-309.

Slife, B. D., Wiggins, B. J., & Graham, J. T. (2005). Avoiding an EST monopoly: Toward a pluralism of philosophies and methods. *Journal of Contemporary Psychotherapy, 35*(1), 83-97.

Smith, Mary Lee; Glass, Gene V; & Miller, Thomas I. (1980). *The Benefits of Psychotherapy*. Baltimore, MD: Johns Hopkins Univ. Press.

Somerfield, M., & Costa, P. (1999). Toward more evidence-based guidelines for psychology. *American Psychologist*, *54*(12), 1131-1132.

Spielmans, G. I., Gatlin, E. T., & McFall, J. P. (2010). The efficacy of evidence-based psychotherapies versus usual care for youths: Controlling confounds in a meta-reanalysis.*Psychotherapy Research : Journal of the Society for Psychotherapy Research, 20*(2), 234-246.

Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *Journal of Clinical Psychology, 63*(7), 611-631.

Strevens, M. (2004). Bayesian confirmation theory: Inductive logic, or mere inductive Framework? *Synthese, 141*, 365-379.

Stricker, G. (2006). The local clinical scientist, evidence-based practice, and personality assessment. *Journal of Personality Assessment, 86*(1), 4-9.

Strupp, H.H., & Howard, K.I. (1992). A brief history of psychotherapy research. In D. K. Freedheim (Ed.), *History of psychotherapy: A century of change (pp. 309-334)*. Washington, D.C: American Psychological Association.

Stuart, R. B., & Lilienfeld, S. O. (2007). The evidence missing from evidence-based practice. *American Psychologist, 62*, 615-616.

Sturdee, P. (2001). Evidence, influence or evaluation? Fact and value in clinical science. In C. Mace, S. Moorey,& B. Roberts (Eds.), *Evidence in the psychological therapies: A critical guide for practitioners* (pp. 61-79). London: Brunner-Routledge.

Sturmey, P. (2009). Behavioral activation is an evidence-based treatment for depression. *Behavior Modification, 33*(6), 818-829.

Sue, S., Zane, N., Levant, R. F., Silverstein, L. B., Brown, L. S., Olkin, R., & Taliaferro, G. (2006). How well do both evidence-based practices and treatment as usual satisfactorily address the various dimensions of diversity?. In J. C. Norcross, L. E. Beutler, R. F. Levant (Eds.) , *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 329-374). Washington, DC US: American Psychological Association.

Tanenbaum, S. (2003). Evidence-based practice in mental health: Practical weaknesses meet political strengths. *Journal of Evaluation in Clinical Practice, 9*(2), 287-301.

Tanenbaum, S. J. (2006). Expanding the terms of the debate: Evidence-Based Practice and public policy. In C. D. Goodheart, A. E. Kazdin, R. J. Sternberg, C. D. Goodheart, A. E. Kazdin & R. J. Sternberg (Eds.), *Evidence-based psychotherapy: Where practice and research meet.* (pp. 239-259). Washington, DC US: American Psychological Association.

The Cochrane Collaboration. (n.d.). *History*. Retrieved from http://www.cochrane.org/about-us/history

Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71*(2), 181-194.

Thorn, B. E., Cross, T. H., & Walker, B. B. (2007). Meta-analyses and systematic reviews of psychological treatments for chronic pain: Relevance to an evidence-based practice. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association, 26*(1), 10-12.

Tibbits,M.K., Bumbarger, B.K., Kyler, S.J., & Perkins, D.F. (2010). Sustaining evidence-based interventions under real-world conditions: Results from a large-scale diffusion project. *Prevention Science, 11*(3), 252-262.

Tonelli, M. R. (2006). Integrating evidence into clinical practice: An alternative to evidence-based approaches. *Journal of Evaluation in Clinical Practice, 12*(3), 248-256.

Treasure, J., & Kordy, H. (1998). Evidence based care of eating disorders: Beware the glitter of the randomised controlled trial.*European Eating Disorders Review, 6*(2), 85-95.

Tucker, J. A., & Roth, D. L. (2006). Extending the evidence hierarchy to enhance evidence-based practice for substance use disorders. *Addiction, 101*(7), 918-918.

Twining, W. L. (2006). *Rethinking evidence: Exploratory essays*. Cambridge: Cambridge University Press.

Upadhyay, U. D., Cockrill, K., & Freedman, L. R. (2010). Informing abortion counseling: An examination of evidence-based practices used in emotional care for other stigmatized and sensitive health issues. *Patient Education and Counseling,81*(3), 415-421.

Veerman, J. W., & van Yperen, T. A. (2007). Degrees of freedom and degrees of certainty: A developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning, 30*(2), 212-221.

Wachtel, P. L. (2010). Beyond "ESTs": Problematic assumptions in the pursuit of evidence-based practice. *Psychoanalytic Psychology, 27*(3), 251-272.

Walker, B. B., & London, S. (2007). Novel tools and resources for evidence-based practice in psychology. *Journal of Clinical Psychology, 63*(7), 633-642.

Waller, G. (2009). Evidence-based treatment and therapist drift.*Behaviour Research and Therapy, 47*(2), 119-127.

Walsh, K., & Hope, D. A. (2010). LGB-affirmative cognitive behavioral treatment for social anxiety: A case study applying evidence-based practice principles. *Cognitive and Behavioral Practice, 17*(1), 56-65.

Wampold, B. (2002). An examination of the bases of evidence-based interventions. *School Psychology Quarterly*, *17*(4), 500-507.

Wampold, B. E. (2000). Outcomes of individual counseling and psychotherapy: Empirical evidence addressing two fundamental questions. In S. D. Brown & R. W. Lent (Eds.), *Handbook of Counseling Psychology* (3rd ed, pp. 711-739). New York: John Wiley.

Wampold, B. E., Goodheart, C. D., & Levant, R. (2007). Clarification and elaboration on evidence-based practice in psychology. *American Psychologist, 62*, 616-618

Wampold, B. E., Ollendick, T. H., & King, N. J. (2006). Do therapies designated as empirically supported treatments for specific disorders produce outcomes superior to non-empirically supported treatment therapies?. In J. C. Norcross, L. E. Beutler, R. F. Levant (Eds.) , *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 299-328). Washington, DC US: American Psychological Association.

Wampold, B., & Bhati, K. (2004). Attending to the Omissions: A Historical examination of evidence-based practice movements. *Professional Psychology: Research and Practice*, *35*(6), 563-570.

Weisz, J. R., & Gray, J. S. (2008). Evidence-Based psychotherapy for children and adolescents: Data from the present and a model for the future. *Child and Adolescent Mental Health, 13*(2), 54-65.

Wendt, D. C. J., & Slife, B. D. (2007). Is evidence-based practice diverse enough? Philosophy of science considerations. *American Psychologist, 62*, 613-614.

Westen, D. (1998). Unconscious thought, feeling, and motivation: The end of a century-old debate. In R. Bornstein & J. Masling (Eds.), *Empirical studies of unconscious processes*. Washington, D.C.: American Psychological Association Press.

Westen, D., & Bradley, R. (2005). Empirically supported complexity: Rethinking evidence based practice in psychotherapy. *Current Directions in Psychological Science, 14,* 266-271.

Wifstad, A. (2008). External and internal evidence in clinical judgment: The evidence-based medicine attitude. *Philosophy, Psychiatry, & Psychology, 15*(2), 135-139.

Wigmore, J. H., 1863-1943. (1913). The principles of judicial proof as given by logic, psychology, and general experience, and illustrated in judicial trials. Boston: Little, Brown & Company.

Wild, T. C. (2006). Social control and coercion in addiction treatment: Towards evidence-based policy and practice. *Addiction (Abingdon, England), 101*(1), 40-49.

Willenbring, M. L., Kivlahan, D., Kenny, M., Grillo, M., Hagedorn, H., & Postier, A. (2004). Beliefs about evidence-based practices in addiction treatment: A survey of veterans administration program leaders. *Journal of Substance Abuse Treatment, 26*(2), 79-85.

Williams, R., & Fulford, K. W. M. (2007). Evidence-based and values-based policy, management and practice in child and adolescent mental health services. *Clinical Child Psychology and Psychiatry, 12*(2), 223-242.

Williamson, T. (1997). Knowledge as evidence. *Mind,106*, 717-741.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

Wilson, J. L., Armoutliev, E., Yakunina, E., & Werth, J. L. (2009). Practicing psychologists' reflections on evidence-based practice in psychology. *Professional Psychology: Research and Practice, 40*(4), 403-409.

Winston, A. S. (1990). Robert Sessions Woodworth and the "Columbia Bible": How the psychological experiment was redefined. *The American Journal of Psychology, 103*(3), 391-401.

Wittgenstein, L. (1953). *Philosophische Untersuchungen.* Frankfurt am Main: Suhrkamp Verlag.

Wolfe, J. (1999). Overcoming barriers to evidence-based practice: Lessons from medical practitioners. *Clinical Psychology: Science and Practice*, *6*(4), 445-448.

Worrall, J. (2010). Evidence: Philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice, 16*(2), 356-362.

Wundt, W. (1913). Die psychologie im kampf ums dasein, zweite Auflage. Leipzig: Kröner. Retrieved from http://philosophiebuch.de/wundtpsy.htm.

Wyer, P. C., & Silva, S. A. (2009). Where is the wisdom? A conceptual history of evidence-based medicine. *Journal of Evaluation in Clinical Practice, 15*(6), 891.

Yamada, A., & Brekke, J. S. (2008). Addressing mental health disparities through clinical competence not just cultural competence: The need for assessment of sociocultural issues in the delivery of evidence-based psychosocial rehabilitation services. *Clinical Psychology Review, 28*(8), 1386-1399.

Yon, A., & Scogin, F. (2007). Procedures for identifying evidence-based psychological treatments for older adults. *Psychology and Aging, 22*(1), 4-7.

Zazzali, J. L., Sherbourne, C., Hoagwood, K. E., Greene, D., Bigley, M. F., & Sexton, T. L. (2008). The adoption and implementation of an evidence based practice in child and family mental health services organizations: A pilot study of functional family therapy in new york state. *Administration and Policy in Mental Health, 35*(1-2), 38-49.

Appendices

Appendix A.

List of Members of the APA Task Force on Evidence-based Practice

1. Carol D. Goodheart, EdD (Chair; Independent Practice, Princeton, NJ)
2. Ronald F. Levant, EdD (ex-officio; University of Akron)
3. David H. Barlow, PhD (Boston University)
4. Jean Carter, PhD (Independent Practice, Washington, DC)
5. Karina W. Davidson, PhD (Columbia University)
6. Kristofer J. Hagglund, PhD (University of Missouri—Columbia)
7. Steven D. Hollon, PhD (Vanderbilt University)
8. Josephine D. Johnson, PhD (Independent Practice, Livonia, MI)
9. Laura C. Leviton, PhD (Robert Wood Johnson Foundation, Princeton, NJ)
10. Alvin R. Mahrer, PhD (Emeritus, University of Ottawa)
11. Frederick L. Newman, PhD (Florida International University)
12. John C. Norcross, PhD (University of Scranton)
13. Doris K. Silverman, PhD (New York University)
14. Brian D. Smedley, PhD (The Opportunity Agenda, Washington, DC)
15. Bruce E. Wampold, PhD (University of Wisconsin)
16. Drew I. Westen, PhD (Emory University)
17. Brian T. Yates, PhD (American University)
18. Nolan W. Zane, PhD (University of California, Davis)
19. Professional American Psychological Association (APA) staff included Geoffrey M. Reed, PhD, and Lynn F. Bufka, PhD (Practice Directorate); Paul D. Nelson, PhD, and Cynthia D. Belar, PhD (Education Directorate); and Merry Bullock, PhD (Science Directorate).

Appendix B.

Interview Guide for APA Task Force Members

Instructions: While responding to these questions, please remember to avoid discussing particular opinions of specific other members of the Task Force and rather focus your responses on your own views, as well as topics discussed by the Task Force in general.

- What was your sense of the historical context that led to the formation of the Task Force in 2006?
- What are some of the political/social/disciplinary/organizational factors that you remember that impacted on the Task Force's conceptualizations of evidence?
- What were the circumstances by which you came to serve as a member of the APA Task Force on Evidence-Based Practice?
- From your perspective, how do you think the Task Force conceptualized the role of evidence in psychological research and practice?
- What were some of the considerations and deliberations of the Task Force concerning the role of evidence in psychological research and practice that you recall?
- How do you conceptualize the role of evidence in psychological research and practice?
- In what ways does your own conceptualization overlap or diverge from the conceptualization of the Task Force?
- How do you think the Task Force would define evidence?
- How would you define evidence? What do you think evidence is or is not?
- How does your own definition of evidence diverge or overlap with that of the Task Force?
- What considerations did the Task Force have to make during its deliberations about the nature of evidence?
- What else do you remember about your experience serving as a member of the Task Force that may be relevant to conceptualizations of evidence?
- What are your thoughts about the manner in which the Task Force's statement has impacted psychology since its publication?
- What do you think the aims of the Task Force were with respect to the role of evidence in psychology? In what ways were those aims accomplished and in what ways have they yet to be achieved?
- What are your thoughts on the current role of evidence in psychology?
- How do you envision the role of evidence in psychological research and practice in the future?