Facing Facts about 'Information' in Psychology

by

Noah Susswein

M.A. (Psychology), Simon Fraser University, 2008
B.A. (Psychology), Rutgers University, 2000

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

Fall 2013

Approval

Name: **Noah Susswein**

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

Title of Thesis: ***Facing Facts about 'Information' in Psychology***

Examining Committee: **Chair:** Robert Ley
Professor

Martin, Jack
Senior Supervisor
Professor

Slaney, Kathleen
Supervisor
Associate Professor

Maraun, Michael
Supervisor
Professor

Robin Barrow
Internal Examiner
Professor
Faculty of Education

Dr. Christopher Green
External Examiner
Professor
Department of Psychology
York University

Date Defended/Approved: November 29, 2013

Partial Copyright License

Abstract

The coherence of theorizing in terms of ‘information’ is generally taken for granted in the cognitive sciences. Despite widespread employment of ‘information’ as an apparently technical concept, no one knows what it means. Historical and conceptual investigations reveal that it is meaningless. A crucial distinction is drawn between ‘signals’ and ‘messages’, respectively, with regards to the role of these concepts in Claude Shannon’s (1948) information theory. Sharpening this woefully neglected distinction reveals the vastness of the gulf separating two bodies of work that are historically associated with the phrase ‘information theory.’ On the one hand, there is Shannon’s (1948) statistical theory of signal transmission; on the other, there are the incoherent, ‘information’-laden speculations of Norbert Wiener (1950) and Warren Weaver (1949).

It is demonstrated that certain of pioneering information-processing psychologist George Miller’s (1951, 1953, 1956) most celebrated works involve gross distortions of Shannon’s concepts, and conflation of Shannon’s (1948) mathematics with Wiener’s (1950) and Weaver’s (1949) respective metaphysics. The widespread misperception that information-processing psychology is substantively related to Shannon’s (1948) information theory is demonstrated to be false. Subsequently, the suggestion that it is reasonable for psychologists to employ ‘information’ and/or ‘information-processing’ as core concepts without defining these terms of art is sharply criticized.

However, throughout these critical investigations, care is taken to distinguish between conceptual clarity and empirical knowledge. It is demonstrated that information-processing psychologists can and have made important contributions to scientific knowledge despite the incoherence that is endemic in their metatheory. On the assumption that a helping hand is preferable to a pointed finger, these critical investigations constitute a therapy rather than an indictment.

Keywords: information-processing; information theory; cognitive revolution; history of psychology; George Miller

Acknowledgements

I would like to thank Jack Martin for being an ideal supervisor, and also, in alphabetical order: Jeremy Carpendale, Mix Bibok, Beverly Davino, Jim Marcia, Sarah Pong, Tim Racine, Kate Slaney, William Turnbull, Anita Turner, and Rene Weideman.

Table of Contents

Approval ii
Partial Copyright License iii
Abstract iv
Acknowledgements v
Table of Contents vi
List of Figures x

Chapter 1. Information as a Core Concept in Psychology 1
Defense of Methods 5
The Logical/Empirical Distinction 6
Clarity of Claims Versus Rules of Language 8

Chapter 2. Distinguishing Among Different Senses of 'information' 10
Shannon (1948) versus Weaver (1949) 11
On Definability and Internal Relations 12
The Everyday Epistemic Sense of 'information' 13
On Sources of 'information' versus Information 14
The Double-life Definition of Everyday Epistemic Information 15
Information, Knowledge, and Potentiality 16
Epistemic versus Semantic 16
Shannon's (1948) Statistical Concept of 'information' 19
Information Theory: The Big Picture 20
The Information-theoretic Sense of Communication 21
Selection Not Semantics 22
Two Crucial Observations 22
Only Signals Are Transmitted 23
Information, in the Statistical Sense, Cannot Be Transmitted 23
The Information-theoretic Sense of *Message* 24
Shannon's (1948) Statistical Sense of *Signal* 26
Shannon's (1948) Statistical Sense of *Encoding* 26
Shannon's (1948) Statistical Sense of 'information' 28
Signal Transmission Versus Message Composition, or on Selecting Versus Being Selected 29
The English Language as a *Message Source* 29
On Shannon's Contributions to Misunderstanding of His Own Work 33
On Applying Shannon's (1948) Concepts to Systems that Do Not Correspond to His Model 36
The ABCs of Weaver (1949) 37
Who Was Warren Weaver? 38
Weaver (1949) on Shannon (1948): Speculation Not Summary 39
Identifying Signals versus Understanding Messages 41
Decision versus Discovery 42

A Theory of Signal Transmission versus a Theory of What Signals Are Signals for 44
Weaver's Levels A and C Can Be Understood Mechanically: His Level B Cannot 45
Untangling Weaver 47
Control Information versus Shannon's Statistical Sense of 'information' 48
Communication as *Control* 49
Human-machine Interaction 50
Control Information versus Everyday Epistemic Information 51
Weaver's (1949) Representational Theory of Mind 52
Hartley's (1928) Representational Theory of Mind 54
Hartley's (1928) Psychologizing 55
Hartley's (1948) De-psychologizing 57
Conclusion 58

Chapter 3. George Miller and Information-processing Psychology 66
Documenting Distortions 67
Shannon's (1948) Math versus Wiener's (1948, 1950) Metaphor 67
Miller's (1951, 1953, 1956) Misunderstanding of Shannon (1948) 71
Perception and Sensation as Shannonian Communication 73
Markoff Processes and the Statistical Structure of Behavior 76
Serially Dependent Actions versus Serially Dependent Events 77
Serial-dependence as Learning 79
Information-theoretic versus Cognitive Uncertainty 80
Engineers' vs. Behavioral Psychologists' use of *Markoff Processes* 84
Mentalistic Interpretations of Behavioral Stereotypy 86
Information Measurement as a Research Method in Psychology 89
Shannon's Stipulations versus Natural Laws 89
'Amount of Information' vs. 'Degree of Informativeness' 92
On the Idea of a Statistical Process 95
What Kind of Information can be Acquired? 96
The Statistical Sense of 'information' Is Not a Quantified Version of Its Everyday Epistemic Sense 97
An Information-theoretic Analysis of a 16-state System 98
Statements versus Messages versus Events 101
On the Idea of Isolating Amounts of Information 103
Statements versus Messages in Shannon's (1948) Information-theoretic Sense 104
From Bits to Chunks: Miller's (1956) Magic 106
Baddeley (1994) on Miller (1956) 106
Information Theory and Information-Processing Psychology: The Continuity Interpretation 107
Miller's (1956) Verbal Innovations 108
Information, in the Statistical Sense, as *Variance* 109
On Information *that Is* in *Things* 111
'Information' and 'Correlation' 112

Research Participants as Communication Channels 112
Epistemic Agents versus Mere Media 114
Channel Capacities: Descriptive or Explanatory*?*..115
Per Unit Time vs per Unit Judgement..118
From Bits to Chunks..119
More- and Less-valid Extensions of Shannon's (1948) Information Theory 120
'Chunking' Clarified 122
Information-theoretic Analyses of Natural Languages 124
The Event of a Sequence versus a Sequence of Events 130
On Containing Bits of 'information' 132
Binary Digits versus Bits..134
On the Idea of Recalling Bits of Information 137
Miller's (1956) Concept of *Recoding*...138
A Language versus a Code 139
Miller (1989) on Miller (1956)...142
Shannon's (1948) Information Theory versus Miller's (1951, 1953, 1956) Information-processing Psychology 143

Chapter 4. Is the Information in *Information-Processing Psychology* Implicitly Defined? 145
Sloman's (2011) Methodological Assumptions..148
Criticism versus Conservatism 150
'Information Architectures' as the Referent of Psychological Predicates................152
Ambiguity as a Vice 152
Ambiguity as a Virtue 153
Defining versus Theorizing...156
Information and Inform-able Agents 157
Mind as Informational...158
On What Bases Can Organisms Be Ascribed 'Minds'? 159
Becoming Informed versus Using Information 160
Does Becoming Informed Involve Processing Information?...................................161
Capabilities versus Mechanisms 165
Information-Processing: Descriptive or Explanatory*?* 166
Bearers of Information versus Sources of Information 167
Information and Representation 169
The Homunculus Fallacy...171
What Is a Stance?...175
On Coin Sorting and Control Information...178
Multiple Distinctions versus Metaphysical Dichotomies 179
Information and Mereology 180
Who Can V? versus What Counts as Ving?..181
No Implicit Definition Found 182

Chapter 5. Conclusion 184

References 190

List of Figures

Figure 2.1. A reproduction of Shannon's (1948) model of a communication system. 21

Chapter 1.

Information as a Core Concept in Psychology

Information is a core concept in the cognitive sciences, especially in cognitive psychology. In 1979, Lachman, Lachman, and Butterfield wrote, “Information-processing psychology has entered a period of normal science. Its revolution has been over since around 1970” (p. 113). By invoking Kuhn’s (1962) notion of *normal science*, these authors assert that psychologists who theorize and hypothesize in terms of ‘information’ *processing*, are principally concerned with accumulating empirical facts expressed within a widely-shared theoretical framework and, by extension, a common vocabulary. Kuhn’s notion of normal science stands in contrast to his concept of *revolutionary science*, which involves challenging or attempting to test the underlying assumptions of a dominant theory. In this respect then, when talking about information-processing in psychology, Lachman and colleagues (1979) appear to be entirely correct in noting its unquestioned hegemony, despite a minority of dissenting voices (Bennett & Hacker, 2003; Brooks, 1991; Dreyfus, 1972; Heil, 1981; Shanker, 1998) and a few efforts at systematic defense (Dretske, 1981; Fodor, 1975, 2008; Sloman, 2011)—most of which come to psychology from other, related disciplines. In short, in contemporary psychology the coherence of theorizing and hypothesizing in terms of ‘information’ is generally taken for granted among psychologists.

It might be expected that a widely employed core concept would be precisely defined, such that when psychologists theorize in terms of ‘information’, it is clear what they are theorizing about. But, despite its widespread employment, there is a paucity of explanation concerning just what is denoted by the information in information-processing psychology. With admirable candor, Dretske (1981) admits that it is “much easier to talk about information than it is to say what it is that you are talking about” (p. ix), noting that the term is “fashionable” (p. viii) in that its “use in telecommunications and computer technology gives it a tough, brittle, technical sound” (p. viii) while it remains “spongy, plastic, and amorphous enough to be serviceable in cognitive and semantic studies” (p. viii). When:

> thinking about information, one tends to think of something objective and quantifiable—the electrical pulses surging down a copper wire, for example—and, at the same time, something more abstract, of the *news* or *message* that these pulses carry. (p. ix)

So, what exactly *is* information in information-processing? Dennett (2005) claims that “theorists have often found it useful to speak, *somewhat* impressionistically, about…information being processed” (p. 15). But what impression do such theorists have? And where did it come from? The prior conceptual and/or ontological question is complemented by the latter historical one: How did information become so fundamental a concept in psychology? Although these two questions are logically distinct, it is most helpful to consider them jointly, and a start date (or at least, a start decade: 1948-1958) for the employment of ‘information’ as a putatively technical term in psychology can be identified.

On the assumption that what the information in information-processing psychology *is* depends on what psychologists *mean* by information and processing, I will investigate the influential works of pioneering information-processing psychologist George Miller (1951, 1953, 1956) in Chapter 3 of this thesis. This method of ontological investigation presupposes what has become, in some quarters, an unfashionable distinction between logical and empirical inquiries.[1] It is also used in Chapter 4 in which Sloman’s (2011) recent attempt to identify an *implicit* definition of (the relevant sense of)

[1] Quine’s (1951) powerful attack on the metaphysical dichotomy between analytic and synthetic truths has been widely regarded as evidence that the ordinary distinction between logical (What is a ‘bachelor’?) and empirical concerns (What is the average height of bachelors in North America?) is misleading and artificial. Although the logical/empirical distinction will be defended here, space does not permit an investigation of the broader philosophical context in which this distinction was overblown and, subsequently, neglected. See Putnam (1989) and Bennett and Hacker (2003) for more on this topic.

‘information’ is reviewed. Throughout this thesis, the logical versus empirical distinction is described and defended. Sharpening this distinction will be helpful in demonstrating (Chapters 2 and 3) that the widespread perception of substantive continuity between Shannon’s (1948) information theory and information-processing psychology (e.g., Baddeley, 1994; Miller, 1951, 1956) is predicated upon gross distortions of ‘information’ theory. For example, consider the following gloss on the history of information-processing psychology:

> After a few years of research using information-theoretic concepts and measures, limitations on the applicability of some of the concepts began to show themselves. Precise measurement of the technical concept of information requires conditions that do not frequently obtain in the cognitive life of human beings…So cognitive psychologists have mostly abandoned the formalisms of ‘information’ theory. While few cognitive psychologists measure “information” in its technical sense nowadays, many still use the term…[at times] nearly as a synonym for knowledge. Contemporary research in the information-theoretic tradition is no longer concerned with quantity of ‘information’, but with the nature of psychological information and its structure … The way had been prepared by information theory. It had prepared us to think about information and knowledge in terms of abstract symbol systems. (Lachman et al., 1979, pp. 74-75)

This passage, which I take as exemplifying widely shared beliefs among psychologists and many other cognitive scientists, suffers from a variety of confusions.

First, it contains a historical error, as the notion that knowledge is to be understood in terms of abstract (mental) symbols has a long pedigree. This representational theory of mind clearly predates, and is substantively irrelevant to, Shannon’s (1948) information theory. Fodor (2008) succinctly clarifies that “a representational theory of mind…is a claim about the metaphysics of cognitive mental states and processes…[that] mental processes are…causal chains of…operations on mental representations” (pp. 3-4). And the representational theory of mind, broadly speaking, has “been the main line of thought among mental realists…arguably since Plato and Aristotle, patently since Decartes and the British Empiricists” (p. 6). In Chapter 2, it will be demonstrated that “to think about information and knowledge in terms of abstract symbol systems” is most definitely *not* to think in terms of Shannon’s

(1948) information theory. In that same chapter, Shannon's (1948) information theory will be clearly explicated, with an emphasis on establishing *the boundaries* of the range of phenomena to which Shannon's *information-theoretic* concepts, such as *channel capacity*, can be coherently applied. This involves describing the highly restricted sense in which such familiar non-technical terms as *communication*, *signal*, and/or *information* figure in Shannon's (1948) work.

Although Shannon's (1948) work itself is explicitly non-psychological, the works of Shannon's collaborators whose names are also closely associated with the phrase information theory—for example Wiener (1948, 1950) and Weaver (1949)—are littered with psychological speculation. Furthermore, 20 years prior, Shannon's predecessor at Bell Laboratories, R.V.L. Hartley (1928), published *Transmission of information,* which is widely regarded as the first reference to information as a measurement term (Bar-Hillel, 1955). Hartley's paper assumes a representational theory of mind in the form of what Harris (1981) describes as a *telementation* model of communication, which roughly amounts to the view that linguistic understanding involves a matching of representational mental states between speaker and hearer. The seeds of subsequent confusion regarding information are sown even in the *title* of Hartley's (1928) paper, as it erroneously conflates signal transmission with information measurement (see Chapter 2).

Chapter 3 then investigates certain influential works of George Miller (Miller, 1951, 1953, 1956; Miller & Frick, 1949). This section will demonstrate that Shannon's information-theoretic concepts were massively distorted by the pioneering information-processing psychologists who adopted Shannon's terminology without observing definitional restrictions associated with his technical *concepts*. In other words, information-processing psychologists borrowed certain of Shannon's words without also borrowing his ideas. So, although the information in information-processing psychology has its historical origins in Shannon's (1948) work, his theorems are not helpful in determining what the information in information-processing psychology is, *pace* distorted status quo accounts of this aspect of the discipline's history. For example, Luce (2003) writes:

> the word *information* has been almost seamlessly transformed into the concept of "information-processing models" in which information theory per se plays no role. The idea of the mind being an information-processing network with capacity limitations has stayed with us, but in far more complex ways than pure information theory. Much theorizing in cognitive psychology is of this type. (pp. 185-186)

It will be demonstrated below that "more complex" really means *less clear*. Bennett and Hacker (2003), to whose analyses the present work is much indebted, succinctly clarified the distinction between the everyday or semantic sense of 'information' and the information-theoretic sense.

> In the semantic sense, information is a set of true propositions…in the engineering sense, information is a measure of the freedom of choice in the transmission of a signal, and the amount of information is measured by the logarithm to the base 2 of the number of available choices. (p. 141)

This concise formulation benefits from elaboration, undertaken below. First, however, I defend my methods.

Defense of Methods

Knowing the history of a scientific discipline is an important aspect of understanding that discipline. However, the historical review below is more critical and analytic than descriptive, and conceptual analysis is a neglected method in disciplinary psychology; it is likely not even recognized as a method, per se. Nevertheless, I assume that in order to determine what the information in information-processing psychology *is,* we must examine the way that the term information is used by information-processing psychologists. Such an approach appears to be an extreme minority undertaking among disciplinary psychologists. Since roughly the mid-20th Century, psychology's principal approach to ontological questions has been quantitative and empirical rather than linguistic and logical (Maraun, 1998). Expressing a point of view that has become almost universally adopted among psychologists, Cronbach and Meehl (1955) wrote that "psychological processes are elusive" (p. 286) and that "scientifically speaking, to 'make clear what something *is'* means to set forth the laws in which it occurs" (p. 290). But how could a researcher set forth the laws in which "C" occurs without first knowing what C is? It is difficult to imagine how an individual could study a phenomenon that was impossible to identify. To define C is not only to explain what the term, 'C' means. It is also to identify phenomena that count as instances of C (Dupre, 1993). That is, the question, What does concept 'C' mean? is akin to a parallel form of the question, What counts as an instance of C? These two questions are not identical, but neither are they entirely distinct. For, to identify an instance of a C is a sensible (although perhaps insufficiently elaborated) response to the question, What does 'C' mean? And conversely, to explain what 'C' means is to explain what counts as an instance of C.

But questions about what something *is* are distinct from questions about what it *is like* (Dupré, 1993). More precisely, conceptual questions of identity are distinct from questions concerning empirical characteristics. If NN does not know what counts as a C, then NN will be unable to study what Cs are like. But not only are questions of identity and questions regarding empirical characteristics logically distinct, they are also ordered with respect to their priority. The identity of C must be established *before* it is possible to establish what C is correlated with, how it is caused, of what it is predictive, etcetera. This "before" should be understood as indicating methodological rather than chronological priority. Of course, researchers refine and redefine concepts in the course of conducting research. However, although empirical investigations regarding C may motivate conceptual innovations that alter the meaning of 'C', this process should not be misconstrued as "discovering what 'C' means."

The Logical/Empirical Distinction

Since Quine's (1951) attack on the philosopher's analytic-synthetic dichotomy, it has been widely supposed that that any attempt to distinguish between logical truths and empirical truths is unsustainable (Putnam, 2002). Dennett (2005), for example, argues that one:

> cannot study [grammar] without asking questions–and even if you only ask *yourself* the questions, you still have to see what you say. The

> conviction that this method of consulting one's (grammatical or other) intuitions is entirely distinct from empirical inquiry has a long pedigree…but it does not survive reflection. (pp. 7-8)

But there is no need to suppose that conceptual analyses are "entirely distinct from empirical enquiry," at least, not in the cartoonish manner that Dennett describes. For his contrast between "grammatical intuitions" and "empirical inquiry" is false. The illuminating contrast is that between grammatical rules and empirical facts (Baker & Hacker, 1984a, 1984b; Bennett & Hacker, 2003; Hacker, 2004, 2007). It is perfectly obvious that no empirical investigation could reveal the existence of a square-shaped circle, because there is no such thing as a square-shaped circle. But the sense in which there is no such thing as a square-shaped circle is different from the sense in which there is no such thing as, say, a unicorn. It is *conceivable* that a unicorn could exist, because it is possible to define 'unicorn' and thereby to identify *what* would exist if a unicorn existed. Accordingly, the statement *Unicorns do not exist* can be said to express an empirical truth. In contrast, it is inconceivable for a figure to be both circular and square-shaped at the same time, and so it impossible to identify what would exist if a square-shaped circle existed. More precisely, it is preferable to replace the concept of 'what would exist ' with that of 'what could exist' and clarify that a square-shaped circle *could not* exist. Because the truth of *There is no such thing as a square-shaped circle* could not be overturned by a discovery of a square-shaped circle (there is no such thing to be discovered), it can be said that *There is no such thing as a square-shaped circle* expresses a logical rather than an empirical truth.

The distinction between (logical) questions regarding how words work and (empirical) questions concerning what the world is like is essential, but that distinction is grossly caricatured in Dennett's (2005) vision of conceptual critics purporting to consult grammatical intuitions in a manner that is somehow "entirely distinct from empirical inquiry" (p. 8). In contrast, it is the application (and misapplication) of concepts that reveals their logical features. And so, with respect to the concepts of an empirical science, logical investigations involve examining *how* those concepts are applied to empirical investigations. Accordingly, my investigation of the 'information' in 'information-processing psychology' involves scrutinizing how information-processing psychologists (e.g., Miller 1953, 1956) employ the concept of ‘information’ in the course of their empirical inquiries. So, investigations regarding what the ‘information’ in ‘information-processing psychology’ means are logical (and not empirical) investigations. However, it is absurd to suppose that investigating the concepts of an empirical science is radically separate from the empirical investigations themselves. In picturesque terms, Dennett’s (2005) conceptual analyst is a straw man faced with an impossible task: investigating how words fit the world without also examining the world to which the investigated words are fit (or misfit). No such absurd undertaking is pursued here. It is not “intuitions…grammatical or otherwise” that constitute my objects of investigation (p. 8). Rather, it is certain particular claims of information-processing psychologists that are scrutinized. So it is easy to agree with Dennett (2005) that “see[ing] what [persons] say” is an essential aspect of the conceptual investigations below (p. 7). However, the fact that these investigations involve examining statements made by empirical scientists provides no reason to doubt their strictly logical (rather than empirical) character.

Clarity of Claims Versus Rules of Language

Hacker (1990) has forcefully argued that “if neurophysiologists, psychologists, artificial-intelligence scientists, or philosophers wish to change existing grammar, to introduce new ways of speaking, they may do so; but their new stipulations must be explained and conditions of application laid down” (p. 148). Hacker frames his critique in terms of ‘rules’. He criticizes “cross[ing a] new ‘technical’ use with the old one [for] this produces a *conflict of rules* and hence incoherence” (pp. 148-149, emphasis added). The reference to “rules” here is easily misunderstood. For example, Dennett (2005) has objected to this argument by observing that:

> …no philosopher has *ever* articulated ‘the rules’ for the use of any ordinary expression. To be sure, philosophers have elicited judgments of deviance by the hundreds, but noting that “we wouldn’t say thus-and-so” is not expressing a rule… Linguists have learned that something may sound a bit odd, smell a bit fishy, but still not violate any clear rule that anybody has been able to compose and defend. (p. 9)

But conceptual investigations are not criminal investigations, and "innocent until proven guilty" does not apply in the way that Dennett supposes. The burden of explaining what a claim *means* lies upon the claimant. Furthermore, no references to rule-violations need be involved.[2] However, it is easy to agree with Dennett (2005) that "it is hard, detailed work *showing* that the terminology used is being misused in ways that seriously mislead the theorists" (p. 15). The difficult detailing of uses and misuses of 'information' in information-processing psychology is precisely what this thesis is about.

[2] To be clear, I regard my arguments as consistent with and much indebted to those of Peter Hacker (e.g., 2004, 2007; Baker & Hacker, 1984a, 1984b; Bennett & Hacker, 2003).

Chapter 2.

Distinguishing Among Different Senses of ‘information’

It will be helpful to begin this investigation of the ‘information’ in ‘information-processing psychology’ by sharply distinguishing among (a) the everyday epistemic sense of ‘information’ that is internally related to the concept of ‘knowing;’ (b) Shannon’s (1948) statistical concept of ‘information’ (*H*, as in $H_i = \Sigma p_i \log_2(1/p_i)$, for which p_i is the probability of message/event *i* being selected from a set of alternatives); and (c) a mass of ambiguous, apparently technical uses of ‘information’ that are historically associated with the phrase information theory, in particular, Hartley (1928), Wiener (1948, 1950), and Weaver (1949). In this chapter, it will be demonstrated that the information in Miller’s (1951, 1953, 1956) seminal work on information-processing involved chronic confusion among the three senses (a)-(c) of ‘information’ identified above. But Miller is hardly alone. As Bar-Hillel (1955) presciently observed, “it is psychologically almost impossible not to make the shift from one sense of information…to the other...” (p. 284). In part, this appears to result from widespread confusion regarding where Shannon’s (1948) mathematics end and ‘information’-laden speculation begins. Norbert Wiener’s (1948, 1950) and Warren Weaver’s (1949) contributions to this confusion are described below. However, as is also demonstrated below, even Shannon’s (1948) seminal paper itself involves certain, subtly confusing habits of language. For example, in places, Shannon (1948) refers to “transmission of information” when he means *transmission of signals* (p. 399). Shannon’s (1948) minor misstatements notwithstanding, this chapter establishes that the widely held contention that “the challenge of extending the concepts of information theory” to the social and cognitive sciences “…is traceable to the writing of its founders” must be qualified considerably (Rapoport, 1956, p. 304). It is essential to distinguish between certain autobiographical and/or historical connections between information theory and information-processing psychology and the widespread misperception (e.g., Broadbent, 1958; Lachman et al., 1979; Miller, 1951, 1953, 1956; Weaver, 1949; Wiener, 1950) of substantive continuity between Shannon's statistical concept of 'information,' measured in units of *bits*, and the ideas of organisms, minds, and/or brains as *information processors* (e.g., Miller, 1951; Wiener, 1950).

Shannon (1948) versus Weaver (1949)

It is essential to distinguish Weaver's (1949) speculative commentary on Shannon's (1948) paper from Shannon's work itself, especially because Shannon and Weaver are often cited together, as if they co-authored the 1949 book that, in fact, includes two individually-authored (and substantively incompatible) chapters (e.g., Adams, 1991; Seow, 2005; see also Ritchie, 1986). Whereas Shannon (1948) created information theory to address engineering questions regarding the relative efficiency and reliability of various electronic communication systems, Weaver (1949) suggested that Shannon's (1948) information-theoretic concepts could also be applied to *reverse-engineering* problems in language research, in which social scientists attempt to determine how human beings communicate and/or use information in solving problems and/or navigating their environments (e.g., Miller, 1951, 1953, 1956). However, Weaver was hardly alone in painting with such broad strokes. Norbert Wiener (1950) refers to "ideas shared by Drs. Claude Shannon, Warren Weaver, and [himself]…" (p. 16). But it is clear that Wiener's concept of 'information' as "a name for the content of what is exchanged with the outer world as we adjust to it and make our adjustment felt upon it" (p. 17) is entirely distinct from Shannon's (1948) statistical concept of 'information' although, as demonstrated below, Wiener, like so many others, is confused about this. The subsequent investigations of Miller's seminal works in information-processing psychology in Chapter 3 demonstrate that it is Weaver's and Wiener's respective versions of Shannon's work that information-processing psychologists unknowingly appropriated in applying information-theoretic concepts to psychological phenomena. The "assumption that [Weaver's speculations] are somehow supported by Shannon's mathematics" (Ritchie, 1986, p. 279) has been rightly called a "most serious source of confusion" (p. 279). Furthermore:

> Confusion has also arisen from confounding the precise technical and statistical usage of words such as "uncertainty," "communication," and "information," with the more common everyday usage of these words…Relationship that are demonstrated to hold for the precisely defined concepts simply cannot be extended to every situation in which the word is used in everyday language. (pp. 279-280)

As Devlin (2001) puts it, "Shannon's theory does not deal with 'information' as that word is generally understood. Instead, it deals with data…" (p. 21). Luce (2003) observes that Devlin's (2001) description makes Shannon's theory "sound akin to what we normally think to be the role of statistics, which is correct" (p. 183). Luce's (2003) claim is correct although somewhat imprecise in that Shannon's (1948) theory is not *akin* to a statistical theory, it *is* a statistical theory. Shannon's (1948) "precise technical and statistical" (Ritchie, 1986, p. 279) usages of 'uncertainty,' 'communication,' and 'information' (as well as a few others) are described below.

On Definability and Internal Relations

As Hartley (1928) warned, "as commonly used, information is a very elastic term" (p. 1). Accordingly, I briefly survey the terrain across which this term typically is stretched, beginning with the *Oxford English Dictionary* entries for 'information,' and adding a supplementary analysis that will be helpful once I turn my attention to information-processing psychology. However, I first anticipate an objection to this approach. Sloman (2011), for example, is suspicious of this method. He observes that linguistic definitions of 'information' inevitably presuppose related concepts:

> …'information'…cannot be explicitly defined without circularity…Attempts to define 'information" by writing down an explicit definition of the form 'information is...' all presuppose some concept that is closely related ('meaning', 'content', 'reference', 'description', etc.). 'Information is meaning', 'information is semantic content', 'information is what something is about' are all inadequate in this sense. (pp. 9-10)

The analysis of the everyday sense of 'information' below will also be vulnerable to this criticism, as it presupposes the concept of 'knowledge.' So it is important to address such concerns, which are easily dissolved. It is far from clear why Sloman (2011) should characterize the internal relations between 'information' and other concepts that are closely related in meaning as establishing that 'information' suffers a "kind of indefinability" (p. 10). The center and radius of a circle, respectively, are mutually co-defining; this is different from their being undefined or indefinable (Bickhard, 2003). Furthermore, the definition developed below is descriptive rather than stipulative. So, rather than *recommending* that 'information' in its everyday sense be understood by reference to the concept of knowledge, I am *observing* that this conceptual connection already obtains.

On the assumption that Lachman and colleagues (1979) are lamentably correct in observing that "while few cognitive psychologists measure 'information' in [Shannon's (1948)] technical sense nowadays, many still use the term…[at times] nearly as a synonym for knowledge" (pp. 74-75) the student of information-processing psychology will do well to be as clear as possible about the relation between 'information' and 'knowledge' respectively (as well as about Shannon's technical sense of 'information'). But only the most prosaic knowledge about 'knowledge' is required for this analysis. For example, if NN has learned *that p,* then NN knows that p, and if NN is surprised to discover that p, then NN did not know that p prior to discovering that p. If that p should turn out to be false, then NN will have mistakenly thought that p, or falsely believed that p, and so on. In short, no grand epistemic theory is prerequisite for this investigation regarding the 'information' in information-processing psychology.

The Everyday Epistemic Sense of 'Information'

The *Oxford English Dictionary* distinguishes between two main senses of 'information.' First, it defines 'information' ("Information," 2010) as "facts provided or learned about something or someone" consistent with Bennett and Hacker's (2003) definition of what they refer to as the "semantic sense" of information, "a body of true propositions" (p. 141). However, the OED's definition foregrounds factivity rather than truth, a subtle but important difference. Truth is a potential property of propositions, and true propositions express facts. But a suitably equipped agent can surmise one fact or another from observing a situation, for example, that the cat is on the mat. That the cat is on the mat can be learned by observing that this is so. So the hypothetical student of cat locations needn't encounter the proposition "the cat is on the mat," and she might never make a statement to this effect. Nevertheless, this unspoken observation would still appear to count as her *gaining information* and/or *becoming informed*. So, very many things that are commonly regarded as sources of information are not propositions.

On Sources of Information versus Information

Of course, a *source of information* is different from an *instance of information*. However, in practice, this subtlety is routinely passed over. "Would you like some flight departure information?," is more familiar than the cumbersome, but perhaps more precise, paraphrase "Would you like a document from which information about flight departure times might be derived?" This appears to reflect a preference for parsimony over precision. And in some cases, the distinction between sources and instances of information, respectively, does not even appear to apply. For example, "a body of true propositions" appears to count *both* as an instance of information as well as a source of information. It counts as a source in that a literate reader can acquire information through exposure to such propositions. On the other hand, specifying *which* information was acquired through such proposition-exposure would appear to require reciting or referencing of those true propositions (or paraphrases thereof). In this sense, true propositions appear to count both as sources of information as well as instances of information.

There are also cases in which information-sources are more easily distinguished from information in-and-of-itself. The OED also defines 'information' as: "what is conveyed or represented by a particular arrangement or sequence of things" ("Information," 2010). This sense of 'information' covers, for example, the geographical information that can be found on a map; the facial expressions and/or gestures of certain animal species that may provide information about their emotional states, intentions, and/or beliefs;[3] the information regarding the existence and/or location of a fire that can be gleaned from observing a plume of smoke; and/or that observing footprints may provide information about the organism that left them, and so on. For these cases, sources and information proper are more readily distinguishable. In order for X to convey or represent Y, X and Y must be logically distinct; it is not the sequence or arrangement of things that counts as information but, rather, *what* is conveyed or represented by such a sequence and/or arrangement. However, as with the mass noun sense, such subtleties are routinely disregarded in the course of casual speech. To insist, *I asked for geographical information and all I got was this map* is probably to joke.

The Double-life Definition of Everyday Epistemic Information

These dictionary definitions of 'information' cast a wide net that appears to cover both (a) statements, objects, and/or events from which knowledge might be derived and (b) knowledge potentially derivable from statements, objects, and/or events. That is, information can designate both a *source* of knowledge that p (e.g., to learn from information) as well as *what* an agent who knows that p knows (e.g., that Vancouver, BC is north of Seattle, WA). Thus, the everyday sense of 'information' is an epistemic sense of 'information': an essential criterion of a statement, object, or event, counting as information, or a source thereof, in that it affords the possibility for an appropriately-equipped agent to derive knowledge from it.[4] Conversely, to be informed and/or possess information is to know something that can potentially be conveyed and/or represented by a statement, object, or event. The everyday sense of 'information' enjoys a double life, designating both sources of knowledge (e.g., [true] propositions, accurate maps, plumes

[3] Of course, for clever animals, such behaviors may not be genuinely informative, as in cases of faking, mimicking, acting, etcetera. Sincerity, rather than truth, per se, is what distinguishes the informativeness of authentic behavior from the disinformation involved in deceptive dissembling.

[4] The potentiality/actuality distinction is crucial. That an agent may, in fact, derive a false belief rather than knowledge from X does not disqualify X as an instance or source of 'information'. This argument is distinct from Sloman's (2011), who argues that information-processing psychology requires some notion of "false information" (see Chapter 4)

of fire-locating smoke, etc.) as well as expressions or manifestations of knowledge (e.g., that the cat is on the mat, that the mat is east of the window, that the mat is on fire, etc.).

Information, Knowledge, and Potentiality

Because this everyday epistemic sense of 'information' is internally related to that of *knowledge*, it is not possible to determine what is and is not information in this everyday sense in absolute terms. The concept of 'knowledge' presupposes that of a 'knower' and possibilities for knowledge-derivation are subject-relative. So what counts as information for one agent might not for another. Consider, for example, receiving driving directions spoken in an unfamiliar, foreign tongue. If one is unable to determine what is being said, one can hardly be said to have acquired information from such an exchange. This gets a little messy, however, as it is certainly possible to recognize that these foreign-language, would-be directions are *potentially* informative, for a linguistically prepared listener that is. Thus, it is possible to identify a source of 'information' by which one is unable to be informed. However, this only underscores the applicability of the definition developed above, as recognizing that the potential of X to inform is constitutive of identifying *X as* source and/or instance of 'information'.[5]

Epistemic versus Semantic

From this point forward, I will refer to the common everyday sense of 'information' described above as its everyday *epistemic* sense. For the purposes of this investigation, this is an improvement over the practice of referring to the everyday sense of 'information,' as its *semantic* sense (e.g., Bar-Hillel, 1955; Bennett & Hacker, 2003;

[5] Would a *false* proposition count as information according to this analysis? This would appear to depend on whether *mis*information is considered to constitute a type of information or, conversely, it is held that misinformation is the opposite of information. There does not appear to be a clear right and wrong here, although the familiarity of describing statements that were once held to be true but which are now considered false as 'outdated information,' suggests that it is preferable to relax the truth requirement on information. Furthermore, if it were false *that p,* and the falsity of *p* were known to agent A, then it would be possible for A to derive the knowledge that *not p*, from *that p*.

Cherry, 1957; Miller, 1953; Shannon, 1948).[6] To be clear, this is a heuristic distinction, not an argument that the everyday sense of 'information' is an epistemic rather than a semantic concept (whatever such a claim might mean).

The convention of characterizing the everyday sense of 'information' as its *semantic sense* is understandable in the light of the history of the statistical sense of 'information.' For if an employment of 'information' is qualified at all, it is typically for the purposes of distinguishing the everyday sense of 'information' from Shannon's (1948) statistical sense of the term. Here, what is generally construed as the most crucial difference is that the everyday sense of 'information' involves a semantic component that the statistical sense lacks. This appears to be consistent with Shannon's statement that the "semantic aspects of communication are irrelevant to the engineering problem" (p. 379) for which his statistical concept of 'information' was developed. So the everyday sense of 'information' has become known as the semantic sense in part through a negation: if Shannon's statistical concept of 'information' is a *non*-semantic one, then the everyday sense of 'information' must be a semantic sense. The practice of contrasting the semantic and information-theoretic concepts of 'information' has become common, and can be found in the works of authors who disagree sharply on a variety of methodological and substantive issues. For example, as quoted above, Bennett and Hacker (2003) contrast the "engineering" sense of 'information' with "the semantic sense" (p. 141). In a very different philosophical pursuit, Bar-Hillel and Carnap (1953) attempted to develop a theory that they "expected…will serve as a better approximation for some future explanation of a psychological concept of information than the concept dealt with in [Shannon's] Communication Theory" (p. 148) and tilted it a *Theory of Semantic Information*.

[6] To be clear, this emphasis on the epistemic aspects of 'information' is not entirely novel. For example, see Dretske's (1981) ambitious, information theory-inspired attempt to develop a non-normative, purely causal account of knowledge as information-caused-belief (p. *x,* see also pp. 85-95). One problem with his argument is that Dretske's definition of 'information' involves the notion of truth, and thus, he does not succeed in reaching his stated goal of developing a purely physical, causal account of knowledge that is free from such normative notions as 'truth.'

The clearest expression of this thesis comes from Dretske (1981) who argues that "to speak of information as *out there*, independent of its actual or potential use by some interpreter…is [often seen as] bad metaphysics" (p. x). But this:

> way of thinking…rests on a confusion...of information with *meaning*. Once this confusion is cleared up, one is free to think about information (though not meaning) as an objective commodity…whose generation, transmission, and reception do not require or in any way presuppose interpretive processes. (p. x, see also pp. 85-95)

In other words, if you sever the connections between everyday 'information' and 'meaning,' you are left with a technical concept of 'information.' The arguments that follow involve a sustained critique of this misleading analysis.

For now, however, the task is merely to observe that the everyday sense of 'information' is an epistemic sense and that, in certain cases, it seems that 'semantic' is used in places for which 'epistemic' would be a preferable substitution. For example, it has been argued that it is "almost inevitable that we make semantic, intentional, [*sic*] projections onto an informational treatment of neural signals" (Barandiaran & Moreno, 2008, p. 683). It seems doubtful that these authors are arguing, absurdly, that an informational treatment of neural signals involves treating neural impulses as *having a meaning* in the same sense that we regard statements such as 'snow is white' as having a meaning. Rather, they appear to be arguing, more reasonably, that it is almost inevitable that we treat neural signals as causal bases for knowledgeable actions of the organism to which those neural signals belong. For example, if neural impulses traveling from through the optic nerve to the visual cortex are viewed as causal correlates of an organism seeing what lies before it, and, in that sense, *knowing* what lies before it, then it seems clearer to say that the theorist who posits neural information is thereby making an epistemic projection via an informational treatment of neural signals.[7]

Shannon's (1948) Statistical Concept of 'information'

We now turn attention to Shannon's (1948) statistical concept of 'information'. As compared to the everyday epistemic sense of 'information', the statistical sense of 'information', Shannon's H, can be defined with relative precision. However, in order to understand the significance of $H = -\sum p_i \log_2 p_i$ it is necessary to survey the network of concepts in which that equation figures. Accordingly, my description of Shannon's (1948) information theory will involve clarifying the restricted information-theoretic senses of a number of common everyday terms (communication, information, message). That is, although it is Shannon's statistical sense of 'information' that constitutes my primary object of investigation, I must also review certain other, related, information-

[7] Bennett and Hacker (2003), following Wittgenstein (1953, 1972), have argued very persuasively that the common, casual habit of treating 'seeing an *X*' and 'knowing that one sees an *X*' as interchangeable is philosophically misleading. In short, their argument is that it only makes sense to speak of *knowing X* in cases when it is sensible to speak of *not knowing X*. But it would be absurd, and/or perhaps psychiatrically alarming, for a person to say "I see an *X* but I do not know that I see an *X*." It would also be absurd (and comically redundant) for a person to say "I see an *X* and I know that I see an *X*," although, this phrase might be used to indicate that, for example, 'I see something before me that I'm certain is an *X* rather than a *Y*.' So it appears that, with certain exceptions, such as cases involving uncertainty regarding the identity of *what* is seen, it typically makes no sense to speak of seeing *X* but *not knowing* that one sees *X* and thus, by implication, that it also (typically) makes no sense to speak of *knowing that one sees an X.* I take Bennett and Hacker's argument to be extremely helpful and entirely correct. However, it is what is conventional, and not what is correct, that I am concerned with here. As subtly incoherent as the practice may be, it is nonetheless true that statements such as 'I know that this is a table that I'm seeing' are a regular feature of philosophical and cognitive-scientific debates.

theoretic concepts in order to grasp its significance.[8] In other words, just as the everyday epistemic sense of 'information' is internally related to the concept of 'knowledge,' Shannon's (1948) statistical, information-theoretic sense of 'information' is internally related to the restricted, information-theoretic senses of communication and message (among others).

Information Theory: The Big Picture

It will be helpful to begin by referring to what is often called (e.g., Ritchie, 1986) Shannon's model of a communication system (see Figure 2.1) and to qualify the role that the concept of 'model' plays in this description. Shannon's model is a graphic definition of what counts as a communication system in information-theoretic terms. In other words, only to what corresponds, more or less exactly, to Shannon's model can information-theoretic concepts such as bits of information, and/or channel capacity be predicated. Shannon (1948) is extremely clear about this:

> By a communication system we will mean a system [that]…consists essentially of five parts…(1) An *information source* which produces a message or sequence of messages to be communicated to the receiving terminal…(2) A *transmitter* which operates on the message in some way to produce a signal suitable for transmission over the channel…(3) The *channel*…the medium used to transmit the signal from transmitter to receiver…(4) The *receiver* [which] ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal…[and] (5) The *destination* [which] is the person (or thing) for whom the message is intended. (p. 380; see also Figure 2.1)

[8] A brief caveat: the review of Shannon's (1948) work that follows will be selectively guided by the ultimate purpose of gaining clarity regarding the 'information' in information—processing psychology. Accordingly, I will review only those aspects of Shannon's theory that are necessary to understand how such pioneering psychologists as Miller (1951, 1953, 1956) distorted Shannon's work in applying his concepts to psychological phenomena. Space will not permit a more exhaustive analysis of Shannon's work (see Ritchie (1986) for an excellent more general analysis). In other words, I will strive to be as clear as possible in saying as little as is needed to correct widespread, foundational confusions regarding the relationship between information theory and information-processing psychology.

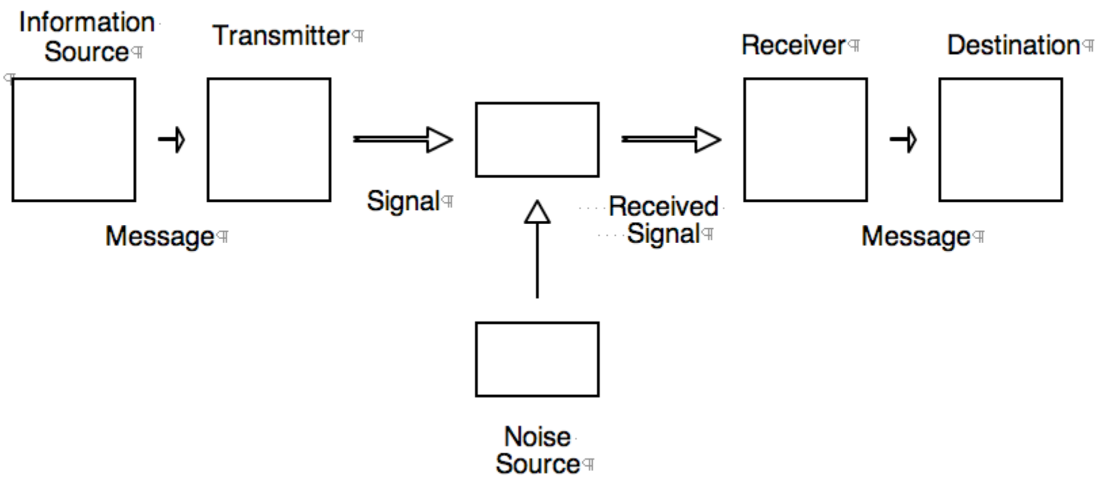

Figure 2.1. A reproduction of Shannon's (1948) model of a communication system

Note. Adapted from Shannon, 1948.

From this point forward, I will sometimes, especially when I want to highlight Shannon's legacy, I will refer to the system described above as a *Shannonian communication system.*

The Information-theoretic Sense of Communication

As Ritchie (1986) observes, Shannon's "theorems constitute not a *communication theory* as students of human communication understand the term, but a general theory of *signal transmission*…[and] both the motivating problem set and intended applications of [his] theory were drawn from the electronics industry" (p. 280). Accordingly, Shannon's (1948) claim that "the fundamental problem of communication is that of reproducing either exactly or approximately a message selected at another point," should be understood as stipulating what counts as communication in information-theoretic terms (p. 379). It is *not* an attempt to specify an essential aspect of communication more broadly construed. So, *pace* Weaver (1949, see below) Shannon is *not* claiming that all cases of communication involve the reproduction of previously-selected messages. Rather, Shannon stipulates that the *only* aspects of communication that are relevant to his information theory are those that involve the reproduction of selected messages.

Selection Not Semantics

Whereas *selection* of messages is essential to the information-theoretic sense of communication, Shannon (1948) clarifies that although "frequently the [selected] messages have *meaning*…" (p. 379) such "semantic aspects of communication are irrelevant to the engineering problem" (p. 379). Thus communication in the information-theoretic sense consists in the *reproduction* of messages; what those messages may or may not mean and/or express is irrelevant. This establishes that Shannon's information-theoretic sense of communication is vastly different from the sense of communication involved in, say, describing As and Bs respective failures to understand one another's statements as A and B *failing to communicate*, or as suffering a *communication breakdown*. Accordingly, the reproduction of, say, HI HOW ARE YOU? and ?OUY ERA WOH IH, respectively, are of equal status with respect to the achievement of communication, in its information-theoretic sense.[9] That the former is straightforwardly interpretable whereas the latter is not is irrelevant (although the latter is quite easy to decrypt into the former, which can then be interpreted).

Two Crucial Observations

First, it is essential that each "actual message is one *selected from a set* of possible messages" (p. 379). So the (misleadingly labeled!) *information source* and *message destination* (see Figure 2.1) must be designed such that each has access to an identical set of possible messages. As Rapoport (1956) observed, "information theory…is fundamentally a theory of *selection*. Something is selected from a well-defined set. To examine the selective process, it is essential to be able to examine the set" (p.

9 In this sentence and subsequently, I will use ALL CAPITAL LETTERS to indicate a message in the information-theoretic sense and italics to indicate the content of such messages. Accordingly, HI HOW ARE YOU? constitutes a message that expresses the interrogative greeting: *Hi, how are you?*

306).[10] The communication engineer determines, that is *decides upon*, the scope of possible messages in accordance with practical goals. For example, a communication system designed to transmit signals pertaining to the results of a radar-based missile detection activity would likely involve a set of possible messages that is different from, say, a system designed to transmit telegraphic (signals corresponding to) natural language messages.

Only Signals Are Transmitted

In order for there to *be* a "fundamental problem of …reproducing either exactly or approximately a message selected at another point," it must be the case that, with respect to information-theoretic communication, what is transmitted are not messages themselves but, rather, signals that correspond to messages, the receipt of which affords the possibility (or, raises the problem) of reconstructing the message previously selected for transmission (Shannon, 1948, p. 379). There can be no message-reconstruction (or *decoding*) without prior, signal-producing *encoding*. If it were messages themselves that were transmitted, then there would be no need to reconstruct anything. In Shannon's terms, the "*transmitter*…operates on the message in some way to produce a signal suitable for transmission over the channel" and it does so by *encoding* messages into signal patterns (p. 380).

Information, in the Statistical Sense, Cannot Be Transmitted

The importance of recalling that it is signals, and *only* signals, that are transmitted by Shannonian communication systems cannot be emphasized enough, especially in light of the fact that Shannon (1948) sometimes misleadingly refers to the *transmission of information*, inviting confusion between the (information-theoretic) statistical and everyday senses of 'information,' respectively. For example, in Shannon's visual depiction (see Figure 2.1) of his model of a communication system, the *message source* is erroneously labelled as an *information source*. And he further confuses the

[10] Rapaport's (1956) paper is an admirably clear analysis of *The Promise and Pitfalls of 'information' Theory* (which phrase constitutes its title) with respect to psychological applications.

reader by referring to a system "in which both the message and the signal are a sequence of discrete symbols" (p. 384) blurring the crucial distinction between messages and signal sequences, that is presupposed by Shannon's core concept of encoding.[11]

Unfortunately, history is of little help here. Shannon (1948) cites Hartley's (1928) paper, titled *Transmission of Information,* in which Hartley introduces a more primitive version of Shannon's *H* statistic (Hartley's *H* can be computed only for sets of equiprobable messages). Now, obviously, a great deal of signal transmission might be accurately described as information transmission, in the everyday epistemic sense of 'information'. For example, the transmission of the signals associated with a message reading BE HOME AT 5 transmits (or, at least, potentially transmits) the (everyday epistemic) information that the message-sending party will be home at 5 o'clock. However, and although in many contexts this degree of rigor would be overkill, the reader who labours to achieve the "psychologically almost impossible" (Bar-Hillel, 1955, p. 284) and speak authoritatively rather than "*somewhat* impressionistically, about…information being processed" needs the sharpest tools he can find (Dennett, 2005, p. 15). Accordingly, it is essential to retain that it is neither messages nor information that are transmitted during communication, in the information-theoretic sense. Rather, it is signal sequences that correspond to messages that are transmitted.

The Information-theoretic Sense of 'Message'

It is difficult to develop a general definition of the information-theoretic sense of message. On the one hand, this concept of message is restricted but, on the other hand, it is very inclusive. This paradoxical difficulty is a result of the fact that such messages:

[11] The error involved is one of imprecision. Shannon (1948) means to distinguish systems in which signals can take only discrete values (e.g., 0 or 1) from those in which signals are continuously variable, for example, as in telephony:

> We may roughly classify communication systems into three main categories: discrete, continuous and mixed. By a discrete system we will mean one in which both the message and the signal are a sequence of discrete symbols. A typical case is telegraphy where the message is a sequence of letters and the signal a sequence of dots, dashes and spaces. A continuous system is one in which the message and signal are both treated as continuous functions, e.g., radio or television. (p. 382)

> may be of various types: (a) A sequence of letters as in a telegraph of teletype system; (b) A single function of time $f(t)$ as in radio or telephony; (c) A function of time and other variables as in black and white television—here the message may be thought of as a function $f(x;y;\ t)$ of two space coordinates and time, the light intensity at point $(x;y)$ and time t on a pickup tube plate; (d) Two or more functions of time, say $f(t)$, $g(t)$, $h(t)$—this is the case in "three-dimensional" sound transmission or if the system is intended to service several individual channels in multiplex; (e) Several functions of several variables—in color television the message consists of three functions $f(x;y;\ t)$, $g(x;y;\ t)$, $h(x;y;\ t)$ defined in a three-dimensional continuum—we may also think of these three functions as components of a vector field defined in the region—similarly, several black and white television sources would produce "messages" consisting of a number of functions of three variables; (f) Various combinations also occur, for example in television with an associated audio channel. (Shannon, 1948, p. 380)

The reader will notice that as "the semantic aspects of communication are irrelevant to the engineering problem" the information-theoretic sense of message excludes very many aspects of what is often meant by message in non-information-theoretic contexts (Shannon, 1948, p. 379). It is probably oversimplified to refer to an everyday (as opposed to information-theoretic) sense of message *in general* as, for example, there is great diversity among, say, voicemail messages left by one person for another, error messages automatically generated by computer programs, and/or the messages that organisms perceive in one another's behavior (e.g., a smile can send a message of approval). However, one common feature of these everyday uses of message is that it makes sense to refer to expressing the same message a different way. For example, a voicemail recording saying, "I'll be home at 6:30," expresses the same message as a text message reading HOME HALF PAST SIX; a computer program might use any arbitrary code to indicate *paper jam*, and a message of approval can be sent through utterances, facial expression, a thumbs up, etcetera. *These* senses of 'message' involve what is frequently expressed metaphorically as the *underlying* meaning or significance of a particular communicative gesture. But it is essential to recall that, with respect to information-theoretic analyses, a message is just that which may be selected from a set of alternatives. The distinction being drawn is that, with respect to the everyday epistemic sense of message, 8TH AVENUE IS WEST OF 7TH and 7TH AVENUE IS EAST OF 8TH might be said to express the *same message*, because 8th Avenue being

West of 7^{Th} is the same as 7^{th} Avenue being East of 8^{th}. On the other hand, 8TH AVENUE IS WEST OF 7TH and 7TH AVENUE IS EAST OF 8TH are entirely distinct messages, in the information-theoretic sense of message although certain of their respective constituent selections (or, values) are identical: _TH AVENUE IS _ _ST OF _TH. In Chapter 3 I will discuss how confusion between the everyday and information-theoretic senses of message plagues certain of Miller's (1951, 1953, 1956) seminal works.

Shannon's (1948) Statistical Sense of *Signal*

Shannon (1948) says very little about signals in his paper, only defining the term implicitly in describing one of the essential components of his model, a transmitter "which operates on the message in some way to produce *a signal suitable for transmission* over the channel" (p. 380, emphasis added). Examples of communication channels include "a pair of wires, a coaxial cable, a band of radio frequencies, a beam of light, etc." (p. 380). Accordingly, a signal in the information-theoretic sense, is the direct object of transmission events whereby the transmitting end of a Shannonian communication channel affects the receiving end (see Figure 2.1). When thinking of signal transmission, it will be helpful to imagine "something objective…—the electrical pulses surging down a copper wire, for example" and *suppress* connotations of the "*message* that these pulses" might be said to "carry" (Dretske, 1981, p. ix), metaphorically speaking. That most readers presumably know what, say, a red-light traffic signal *means* is a serious liability here.

Shannon's (1948) Statistical Sense of *Encoding*

Encoding is the process that mediates between messages and transmittable signal sequences. Encoding procedures can take various forms, but common to them all is that they involve *transformation rules*:

> In telephony [the encoding] operation consists merely of changing sound pressure into a proportional electrical current. In telegraphy we have an encoding operation which produces a sequence of dots, dashes and spaces on the channel corresponding to the message. In a multiplex

> PCM system the different speech functions must be sampled, compressed, quantized and encoded, and finally interleaved properly to construct the signal. Vocoder systems, television and frequency modulation are other examples of complex operations applied to the message to obtain the signal. (Shannon, 1948, p. 380)

Shannon's information-theoretic sense of encoding is, in certain important aspects, consistent with non-information-theoretic applications of the concept of a code. For example, Bennett and Hacker (2003) define "a code [as] a method of encrypting linguistic expression (or any other form of representation) according to conventional rules" (p. 167). Being easily accessible to readers, Shannon (1948) refers frequently to examples of linguistic encoding, as in the following passage, which also explains that the relative efficiency of different encoding procedures is of special interest to the information-theoretic analyses of engineers.

> In telegraphy, for example, the messages to be transmitted consist of sequences of letters. These sequences, however, are not completely random. In general, they form sentences and have the statistical structure of, say, English. The letter E occurs more frequently than Q, the sequence TH more frequently than XP, etc. The existence of this structure allows one to make a saving in time (or channel capacity) by properly *encoding the message sequences into signal sequences*. This is already done to a limited extent in telegraphy by using the shortest channel symbol, a dot, for the most common English letter E; while the infrequent letters, Q, X, Z are represented by longer sequences of dots and dashes. (p. 383, emphasis added)

So, in seeking to encode messages into transmittable signal form, engineers are interested in anticipating the statistical properties of the signal transmission patterns. And although not all messages consist in symbol sequences, for those that do, it is typically possible to detect statistical regularities in the signal transmission patterns associated with the communication of such messages. In general, the engineer's goal is to design communication systems with ever increasing channel capacities, that is, systems that can transmit signals with increasing speed and accuracy. Shannon (1948) defines 'channel capacity' (C) as "$C = \text{Max}[H(x) - H_y(x)]$." But before we can understand what this means, we must introduce the star of the show, H (p. 401).

Shannon's (1948) Statistical Sense of 'information'

Bearing in mind that Shannon's "information theory…is fundamentally a theory of *selection*" (Rapoport, 1956, p. 306) and that much "confusion has…arisen from confounding the precise technical and statistical usage of…'information'…with [its] more common, everyday usage" (Ritchie, 1986, p. 279), we are now prepared to investigate Shannon's statistical sense of 'information'. Shannon (1948) stipulates:

> if the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as *a measure of the information produced* when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley [1928] the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message.
>
> (Shannon, 1948, p. 379, emphasis added)

How does Shannon (1948) consider the influence of statistics? He ingeniously observes that:

> we can think of a discrete source as generating the message, symbol by symbol. It will choose successive symbols according to certain probabilities depending, in general, on preceding choices as well as the particular symbols in question. A physical system, or a mathematical model of a system which produces such a sequence of symbols governed by a set of probabilities, is known as a stochastic process. We may consider a discrete source, therefore, to be represented by a stochastic process. Conversely, any stochastic process which produces a discrete sequence of symbols chosen from a finite set may be considered a discrete source. This will include such cases as….[n]atural written languages such as English, German, Chinese [and/or] [c]ontinuous information sources that have been rendered discrete by some quantizing process. For example, the quantized speech from a PCM transmitter, or a quantized television signal. (p. 385)

It is important to note that Shannon's (1948) claim is that such message sources can *be represented by* a stochastic process. This is different from claiming that say, the behavior of a contrite lover who frets over how to word his text-messaged apology *is engaged in a stochastic process*. Rather, Shannon's claim is merely that, for example, because, in English, readers encounter the "sequence TH more frequently than XP…"

(p. 383) this hypothetical text-messenger is more likely to send I WANT TO BE WITH YOU than I WANT TO BE WIXP YOU. So it will be helpful to observe that message sources *under an information-theoretic description* are represented as stochastic processes. Anticipating TH rather than XP in this example is important because the "main point at issue" for the engineer "is the effect of statistical knowledge about the source in reducing the required capacity of the channel, by the use of proper encoding" of messages (pp. 383-384). Most readers of this text will likely be familiar with such automated, statistical anticipations of symbol sequences in the form of word-complete, or T9 functions on portable data devices.

Signal Transmission Versus Message Composition, or on Selecting Versus Being Selected

It is essential to distinguish between signal-transmission *events* and message-composition *actions*, with actions having agentive and semantic connotations of which events is free. The distinction being drawn is that between a communication system operator *composing* a message, that is *selecting* as in *deciding* which signals to transmit, and the event of a particular message *being selected* from a set of alternatives. *The distinction between selecting and being selected is crucial.* Shannon's (1948) work on serially-dependent signal sequences was not intended to aid in the prediction of which signal a message-constructing agent might *choose* to transmit. Rather, the question that concerned Shannon was the statistical structure of which signals *are selected* by particular message sources. His interest in the statistical structure of English-language symbols considered as communication signals was motivated by a desire to construct maximally efficient electronic communications systems, not to reverse-engineer the linguistic capacities of potential English-language message composers.

The English Language as a *Message Source*

For the purposes of illustrating how sources generating natural language messages can be profitably construed as stochastic processes, Shannon (1948) describes a series of approximations to the English language:

> To give a visual idea of how this series of processes approaches a language, typical sequences in the approximations to English have been constructed and are given below. In all cases we have assumed a 27-symbol "alphabet," the 26 letters and a space.
>
> 1. Zero-order approximation (symbols independent and equiprobable). FOML RXKHRJFFJUJ ZLPWCFWKCYJ FJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.
> 2. First-order approximation (symbols independent but with frequencies of English text). OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.
> 3. Second-order approximation (digram structure as in English). ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.
> 4. Third-order approximation (trigram structure as in English). IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.
> 5. First-order word approximation. Rather than continue with tetragram, : : : , *n*-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.
>
>   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TOOF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.
> 6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.
>
>   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.
>
> (Shannon, 1948, p. 386)

Shannon (1948) observes that "The resemblance to ordinary English text increases quite noticeably at each of the above steps…The particular sequence of 10 words 'attack on an English writer that the character of this' is not at all unreasonable" (p. 386). He concludes: "a sufficiently complex stochastic process *will give a satisfactory representation* of a discrete source" (p. 386, emphasis added). This sufficient complexity is attributable to Shannon's recognition of *serial dependencies* among successive symbols. For example, T is more likely to be followed by H than by X in English–language messages. Shannon observes that such serially dependent relations among successive events:

> are known mathematically as discrete Markoff processes…The general case can be described as follows: There exist a finite number of possible "states" of a system; *S*1;*S*2,…,*Sn*. In addition there is a set of transition probabilities; *pi*(*j*) the probability that if the system is in state *Si* it will next go to state *Sj*. To make this Markoff process into a [message] source we need only assume that a letter is produced for each transition from one state to another. The states will correspond to the "residue of influence" from preceding letters. (p. 385)[12]

Having demonstrated the utility of "represent[ing] a discrete [message] source as a Markoff process" (p. 389), Shannon then asks, "Can we define a quantity which will measure, *in some sense*, how much information is "produced" by such a process, or better, at what rate information is produced?" (p. 389, emphasis added).[13]

Finally, we are prepared to bring this statistical sense of 'information' into view.

> Suppose we have a set of possible events whose probabilities of occurrence are p_1; $p_2 \ldots p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome? (Shannon, 1948, p. 389)

The careful reader will notice that, in the preceding passages Shannon has enclosed both *produced* and *choice* in quotation marks. This should be understood as an expression of the restricted sense in which these terms figure in Shannon's work. Shannon characterizes his *H* in terms of 'information' but also in terms of uncertainty and choice: "Quantities of the form $H = -\sum p_i \log_2 p_i$…play a central role in information theory as measures of 'information', *choice and uncertainty*" (p. 390, emphasis added).

12 This quote has been altered for expository clarity. The original refers to an information source, exemplifying Shannon's subtly confusing habit of identifying *message sources* with *information sources*.

13 See Footnote 12.

Rapoport (1956) describes the manner in which 'information,' 'choice,' and 'uncertainty' fit together here:

> If one selects a message from a source of *n* messages, each selection is a 'configuration' characterized by a certain probability. Then *H* is the uncertainty (per message) associated with the source. The receipt of the message transmitted without error 'destroys' the uncertainty of the recipient, with regard to which message will be chosen. Therefore *H* measures also the amount of 'information' per message. (p. 304)

Although Rapoport's gloss is criticized below (Chapter 3), it has the virtue of orienting the reader to the *restricted* sense of uncertainty to which information-theoretic analyses pertain. What unit of 'information', in this statistical sense, is chosen and measured? Shannon (1948) explains that the "choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits,* a word suggested by J. W. Tukey" (p. 379).[14] So, the amount of 'information', in Shannon's (1948) statistical sense associated with a particular message, *M*, is a joint function of (a) the frequency with which that message is expected to be transmitted and (b) the number of alternative messages eliminated by *M*'s selection for transmission.[15] For illustrative purposes, consider the comparison of two familiar activities that may be heuristically construed as information-theoretic message sources: a coin toss and the roll of a die, respectively. The probability of a heads-side-up outcome in a fair coin toss is $p = 0.5$, whereas the probability of

[14] In the Chapter 3, I will explain how different senses of *bit* have contributed to confusion regarding information.

[15] From here on, I will refer to Shannon's (1948) technical concept of 'information' as the *statistical* sense of 'information' rather than the *information-theoretic* sense for two reasons. First, the information-theoretic sense of 'information' suggests circularity. Second, for the purposes of clarifying that it is not bits of 'information', in Shannon's sense of the term, that are transmitted by communication channels (but rather, signals that are transmitted), it is helpful to recognize that particular values of Shannon's *H*, say 2.4 bits, are *values of a statistic*, not measurements of a property that can be described in non-information theoretical terms. It is easy to understand that it makes no sense to speak of transmitting a value of a statistic, say $r = .07$, although it is perfectly clear to refer to the transmission of a signal which corresponds to a message, R .07, expressing the information, in the everyday epistemic sense, that a particular Pearson *r* =. 07.

rolling a, say, six is approximately $p \approx 0.17$. Accordingly, the amount of 'information', in the statistical sense, associated with a fair coin toss is exactly 1 bit.

$H_i = -\log_2 p_i$; $p_i = 0.5$

$- \log_2 (0.5) = \log_2 (1/.05) = \log_2 (2) = 1.$

The less-likely event of rolling a six, for example, is associated with the greater amount of approximately 2.6 *bits*:

$H_i = -\log_2 p_i$; $p \approx 0.17$

$-\log_2 (0.17) = \log_2 (1/0.17) \approx \log_2 (5.9) \approx 2.6.$

Shannon (1948) clarifies that "justification of [his] definitions…will reside in their implications" (p. 390). However, it is probably clearer to say that he considered the justification for the restricted and interrelated senses of 'information', 'choice,' and 'uncertainty' to reside in their *applications* to engineering problems.[16] Shannon (1948) contains many clear examples of such applications. Rather than recapitulate these examples here, it will be more helpful for our purposes to identify certain potentially confusing aspects of Shannon's paper.

On Shannon's Contributions to Misunderstanding of His Own Work

I have emphasized that it is *signals* rather than *messages, symbols*, or *information*, in the statistical sense (i.e., *H*), that are transmitted by Shannonian communication systems. However, throughout his paper, Shannon repeatedly writes in a

[16] It is clear from certain of Shannon's (1956) later statements that he did not anticipate many of the *implications* of his choice of terminology here. For example:

> Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. In part, this has been due to connections with such fashionable fields as computing machines, cybernetics, and automation; and in part, to the novelty of its subject matter. As a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments. (p. 3)

manner that suggests otherwise, even though it is perfectly clear that he was not substantively confused about this. For example, Shannon (1948) states that:

> If [a] channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with *certainty* by any operation on the received signal *E*. There are, however, ways of *transmitting the information* which are optimal in combating noise.
>
> (p. 399, emphasis added)

It is clear that Shannon means to say that there are ways of transmitting the signal that are optimal in increasing the probability that the received signal corresponds to the transmitted one. Yet, in the very next sentence, Shannon refers to a hypothetical system "transmitting at a rate of 1000 *symbols* per second" (p. 399, emphasis added). But it is *signals* which, in certain cases, are *correlated with symbols* via the transformation rules of a particular encoding procedure, that are transmitted. Along these misleading lines, Shannon also labels the message source in his graphic depiction, or definition, of a communication system (see Figure 2.1) as an information source. There are many examples of such imprecision throughout Shannon's (1948) paper, for example:

> The question we now consider is how one can measure the capacity of such a channel to *transmit information*. In the teletype case where all symbols are of the same duration, and any sequence of the 32 symbols is allowed the answer is easy. *Each symbol represents five bits of 'information'*. If the system transmits *n* symbols per second it is natural to say that the channel has a capacity of 5*n* bits per second.
>
> (p. 382, emphasis added)

To be clear, to identify imprecision on Shannon's part is not to justify misunderstandings of his theory (e.g., Miller, 1951, 1953, 1956; Weaver, 1949). As Shannon (1956) protested, his information theory was quite clearly "aimed in a very specific direction, a direction that is not necessarily relevant to such fields as psychology" (p. 1). However, it is also true that certain of Shannon's core concepts do invite misunderstandings and encourage confusion between *signal transmission*, to which his paper does apply, and *information transmission*, to which it does not. Shannon (1948) is silent on matters concerning what transmitted signals are signals *for*; whatever information, in the everyday epistemic sense, can be derived from a message that corresponds to a transmitted signal sequence is “irrelevant to the engineering problem” (p. 379) that concerns him. That Shannon (1948) himself and, before him, Hartley (1928) refer so frequently to the *transmission of information* makes this point extremely difficult to bear in mind. It seems clear that such misleading habits of language contribute to it being “psychologically almost impossible” (Bar-Hillel, 1955, p. 284) to maintain a sharp distinction between the statistical and everyday epistemic senses of ‘information’, respectively.

Most confusingly, it appears that Shannon’s (1948) core concept of ‘channel capacity’ is impossible to describe in words without at least *appearing* to suggest that communication channels transmit information, in the statistical sense of ‘information’. That is, *C* indicates “the maximum possible rate of transmission” of reliably identifiable signals (p. 401).[17] However, *C* is expressed in terms of amounts of *bits* per unit of time, which makes it awfully difficult to suppress the idea or *bits* of information traveling through the channel.[18] Nonetheless, this is obviously not the case, as a value of a statistic is not something that can travel through “a pair of wires, a coaxial cable, a band of radio frequencies, [or] a beam of light” (Shannon, 1948, p. 380).

Finally, and perhaps most egregiously, on the basis of a series of approximations to English language (quoted above), Shannon (1948) concludes that “when we write English half of what we write is determined by the structure of the language and half is chosen freely” (p. 393). It is difficult to understand this comment as anything other than a psychological conclusion drawn from an information-theoretic premise. So it would be overstated to claim that psychologists’ interest in applying Shannon’s ideas to psychological phenomena were unprecedented and entirely inconsistent with Shannon’s

17 $C = \mathrm{Max}[H(x) - H_y(x)]$” (p. 401) for which $H(x)$ is the average amount of ‘information’, in the statistical sense, per message associated with a source (x) and “$H_y(x)$ is…[roughly]…the amount of additional information that must be supplied per second at the receiving point to correct the received message” (Shannon, 1948, p. 400).

18 In his invaluable effort to correct common misperceptions regarding Shannon’s work, even a writer as lucid as Ritchie (1986) recapitulates certain of Shannon’s terminological errors: “…it is impossible to exceed a *symbol* [emphasis added] transmission rate of *C/H* [emphasis added]” (p. 282).

work.[19] Furthermore, it is clear that Shannon is interested in the possible implications of technological progress for psychological concepts. For example, in a paper concerning programming a computer to play chess, Shannon (1950) observes that "chess is generally considered to require 'thinking' for skilful play" (p. 266), and that the existence of a chess playing-computer "will force us either to admit the possibility of a mechanized thinking or to further restrict our concept of 'thinking'" (p. 266). However, that Shannon (1950) expresses interest in artificial intelligence does not entail that the concept of 'information' that pertains to his 1948 mathematical theory of communication has anything to do with knowledge or intelligence; it very clearly does not.

On Applying Shannon's (1948) Concepts to Systems that Do Not Correspond to His Model

Ritchie (1986) has argued with admirable vigor and clarity that there is no:

> point in trying to use Shannon's theory without his model; his theory was developed specifically for the system described by the model…There is nothing in Shannon's paper to justify use of [his] formula[e] under any other circumstances. (pp. 280-281; see also Figure 2.1)

Ritchie's conservativism is reasonable and might be viewed as an admirable corrective to widespread confusion among information-processing psychologists (and beyond) regarding where Shannon's (1948) information theory ends and 'information'-based speculation begins. However, for the purposes of investigating the relationship between Shannon's information theory and information-processing psychology, it will be helpful to accept, at least temporarily, that correspondence to Shannon's model comes in degrees. For the history of information-processing psychology includes both coherent (e.g., Garner, 1953; Pollack, 1953a, 1953b) and incoherent (e.g., Miller, 1953, 1956) applications of Shannon's (1948) concepts to psychological phenomena.

[19] In rightfully emphasizing the contrast between Shannon (1948) and Weaver (1949), Ritchie (1986) may underestimate Shannon's own contribution to the abuse of his information-theoretic concepts.

Shannon (1948) also foregrounds a non-engineering application of his statistical concept:

> The form of [his]*H* [statistic] will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell *i* of its phase space. *H* is then, for example, the *H* in Boltzmann's famous *H* theorem. (p. 390)

So, Shannon's statistical concept of 'information', (or his information statistic, *H)* certainly has valid applications that do not involve 'communication systems,' as defined in his (1948) paper (see Figure 2.1). However, there are also limits on the range of cases to which information-theoretic concepts can be applied. These limits are foregrounded by the confusion that ensues when they are transgressed, as by Weaver (1949).

The ABCs of Weaver (1949)

Shannon's (1948) paper appears in the *Bell System Technical Journal.* It was later reprinted in book form, along with a commentary written by Warren Weaver (1949). Shannon's and Weaver's respective chapters are entirely distinct; there is no co-authored text. Weaver describes his chapter as "an interpretation of mathematical papers by Dr. Claude E. Shannon" (p. 1). However, Weaver's is a *mis*interpretation, as revealed by his misdescription of the relation between Shannon's (1948) statistical sense of 'information' and the concept of 'entropy' as used in statistical mechanics: "entropy is related to 'missing information'…inasmuch as it is related to the number of alternatives which remain possible to a physical system after all *the macroscopically observable information concerning it has been recorded*" (p. 1, emphasis added). But the kind of information that may be macroscopically observable and which may be recorded is everyday epistemic information; there is obviously no such thing as, say, 2.6 macroscopically observable *bits*. Of course, it is possible to record the information that, for example, in construing a fair die toss as an information-theoretic message source, rolling six, for example, is associated with approximately 2.6 bits. But what is recorded in recording that 'rolling a six involves approximately 2.6 bits' is everyday epistemic information *about* the amount of information, in the statistical sense, associated with a die roll of six. Weaver's (1949) is a fundamentally unsound introduction to Shannon (1948). However, before further investigating Weaver's confusion, it will be helpful to set his work in context.

Who Was Warren Weaver?

A physicist working for the Sloan Foundation, Warren Weaver was a broad-minded thinker who pioneered interest in the use of digital computers to translate among natural languages, writing a memorandum on the topic in 1949 that quotes passages from a letter penned to Norbert Wiener in 1947. Weaver's *Translation* memorandum was subsequently reproduced in an edited volume (Weaver, 1955). His interest in machine translation primed him to seek generality or commonalities across different spoken languages:

> All languages—at least all the ones under consideration here—were invented and developed by men; and all men, whether Bantu or Greek, Islandic or Peruvian, have essentially the same equipment to bring to bear on this problem. They have vocal organs capable of producing about the same set of sounds (with minor exceptions, such as the glottal click of the African native). Their brains are of the same general order of potential complexity. The elementary demands for language must have emerged in closely similar ways in different places and perhaps at different times. One would expect wide superficial differences; but it seems very reasonable to expect that certain basic, and probably very non obvious, aspects be common to all the developments.
>
> (Weaver, 1955, pp. 16-17)

One commonality across languages is that meaning and/or significance of spoken or written words are influenced by the context in which they occur. But how could a machine tell the difference between the 'banks' in "bank of the Mississippi River" and "Bank of Canada," respectively? Weaver (1955) observes that, in the case of printed text, the influence of context can be thought of in terms of serial dependencies among words:

> Let us think of a way in which the problem of multiple meaning can, in principle at least, be solved. If one examines the words in a book, one at

> a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words. "Fast" may mean "rapid"; or it may mean “motionless"; and there is no way of telling which. But if one lengthens the slit in the opaque mask, until one can see not only the central word in question, but also say *N* words on either side, then if *N* is large enough one can unambiguously decide the meaning of the central word. (p. 24)

Although Weaver (1955) does not mention Shannon’s (1948) work on Markoff processes, it appears all but certain that his interest serially dependent *semantic* relationships influenced his reading of Shannon. It is certainly true that Weaver (1949) mistakenly interprets Shannon’s work as relevant to semantic questions of meaning. However, I can only speculate that Weaver’s prior interest in the contextual dependency of word meaning may have influenced his misreading of Shannon’s (1948) work on serially-dependent frequency relationships among messages, in the restricted, information-theoretic sense of ‘message.[20]

Weaver (1949) on Shannon (1948): Speculation Not Summary

Weaver (1949) very clearly reads Shannon’s (1948) work in light of his own, non-engineering interests. He begins his treatment of Shannon’s work with unfounded speculations regarding the significance of Shannon’s theorems for communication in general: “The word *communication* will be used here in a very broad sense to include all of *the procedures by which one mind may affect another*” (Weaver, 1949, p. 1, emphasis added). Thus, Weaver’s introduction does the *exact opposite* of Shannon’s (1948). Shannon explicates that, under an information-theoretic analysis, the “semantic aspects [of communication] are irrelevant” (p. 379). Shannon is concerned with one *mechanism* affecting another, in particular, a receiver reproducing a message selected for encoding and transmission by a transmitter. Weaver (1949), in astonishing contrast, casts his net far and wide, observing that communication “involves not only written and oral speech,

[20] Weaver’s (1955) stated interest in the possibility of machine translation is admirably humanitarian: “A most serious problem, for UNESCO [United Nations Education, Scientific and Cultural Organization] and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples” (p. 20).

but also music, the pictorial arts, the theatre, the ballet, and in fact all human behavior" (p. 1). Most puzzlingly, he construes Shannon's, restricted, information-theoretic sense of 'communication' as the most inclusive:

> In some connections it may be desirable to use a still broader definition of communication, namely, one which would include the procedures by means of which one mechanism (say automatic equipment to track an airplane and to compute its probable future positions) affects another mechanism (say a guided missile chasing this airplane).
>
> (Weaver, 1949, p. 1)[21]

Weaver posits three interrelated levels of communication problems, the first of which, despite some verbal imprecision, is easily recognizable as related to Shannon's (1948) information-theoretic concerns:

> Relative to the broad subject of communication, there seem to be problems at three levels. Thus it seems reasonable to ask, serially:
> **LEVEL A.** How accurately can the symbols of communication be transmitted? (The technical problem.)
> **LEVEL B.** How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)
> **LEVEL C.** How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)
>
> (Weaver, 1949, p. 2)

Although the above passages involve the nearly ubiquitous error of misidentifying what it is actually transmitted, Weaver (1949) is elsewhere commendably clear about this: "The *transmitter* changes [the] *message* into the *signal* which is actually sent over the c*ommunication channel* from the transmitter to the *receiver*" (p. 3). He also correctly recognizes that "The mathematical theory…[of] Claude Shannon at the Bell Telephone Laboratories, admittedly applies in the first instance only to problem A," but goes on to argue that Shannon's work has:

> a deep significance….[which] comes from the fact that levels B and C, above, can make, use only of those signal accuracies which turn out to be possible when analyzed at Level A. Thus any limitations discovered in

[21] The implication is that minds are mechanisms, which Weaver uncritically assumes.

> the theory at Level A necessarily apply to levels B and C. But a larger part of the significance comes from the fact that the analysis at Level A discloses that this level overlaps the other levels more than one could possibly naively suspect. Thus the theory of Level A is, at least to a significant degree, also a theory of levels B and C.
>
> (Weaver, 1949, p. 3)

Identifying Signals versus Understanding Messages

What sense can be made of Weaver's (1949) contention that "levels B and C…can make use only of those signal accuracies which turn out to be possible when analyzed at Level A" (p. 3)? First, it is helpful to observe that this statement only applies to certain types of messages, in the information-theoretic sense. That is, it only makes sense to ask how "precisely…the [communicated] symbols convey the desired meaning" with respect to messages that consist in symbols (p. 2). However, a paraphrase of Weaver's (1949) questions might also apply to messages that, for example, do not consist in symbols but rather are "thought of as a function $f(x;y;\ t)$ of two space coordinates and time, the light intensity at point $(x;y)$ and time t on a pickup tube plate…" (Shannon, 1948, p. 380) such as a television program. Consider, for example, a television program concerning the health risks associated with smoking tobacco. It is sensible to wonder how precisely such a program communicates its anti-smoking message, and to wonder if the receipt of this message has any effect on the smoking-behavior of its audience. And, of course, asking such questions presupposes that the signals associated with the reproduction of this hypothetical anti-smoking program were properly encoded, transmitted, and decoded. If, for example, what its would-be viewers actually viewed involved a glitchy, signal-scrambled, mostly inaudible mess, it would not be reasonable to wonder how precisely its anti-smoking message (in the everyday sense) was communicated (in the everyday sense) and/or whether or not viewership effected behavioral change.

So, there is at least one sensible way to interpret Weaver's (1949) contention that "levels B and C, above, can make, use only of those signal accuracies which turn out to be possible when analyzed at Level A" (p. 3). Should the reader therefore accept that "any limitations discovered in the theory at Level A necessarily apply to levels B and C"

(p. 3) and, that, therefore, a theory of Level A is “also a theory of Levels B and C” (p. 3)? Certainly not. For Weaver only establishes that Levels B and C investigations presuppose certain conditions at Level A. For example, as in our example above, entertaining the Level B question of whether or not a message was understood as intended presupposes that the message was received in an interpretable form. Furthermore, both levels B and C (concerning “How effectively…received meaning affect[s] conduct in the desired way,” p. 2) also presuppose that communication involves agents that are capable of understanding or not understanding a particular statement.

In light of Weaver’s (1949) recognition that “two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information [in that statistical sense] (p. 6),” his speculations regarding the equivalence of his Levels A, B, and C, are extremely puzzling. He correctly observes that the meaning of a particular message is logically independent of the amount of information, in the statistical sense, which is associated with the message. “But,” Weaver goes on to say, “this does not mean that the engineering aspects are necessarily irrelevant to the semantic aspects” (p. 6). This is apparently contradictory, as avenues of (ir)relevance typically run both ways. That is, if *X* is *un*related to *Y*, then it is not clear how *Y* can be related to *X.* Is Weaver merely suggesting that, as in our example of an anti-smoking television program above, it only makes sense to ask if the anti-smoking message was understood by viewers if those viewers are known to have viewed the program in the form it was intended to be viewed?

Decision versus Discovery

Clearly, Weaver (1949) has something more profound, but also much more vague, in mind. Weaver argues that “when one meets the concept of entropy in communication theory, he has a right to be rather excited—a right to suspect that one has hold of something that may turn out to be basic and important” (p. 7). In his rather excited state, Weaver’s remarks suffer from confusing a decision with a discovery:

> When we have…[a] source which is producing a message by successively selecting discrete symbols (letters, words, musical notes,

> spots of a certain size, etc.), the probability of choice of the various symbols at one stage of the process being dependent on the previous choices (i.e., a Markoff process), what about the information associated with this procedure? The quantity which uniquely meets the natural requirements that one sets up for "information" turns out to be exactly that which is known in thermodynamics as *entropy.* (p. 7)

However, this puts the informational cart before the Markoffian horse. For it is not the case that Shannon discovered that the "natural requirements" for information measurement were met by the thermodynamic concept of 'entropy.' Conversely, Shannon (1948) demonstrated that deciding to represent such message sources as Markoff processes was useful for engineering purposes because it afforded a mathematical analysis of communication, in his explicitly restricted, information-theoretic sense of the term 'communication.' Weaver (1949), however, anthropomorphizes Shannon's strictly mechanical, information-theoretic concepts, as can be seen in the following passage, which begins by misleadingly referring to a message source as an information source:

> the information source is free to choose only between several definite messages—*like a man picking out one of a set of standard birthday greeting telegrams*. A more natural and more important situation is that in which the information source makes a sequence of choices from some set of elementary symbols, the selected sequence then forming the message. Thus *a man may pick out one word after another, these individually selected words then adding up to form the message.* (p. 6, emphasis added)

However, Shannon (1948) did not discover a basic and important similarity between "a man picking out one of a set of standard birthday greeting telegrams" and a thermodynamic system (Weaver, 1949, p. 6). Rather, Shannon (1948) realized that a sequence of choices from some set of possible values could be profitably *represented as* a Markoff chain, because, for example, HAPPY is a more probable sequence than HAPPX. But Shannon's theorems pertain to the behavior of signal transmitting systems, *not* to the behavior of those who choose which sequences are transmitted.

A Theory of Signal Transmission versus a Theory of What Signals Are Signals for

From the diversity of what messages may be reproduced from transmitted signals, Weaver (1949) draws sweeping and illegitimate conclusions about the generality of Shannon's theorems:

> [Information theory] is a theory so general that one does not need to say what kinds of symbols are being considered—whether written letters or words, or musical notes, or spoken words, or symphonic music, or pictures. *The theory is deep enough so that the relationships it reveals indiscriminately apply to all these and to other forms of communication.* This means, of course, that the theory is sufficiently imaginatively motivated so that it is dealing with the real inner core of the communication problem—with those basic relationships which hold in general, no matter what special form the actual case may take. (p. 14, emphasis added)

It is true that in developing an information-theoretic analysis, one "does not need to say what kinds of symbols are being considered." But this is most definitely *not* because Shannon's theory *"indiscriminately appl[ies] to all…forms of communication*" (as cited in Weaver, 1949, p. 14). Weaver mistakenly regards the breadth of what signals *can be signals for* (e.g., "written letters or words, or musical notes, or spoken words, or symphonic music, or pictures…") as evidence that Shannon's (1948) theorem's reveal "basic relationships which hold [for communication] in general, no matter what special form the actual case may take" (Weaver, 1949, p. 14). But this is incorrect, because Shannon's (1948) theorems apply exclusively to certain *aspects* of certain *cases* of communication—those involving the reproduction of messages selected from a set of alternatives. In that respect, "a man picking out one of a set of standard birthday greeting telegrams" *can* be reasonably construed in information-theoretic terms (Weaver, 1949, p. 6). However, an information-theoretic description describes only certain *aspects* of that situation, such as the probability of a particular message being selected from the set of standard birthday greetings. Shannon's (1948) information theory is absolutely silent regarding the questions constituting Weaver's (1949) "**LEVEL B.** How precisely do the transmitted symbols convey the desired meaning? (The semantic problem) [and/or]

LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem)" (p. 2).

Weaver's Levels A and C Can Be Understood Mechanically: His Level B Cannot

Weaver's (1949) Level A (the technical problem, i.e., Shannon's [1948] problem of signal accuracies in reproducing previously selected messages) relates to his Level C (the effectiveness problem), but this relation does not pass through his Level B (the semantic problem) in the manner that he supposes. Weaver (1949) asks "How effectively does *the received meaning* affect conduct in the desired way?" (p. 2, emphasis added). But the effectiveness problem concerns how the received *signals* affect the behavior of that which receives them. So Weaver is incorrect in claiming that "the effectiveness problem is closely interrelated with the semantic problem" (p. 3). For, as Ritchie (1986) observes:

> In engineering, interpretation or understanding has no meaning apart from action, and effectiveness has no meaning apart from control…If we send a signal from the helm of a ship specifying "right ten degrees rudder," we would indeed evaluate the effectiveness of the communication system according to how near the rudder comes to stopping at precisely ten degrees right. We would scarcely think to speak of a rudder *understanding* a message, *choosing* whether to obey a message or not, *resenting, appreciating,* or *considering* a message…The idea that an individual might *understand* a message but choose to disregard it, or that a *message* might, as with the *Mona Lisa*, have enduring value precisely because it calls for an ambiguous or open-ended interpretation, has no place in an engineering model.
>
> (pp. 284-285)

So, in an engineering context, it is reasonable to conceive of the Level A technical and Level C effectiveness problems as interrelated. For example, an engineer might investigate both how reliably the command signal for "ten degrees right" (Ritchie, 1986, p. 284) is transmitted as well as how accurately the rudder-turning mechanism responds. However, the type of communication that occurs as a control lever transmits command signals to a ship rudder obviously does not involve the semantic, agentive, epistemic, and aesthetic aspects of non-information-theoretic communication identified in

Ritchie's counterexample. Rather, the relation between a control lever and the rudder it controls is strictly causal, and the distinctions among 'confused,' 'disobedient,' and 'dysfunctional' do not apply. So the relation between Levels A and C, respectively, does not pass through the semantic Level B, so to speak.

In failing to distinguish between signals (which are transmitted) and symbols (which are not), Weaver (1949) semanticizes the relation between Levels A and C: He describes his "**LEVEL B…**semantic problem" (p. 2) as one of determining "How precisely…the *transmitted symbols convey the desired meaning*?" (p. 2) and Level C, in turn, as asking "How effectively does *the received meaning* affect conduct in the desired way? (The effectiveness problem)" (p. 2). But the degree of accuracy with which a ship's rudder responds to a *received command signal* obviously does not depend upon the degree to which the rudder *understands* that signal (or the symbol(s) to which that signal may correspond). Rather, the only problem of communication that a ship's rudder can have is that of identifying *which* signal it has received. This is different from identifying what that signal might *mean* to some agent who knows what it is a signal *for*.

How does Weaver (1949) imagine that his Level B semantic problem mediates between Levels A and C? By substituting conduct for control, Weaver blurs the distinction between mechanical and interpersonal senses of 'commands,' as if a ship captain's command to a crew member to execute a 10-degree turn to the right is identical with a command signal sent from a steering mechanism to a rudder-turning mechanism. In confusingly collapsing the very distinctions that he draws between Levels A, B, and C, respectively, Weaver stipulates that the effectiveness problem should be understood very broadly.

> The problem of effectiveness involves aesthetic considerations in the case of the fine arts. In the case of speech, written or oral, it involves considerations which range all the way from the mere mechanics of style, through all the psychological and emotional aspects of propaganda theory, to those value judgments which are necessary to give useful meaning to the words "success" and "desired" in the opening sentence of this section on effectiveness. (p. 3)

In place of an argument describing how Level B semantic questions of interpretation relate to Level C effectiveness questions of control, Weaver construes Level C so broadly as to collapse the distinctions among conveying meaning, exerting control, and affecting conduct.

Untangling Weaver

In order to untangle Weaver (1949) and diagnose his distortion of Shannon, it is helpful to review the respective works of Hartley (1928) and Wiener (1948, 1950). Weaver (1949) begins his commentary on Shannon's 1948 paper by observing that:

> Dr. Shannon's work connects…directly with certain ideas developed some twenty years ago by…R. V. L. Hartley…and Dr. Shannon has himself emphasized that communication theory owes a great debt to Professor Norbert Wiener for much of its basic philosophy. (Weaver, 1949, p. 1)

Accordingly, I investigate Hartley's (1928) and Wiener's (1948, 1950) relevant works below. First it will be demonstrated that Wiener's (1950) work involves a more elaborate version of the same confusions exhibited by Weaver (1949). Then, Hartley's (1928) foundational contribution to the 'information'-related confusion will be investigated. In particular, it will be demonstrated that Hartley posited a chimerical link between the what would later be known as Shannon's (1948) information theory on the one hand and a class of psychological theories on the other (in particular, representational theories of mind).[22]

[22] As noted above, representational theories of mind are those that assume that such mental (particularly, epistemic) processes as perceiving, believing, thinking, inferring, and so on involve representations, which mediate between the epistemic subject and the world (e.g., Fodor, 1975, 2010; Hobbes, 1994 circa 1650). For devastating critiques of representational theories of mind see Baker and Hacker (1984a), Bennet and Hacker (2003), Kenny (1972), and Wittgenstein (1953).

‘Control Information’ versus Shannon’s Statistical Sense of ‘Information’

Miller (1953) describes Norbert Wiener’s (1948) *Cybernetics* as “more stimulating than intelligible” (p. 11) while describing Wiener’s lesser-known 1950 work, *The Human Use of Human Beings* as “highly readable prose” (Miller, 1953, p. 11). Given that certain of Miller’s (1951, 1953, 1956) publications are of particular interest in Chapter 3, Wiener’s (1950) later remarks on information in particular are investigated below.

In places, Wiener’s (1950) use of ‘information’ is either metaphorical or metaphysical: “Information is a name for *the content of what is exchanged* with the outer world as we adjust to it, and make our adjustment felt upon it” (p. 17, emphasis added). He alludes to a substance that is exchanged in the interaction between individuals and their physical environment, but which also flows through persons physiologically, in through the eyes and ears and out to the muscles:

> Man is immersed in a world which he perceives through his sense organs. *Information that he receives is co-ordinated through his brain and nervous system until, after the proper process of storage, collation, and selection, it emerges through effector organs, generally his muscles.* (p. 17, emphasis added)

However, information is also said to be exchanged between communicating parties: “When I communicate with another person, I impart a message to him, and when he communicates back with me he returns a related *message which contains information* primarily accessible to him and not to me” (p. 16, emphasis added). So, in the communicative exchange of Wienerian information, a kind of radical privacy obtains such that that information contained in a transmitted message is primarily accessible to the sender and, somehow, only secondarily accessible to the receiver. In this formulation, information appears to be roughly synonymous with meaning, as it is common to observe a difference between the intended meaning and the received meaning of a particular statement. And it is very difficult to understand how the kind of information that is contained in and exchanged through messages might also flow “through effector organs” (Wiener, 1950, p. 17).

Wiener (1950) wrongly suggests that his confusing concept of ‘information’ is consistent with Shannon’s (1948) statistical sense of the term. He conflates quantitative, information-theoretic analyses with critical, qualitative ones: “The more probable the message, the less information it gives. Clichés, for example, are less illuminating than great poems” (Weaver, 1950, p. 21). But Shannon’s *H* does not measure degrees of illumination and cannot be used to distinguish illuminating from hackneyed poetry. However, one defining feature of clichés is that they are common. And, if we view written English language as a message source, more probable messages are associated with less information, in Shannon’s (1948) statistical sense, by definition. But, upon reflection, it is perfectly clear that not every uncommon message would be “illuminating,” for example, FFFFFFFFFF.

Communication as *Control*

Wiener (1950) intentionally and instrumentally collapses the distinction between ‘commands’ in the sense of ‘imperative utterances’ (e.g., “Turn ten degrees right”) on the one hand and, on the other, ‘command signals’ such as the signals transmitted by a steering mechanism to a mechanical rudder (Ritchie, 1986). He compares *interpersonal* with *mechanical* communication, describing a coercive vision of communication as control: “When I *control* the actions of another person, I *communicate* a message to him” (Wiener, 1950, p. 16). This communication-as-control perspective applies quite naturally to the engineering sense of ‘communication’ that occurs in Ritchie’s (1986) above-cited example of a steering mechanism communicating with the rudder of a boat. In Weaver’s (1949) terms, Ritchie’s rudder example would be said to involve Level’s A (the technical problem, Did the rudder receive the same signal that the steering mechanism transmitted?) and C (the effectiveness problem, Did the steering mechanism steer as directed by the transmitted signal?). No question of understanding applies here, although it is possible for signals to be distorted during transmission and/or for messages to be improperly encoded/decoded. But mechanical failures are not misunderstandings. There is a difference between misidentifying *which signal* has been transmitted, on the one hand, and misinterpreting what a transmitted signal is a *signal for*, on the other. However, in Wiener’s (1950) communication-as-control analysis, the distinction between semantic agents and mechanical components, between causing N to *V* and demanding of N that he *V,* is blurred: “If my control is to be effective, I must take cognizance of any messages from him which may indicate that the order is *understood* and has been *obeyed”* (p. 16, emphasis added).

Human-machine Interaction

It is essential to understand that Wiener (1950) “classed communication and control together” (p. 16) for very particular sociological and social-engineering purposes. He was particularly interested in human-machine interaction in the dawning of the computer age. Wiener (1950) appears to have correctly anticipated the extent to which such technology would change society and/or culture at large:

> Let us go on now to picture a more completely automatic age. Let us consider what for example the automobile factory of the future will be like; and in particular the assembly line, which is the part of the factory that employs the most labor. (p. 154)

His argument is (needlessly) reductive for claiming that “Society can *only be* understood through a study of messages and the communication facilities which belong to it” (Wiener, 1950, p. 16, emphasis added). Nevertheless, Wiener correctly predicts that:

> in the future development of these messages and communication facilities, messages between man and machines, between machines and man, and between machine and machine, are destined to play an ever-increasing part. (p. 16)

Accordingly, Wiener (1950) is expressly interested in analyzing systems from a particular point of view, which he called *cybernetics*, named after the “Greek word *kubernëtës*…[for] ‘steersman,’ the same Greek word from which we eventually derive our word ‘governor.’” (p. 14). Wiener is interested in communications that consist in *cause-and-effect* relationships, for example:

> The automatic photo-electric door opener is known to every person who has passed through the Pennsylvania Station in New York, and is used in many other buildings as well. When a message consisting of the interception of a beam of light is sent to the apparatus, this message actuates the door, and opens it so that the passenger may go through. (p. 23)

It is helpful to bear the examples of the automatic photo-electric door opener and automobile factory assembly line in mind when confronted with certain of Wiener's (1950) claims. For example, he writes, "When I give an order to a machine, the situation is not essentially different from that which arises when I give an order to a person" (p.16). This might be read as a bizarre denial of human agency. However, it can also be read charitably as describing a cybernetic point of view, for example, it might be useful to think cybernetically and conceive of an automobile factory as involving the exchange of causally-efficacious messages among persons and machines, respectively, because the behavior of human workers on an assembly line is predictable, and therefore *idealizable* in casual terms in a way that other instances of communication, say, conversation among automobile factory workers on their lunch breaks, is not.

'Control Information' versus Everyday 'Epistemic Information'

Wiener (1950) suspects that cybernetic analyses will shed light not only on human-machine interaction, but, much more broadly, on the functioning of biological organisms in general:

> …the physical functioning of the living individual and the operation of some of the newer communication machines are precisely parallel in their analogous attempts to control entropy through feed-back. Both of them have sensory receptors as one stage in their cycle of operation: that is, in both of them there exists a special apparatus for *collecting information from the outer world* at low energy levels, and for making it available in the operation of the individual or of the machine. In both cases these external messages are not taken neat, but through the internal transforming powers of the apparatus, whether it be alive or dead. The *information is then turned into a new form* available for the further stages of performance. (pp. 26-27, emphasis added)

This is perhaps the earliest expression of the inchoate idea of organisms as information-processors. However, Wiener (1950) clarifies that *his* concept of 'information', as applied to behavioral analyses of organisms, must be understood as distinctly *non-volitional* and *non-representational:*

> If I pick up my cigar, *I do not will* to move any specific muscles…What I do is to turn into action a certain feedback mechanism; namely, a reflex

> in which the amount by which I have yet failed to pick up the cigar is turned into a new and increased order to the lagging muscles, whichever they may be…Similarly, when I drive a car, *I do not follow out a series of commands dependent simply on a mental image of the road and the task I am doing.* If I find the car swerving too much to the right, that causes me to pull it to the left. This depends on the actual performance of the car, and not simply on the road…. (p. 6, emphasis added)

In this passage, Wiener (1950) is at pains to establish that information transfer between organism and environment is not mediated by any mental representation ("...I do not follow out a series of commands dependent simply on a mental image of the road…") (p. 6). In Weaver's (1949) terms, we might say that Wiener is at pains to clarify that, under a cybernetic description, communication, the exchange of messages that contain information, involves Weaver's levels A and C, but not B. However, as quoted above, Wiener (1950) also claims that if "control is to be effective, [one] must take cognizance of any messages from [others] which may indicate that [one's] order is *understood* and has been *obeyed"* (p 16, emphasis added).

It has been demonstrated that Shannon's statistical sense of 'information' is *not* identical with Wiener's metaphysical sense of 'information' as "a name for the content of what is exchanged with the outer world as we adjust to it and make our adjustment felt up on it" (as cited in Wiener, 1950, p. 17). Nor is Shannon's *H* statistic synonymous with Wiener's more narrow concept of control information. It is further demonstrated below that Hartley (1928) is a primary source of confusion regarding the statistical and everyday epistemic senses of 'information'. By conceiving of his more-primitive-than-Shannon's *H* statistic as a measure of 'information', Hartley conflated signal identification with semantic understanding. However, I will work backwards to Hartley (1928) and begin with Weaver's (1949) subsequent confusion in this regard.

Weaver's (1949) Representational Theory of Mind

Whereas Shannon (1948) emphasizes the "irrelevance" of the "semantic aspects of communications" (p. 379), Weaver (1949) confuses the epistemic and/or semantic sense of ‘uncertainty’ (for example, Mr X not *understanding* what Mr. Y *means* by saying such-and- such) with Shannon’s (1948) *information-theoretic* sense of ‘uncertainty:’

> If Mr. X is suspected not to understand what Mr. Y says, then it is theoretically not possible, by having Mr. Y do nothing but talk further with Mr. X, completely to clarify this situation in any finite time. If Mr. Y says “Do you now understand me?” and Mr. X says “Certainly, I do,” this is not necessarily a certification that understanding has been achieved….this basic difficulty is, at least in the restricted field of speech communication, reduced to a tolerable size (but never completely eliminated) by “explanations” which (a) are presumably never more than approximations to the ideas being explained, but which (b) are understandable since they are phrased in language which has previously been made reasonably clear by operational means. (Weaver, 1949, p. 2)

But there is a difference between asking *Which message M did Mr. X select?* and asking *What did Mr. X mean by selecting message M?* And this is the very distinction that Weaver draws in distinguishing his Level A from his Level B. Furthermore, Weaver uses Mr. X and Mr. Y to draw a false analogy between the impossibility of noiseless signal transmission and conceptual (or philosophical) problems relating to meaning and mind. For not only are Misters X and Y prohibited from explaining themselves *to one another*, they are each also unable to *explain themselves,* as explanations are assumed to be *approximations to* or *of ideas*. Weaver’s (implicitly) representational theory of mind assumes a radical privacy of epistemic phenomena (in this case, understanding), which prohibits, for example, a genuine *exchange of ideas*.

In Weaver’s picture of Mr. X and Mr. Y, signals are to messages as statements are to ideas. And, for example, even though what Mr. X means by ‘red’ (i.e., his *idea* of red) might be different from what Mr. Y means by ‘red’ (i.e., *his* idea of red), so long as both Mr. X and Mr. Y agree that, for example, STOP signs are red, mutual good faith efforts at co-understanding shrink their problem of other minds to a tolerable size. In Chapter 3, it will be demonstrated that Miller’s (1951, 1953, 1956) influential work in information-processing psychology involved mistakenly supposing that Shannon’s (1948) theory of signal transmission could be used to explain (aspects of) *information* transmission in the sense that, say text-messaging driving directions transmits information to the recipient *who understands them*. But Miller's (1951, 1953, 1956) misapprehension in this regard is anticipated by Weaver (1949).

Hartley's (1928) Representational Theory of Mind

The most glaring 'information'-related confounding of theories of signal transmission and theories of mind is found in Hartley's (1928) paper, *Transmission of Information*. This paper introduced a less-flexible ancestor of Shannon's (1948) widely-known statistical concept of 'information' described above. Hartley's (1928) *H* pertained only to selections from sets of *equiprobable* alternatives: $H = log\ s^n$ for which *n* indicates the number of selections from a set of *s* equiprobable symbols.[23] But why did Hartley conceive of the quantity $H = log_2\ s^n$ as a 'measure of information'? As, Bar-Hillel (1955) observes, "it is hardly good English to talk about 'measure of signal sequences' or 'amount of signal'" (p. 285). Bar-Hillel helpfully explicates confusing consequences of this decision:

> we see over and over again that, in spite of the official disavowal of the interpretation of "information" as "what is conveyed by a signal sequence", "amount of 'information'", officially meant to be a measure of the rarity of kinds of transmission of signal sequences, acquires also, and sometimes predominantly, the connotation of a measure (of the rarity or improbability) of the kinds of facts (events, states) designated by these signal sequences. (p. 285)

However, Bar-Hillel (1955) also suggests that "this confusion is [not] only a result of an unfortunate terminology. To a certain degree, at least, it seems that the confusion was rather the cause of the misleading terminology. This can be seen in Hartley's paper

[23] Unlike Shannon's (1948) *H*, Hartley's (1928) *H* is *not* Boltzman's *H* (circa 1870, see Shannon, 1948). The relation between Boltzman's concept of 'entropy' and Shannon's concept of 'information' is that "entropy is related to 'missing information,' inasmuch as it is related to the number of alternatives which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded" (Weaver, 1949, p.1). Shannon's (1948) *H* is a descendent of both Hartley's (1928) and Boltzman's *H*s (see Shannon, 1948); like Hartley's, Shannon's *H* pertains to situations involving message reproduction; and, like Boltzman's, Shannon's *H* concerns the prediction of system states (i.e., the state of transmitting one signal rather than another).

itself" (p. 285, emphasis added). Bar-Hillel does not specify just what confusion might have caused Hartley to choose such misleading terminology. Without assuming that Bar-Hillel would necessarily sanction the diagnoses below, I will follow his lead and elaborate on what confusions may have led Hartley (1928) to conceive of $H = \log_2 s^n$ as a 'measure of information.'

Hartley's (1928) Psychologizing

It appears that Hartley's (1928) choice of words was influenced by his assuming a representational theory of mind (RTM):

> Let us consider what factors are involved in communication; whether conducted by wire, direct speech, writing, or any other method. In the first place, there must be a group of physical symbols, such as words, dots and dashes or the like, which by general agreement convey certain meanings to the parties communicating. *In any given communication, the sender mentally selects a particular symbol* and by some bodily motion, as of his vocal mechanism, *causes the attention of the receiver to be directed to that particular symbol*. By successive selections a sequence of symbols is brought to the listener's attention. At each selection there are eliminated all the other symbols which might have been chosen. As the selections proceed more and more possible symbol sequences are eliminated, and we say that the information becomes more and more precise. For example, in the sentence, "Apples are red," the first word eliminates other kinds of fruit and all other objects in general. The second directs attention to some property or condition of apples, and the third eliminates other possible colors… Inasmuch as the precision of the information chosen depends upon what other symbol sequences might have been chosen it would seem reasonable to hope to find in the number of these sequences the desired quantitative measure of communication. (p. 536)

Hartley (1928) uncritically assumes that *all* cases of communication involve selections from a set, with "mental symbols" playing the part of messages and spoken and/or written words playing the part of the transmitted signals (p 536). However, this particular *telementation model of communication* (Harris, 1981) confuses 'selecting from a set' with 'determining what to say.' Determining what to say involves a consideration of what one *means* to say. But, of course, Hartley's statistical, sense of 'information' is not related to what any mental or mechanically selected symbol *means*. Rather, 'amount of information,' in Hartley's statistical sense, is a joint function of a number of possible selections, *s*, and the number of actual selections, *n.* And what counts as a selection in this sense is causing some communication device to be in a particular state at a particular time, for example, the state of transmitting the signal that corresponds to the symbol, S.

But how does Hartley (1928) get from 'selection' to 'information'? On the assumption that generating a sentence such as *Apples are red* involves selecting mental symbols, he argues that "the precision of the information chosen" in deciding to say, for example, *Apples are red* "depends upon what other symbol sequences might have been chosen" (p. 536). But this involves an entirely novel concept of 'precision of information', for values of Hartley's *H* have nothing to do with the relative accuracy or specificity of particular statements such as *Apples are red*. As Bar-Hillel (1955) observes "the phrase, 'to transmit information,' used by Hartley is highly ambiguous, but in this context, the term 'information' has certainly nothing to do with what we might call the *semantic content* of the signals" (p. 283). The kind of information that can be more or less precise (that is, everyday epistemic information) is just not the kind of information that is measured by $H = log\ s^n$. For example, no quantitative calculations are involved in arguing that the statement *Apples are red* is quite imprecise in that many apples are green and/or yellow in color. And this observation regarding the imprecision of *Apples are red* is in no way expressed in the following calculations of Hartley's *H* value associated with APPLES ARE RED, for which we must assume that each of A…Z (plus space) is an equiprobable selection.

$H = log\ s^n$ for which *n* indicates the number of selections from a set of *s* equiprobable symbols. In keeping with Shannon (1948), log base 2 will be used.

Let *s* = A…Z (plus space) = 27

Let *n* = the number of selections in APPLES ARE RED = 14

$H = \log_2 (27)^{14}$

$= \log_2 (109{,}418{,}989{,}131{,}512{,}370{,}000)$

= 66.568425 "Hartley-*bits*"

There is obviously no relation between the quantity 66.568425 and the amount of everyday epistemic information associated with the utterance *Apples are red.* The non-calculable amount of (everyday epistemic) information associated with *Apples are red* depends on the context in which the statement appears.

Hartley's (1948) De-psychologizing

Hartley loosely and needlessly fastens his measurement concept, *H,* to a representational theory of mind through the concept of a 'mental symbol' and to a telementation model of communication (Harris, 1981). He claims that in understanding an utterance *a sequence of symbols is brought to the listener's attention*. Furthermore, confusing 'statements' with 'selections from a set' leads Hartley to argue that since audible speech "consists of an acoustic or electrical disturbance which may be expressed as a magnitude time function…We have then to examine the ability of such a continuous function to *convey information*" (pp. 542-543, emphasis added). Hartley regards 'convey' and 'transmit' as interchangeable. But in explaining what $H = log\ s^n$ *means*, Hartley (1928) has to sever these ties, because the relation between Hartley's *H* and these metaphysical theses is strictly rhetorical. So, Hartley is also at pains to clarify that what Shannon (1948) later designated as the "irrelevant…semantic aspects" of communication do not pertain to his *H* either (p. 379).

Hartley (1928) observes that "as commonly used, information is a very elastic term," and accordingly he finds it "necessary to set up for it a more specific meaning …" (p. 536). This *specific meaning* for 'information' is developed for engineering purposes. Hartley's intention is to develop a quantity that is "useful for estimating the possible increase in performance which may be expected to result from improvements in apparatus or circuits, and also for detecting fallacies in the theory of operation of a proposed system" (p. 535). However, and most confusingly, Hartley (1928) describes his $H = log\ s^n$ as an "*expression for the information content of the symbols* at the sending end" (p. 544, emphasis added). He then demonstrates the irrelevance of psychological and/or philosophical questions by emphasizing that, from his engineering perspective, there is no distinction between an intentional act of communication and a random event. He describes a system for which signals are constituted by "disturbance[s] transmitted over [an electrical] cable…" and observes that in certain cases, such electrical disturbances might be the "result of a series of conscious selections. However," he continues to say, "a similar sequence of arbitrarily chosen symbols might have been sent by an automatic mechanism which controlled the position of the key in accordance with a series of chance operations such as a ball rolling into one of three pockets" (p. 537). In other words, Hartley's (1928) concern is with *selections*, not the meaning of and/or the thinking (or lack thereof) responsible for said selections. So *H* values have nothing to do with *mentally* selecting any symbols, that is, *determining which symbols to select*. There is a difference between *selecting message M* and *message M being selected*.

Whereas his choice of the term 'information' for his *H* is confusing, Hartley's arguments for the exclusion of psychological factors are clear and compelling:

> In estimating the capacity of the physical system to transmit information we should ignore the question of interpretation, make each selection perfectly arbitrary, and base our result on the possibility of the receiver's distinguishing the result of selecting any one symbol from that of selecting another. By this means the psychological factors and their variations are eliminated and it becomes possible to set up a definitive quantitative measure of 'information' based on physical considerations alone. (p. 538)

Why is Hartley concerned with "the possibility of the receiver's distinguishing the result of selecting any one symbol from that of selecting another"? His goal is to "evaluate a [signal] transmission system in terms of how well the wave received over it permits distinguishing between the various possible symbols which are available for each selection" (p. 544). In other words, Hartley (1928), like Shannon (1948), is exclusively concerned with identifying which signal has been transmitted, not with what transmitted signals are signals *for*.

Conclusion

In this second chapter, the everyday epistemic sense of 'information' was sharply distinguished from Shannon's (1948) statistical concept of 'information'. (*H*). In tracing this distinction, I identified manifold substantive incompatibilities between Shannon's (1948) *Mathematical Theory of Communication* and Weaver's (1949) speculative commentary on Shannon's (1948) mathematical theorems (Ritchie, 1986). The widespread practice of attributing *information-theory* to "Shannon and Weaver" — as if they co-authored a single work (e.g., Adams, 1991; Seow, 2005; see also Ritchie, 1986) — was thereby shown to be highly misleading. There is a world of difference between Shannon's (1948) mathematical analyses of mechanical concerns, on the one hand and, on the other, Weaver's (1949) far-reaching (and unfounded) musings regarding the significance of Shannon's work for social scientific investigations of communication, more broadly construed.

On the one hand there are the restricted, information-theoretic senses of 'information, 'communication,' 'signal,' and 'message.' These concepts apply exclusively to phenomena that correspond, in at least certain respects to Shannon's (1948) model (Fig 2.1).[24] However, it would be overstated to say that information-theoretic concepts can only be validly applied to electronic communication systems. For example, a fair coin toss can be coherently described as an event involving 1 *bit* of information, and a fair die roll is demonstrated above to involve approximately 2.6 *bits*. So, there are valid applications of certain information-theoretic concepts to phenomena other than those for which Shannon (1948) developed information-theory. However, the validity of such non-Shannonian applications of information-theoretic concepts hinges upon the existence of relevant correspondences between the Shannonian and non-Shannonian applications, respectively.

With respect to the coin toss and/or die roll, the coherence of associating such events with amounts of *bits* is established by the facts of (a) a coin toss and/or die roll involving a selection from a set of pre-defined alternatives (head-or-tails, for the coin toss and 1…6 for the die roll) and, (b) accepting assumptions regarding the probability of

[24] The phrase 'Shannon's model of a communication system' might suggest that Shannon (1948) intended his graphic (Figure 1) as representing an essential property common to *all* communication systems. High-profile instances of this misinterpretation include Miller (1951, 1953, 1956) and Weaver (1949).

each alternative (p = 0.5 for either both head or tails; $p \approx 0.17$ for each of 1…6). So, although neither a coin toss nor a die roll constitute a communication system as defined by Shannon (1948, see also Figure 2.1), both events do involve a selection from a set of alternatives, and therefore, the amount of information (*H*) in Shannon's (1948) statistical sense associated with each event can be computed.

The information-theoretic sense of 'communication' entails reproducing at $Point_2$ a selection made at $Point_1$. The reproducible selection must be made from a predefined set of alternative messages, for which 'message' must be understood as a particular value of a (categorical and/or continuous) variable (for example heads and tails are two possible values for a coin toss-as-message-set). For illustrative purposes, and with Shannon's (1948) model in mind (see Figure 2.1), imagine that the set {A…X} constitutes a set of alternative messages that may be selected and encoded into a transmittable signal form, say, the Morse Code. 'Communication,' in this information-theoretic sense is constituted by a signal-receiving module correctly identifying the signal sequence (i.e., the dot-and-dash combination) transmitted to it and correctly decoding that signal sequence back into message form, for example, correctly decoding the Morse Code signal sequence "dot-dash" into the message A. The amount of information, in the statistical, *H*-value sense, that is associated with the event of this transmission is a joint function of the probability of message A being selected and the number of alternative messages eliminated by its selection (e.g., message A being selected from the set {A…X} eliminates 25/26ths of the set). Despite certain highly misleading word-choices on the parts of Hartley (1928) and Shannon (1948), it is essential to understand that information, in the statistical sense of 'information,' cannot be transmitted. Neither are messages transmitted. Rather, only signals are transmitted, and the distinction between encodable messages and transmittable signals is crucial. For if messages were transmitted and received, then there would be nothing for signal-receivers to reproduce, and therefore there would be no such thing as communication, in the information-theoretic sense of 'communication.' For, in the information-theoretic sense, 'communication' means nothing but "reproducing at one point either exactly or approximately a message selected at another point" (Shannon, 1948, p. 379.)

Shannon (1948) explains that the “semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages” (p. 379). But it is very difficult to bear in mind the highly restricted senses in which such familiar terms as ‘information, ‘communication,’ ‘signal,’ and ‘messages’ are employed in information-theoretic analyses. Whereas there is such a thing as ‘transmitting information’, in the everyday sense of ‘information,’ (for example, by sharing computer files and/or broadcasting news reports) there is no such thing as ‘transmitting information’ for the information-theoretic sense of ‘information.’ And whereas the everyday, non-information-theoretic senses of ‘signal’ and ‘message’ can often be used interchangeably (for example, a flashing red light on a device might be described as either/both an ‘error message’ and/or ‘an error signal’), the respective information-theoretic senses of these terms must be sharply distinguished.

What makes it so difficult to respect the distinctions between the everyday and information-theoretic senses of these terms? For one thing, it is easy to confuse message *composition* with message *selection*. That is, it is tempting to imagine that, in general, ‘deciding what to say’ is analogous to ‘selecting which message to encode and transmit.’ Now, in certain cases, the identification of ‘deciding what to say’ with ‘selecting which message to transmit’ is justifiable. For example, Weaver (1949) refers to “a man picking out one of a set of standard birthday greeting telegrams” (p. 6). For the selector of a standard birthday-telegram, deciding what to say *is* deciding which message to select. However, it is obviously false to suppose that all cases of deciding what to say involve selections from a pre-defined set of alternatives, at least, not in the relevant sense of ‘pre-defined set of alternatives.’ The set of all possible utterances is infinite, and there is no such thing as a proportion of infinity, so the event of a particular utterance cannot be coherently construed as eliminating a precise quantity of alternatives in the same manner that, say, the selection of greeting_2 from the set $\text{greeting}_1 \ldots \text{greeting}_{10}$ involves an eliminating 9 out of 10 alternatives. Shannon (1948) is perfectly clear on the requirement of finitude: “If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely”

(Shannon, 1948, p 379). So, whereas the event of a particular standard telegram being selected can be associated with a precise quantity of information (*H*) in Shannon's (1948) statistical sense, it is not clear that the event of a person composing a message from scratch, so to speak, can be understood information-theoretically.

Of course, it is possible to construe the English language (or any other spoken language in which a verbal message might be composed) as a message source, and associate via stipulation particular probabilities with particular words, phrases, letters, phonemes, etc. (Shannon, 1948). The possibility of analyzing any sample of verbal communication in such information-theoretic terms might appear to suggest that, although the concepts of 'message composition' and 'message selection' are distinct, any case of message composition can be reasonably transformed into a case of message selection. In some sense, this is true. However, it is essential to understand that describing an instance of message composition as information-theoretic message-selection entails construing the compositional process as a stochastic process. That is, information-theory is concerned not with the act of selecting a message but with the fact of a message being selected. That is, information-theoretic concepts pertain not to deciding which signals to transmit but with identifying which signals have been transmitted. This crucial distinction was described above as that between message composition *actions* and signal transmission *events*. In other words, information-theory is a theory of selections, not selecting.

An important source of confusion regarding the strictly mechanical, non agentive character of Shannon's (1948) information-theory is the association of the phrase 'information-theory' with the work of Norbert Weiner (1948, 1950). Weiner, whose principal interest was in human-machine interaction, developed a point of view that he called *cybernetics* and which collapses the distinctions among the following: interpersonal communication (e.g., NN commanding MM to steer towards the right); machine-to-machine communication (e.g., a steering mechanism transmitting a turn signal to a mechanical rudder); human-to-machine communication (e.g., NN causing a steering mechanism to transmit a turn-right signal to a rudder); and machine-to-human (e.g., NN understanding the navigational information displayed by a vessels control panel). Despite the appearance of very many ambiguities in Weiner's (1950) employment of 'information,' provided that his work is read in context, it is easy to appreciate why he "classed communication and control together" (p. 16). In this respect, Weiner's (1948, 1950) interest in communication corresponds to Shannon's (1948). That is, to reproduce at $Point_2$ a message selected for encoding-and-transmission at $Point_1$, *is* for the signal-transmitting mechanism to control the behavior of the signal-receiving mechanism. So, in certain respects, Wiener's (1948, 1950) employment of 'information' is consistent with Shannon's (1948).

However, in places, Weiner's (1950) work involves radical departures from Shannon's (1948) narrow focus on engineering concerns. For example, Wiener (1950) remarks:

> Man is immersed in a world which he perceives through his sense organs. Information that he receives is co-ordinated through his brain and nervous system until, after the proper process of storage, collation, and selection, it emerges through effector organs, generally his muscles. (p. 17)

Clearly, Weiner (1950) does not have *H*-values in mind when he refers to information being "co-ordinated through [the] brain and nervous system…" (p. 17). But neither does Weiner appear to have the everyday, non-information-theoretic sense of 'information' in mind. Rather, Wiener appears to be referring to the undefined sense of 'information' that is involved in 'information-processing psychology.'

The clarifications above regarding what I referred to as the double-life definition of everyday, epistemic 'information' help to explain the intuitive appeal of this apparently, but superficially, technical sense of 'information.' As demonstrated above, the everyday sense of 'information' is an essentially epistemic concept that can be used to refer to either/both *sources* of knowledge and *instances* of knowledge. Instances of information include both (1) statements, objects, and/or events from which agents *become informed* and from perceptions of which *knowledge might be derived;* and (2) *knowledge potentially derivable* from perceptions of statements, objects, and/or events *also counts* *as* information, in the everyday sense. Both a *source of* knowledge *that p* as well as an *instance* of knowledge *that p* can count as information.

In referring to “information that” is “receive[d and] co-ordinated through…[the] brain and nervous system” Wiener (1950, p. 17) drew imaginative, apparently metaphysical conclusions from the grammatical fact of ‘information’ denoting both sources and instances of knowledge. To set the record straight, on the one hand there is the kind of information that exists in the environment, the kind of information that Wiener refers to in referring to *receiving information* that is “perceive[d] through…sense organs” (p. 17). However, there is also the kind of information that is possessed by knowledgeable agents who know “what is conveyed or represented by a particular arrangement or sequence of things” (“Information,” 2010). Now, it is clear that neurological activity is related to informed and/or knowledgeable behavior, for example, it is by virtue of the neurological activity associated with the exercise of sensory and perceptual faculties that organisms become informed of the features of their environments. However, Wiener (1950) also supposes that the neurophysiological causal correlates of sensing and/or perceiving involve the “process[ing]” of information: “Information that [is] receive[d] is co-ordinated through [the] brain and nervous …” (p. 17). But the information that informed agents possess is distinct from the causal correlates of their capacity to be informed or uninformed.

In Chapter 4, the distinction between capabilities and mechanisms is further investigated. In this Chapter 2, historically significant sources of confusion with regard to the ‘information’ in ‘information-processing psychology’ were identified in the works of Hartley (1928), Weaver (1949), and Weiner (1950). In particular, it was demonstrated that there is a tradition of assuming that Hartley’s (1928) and Shannon’s (1948) respective theories of signal transmission are somehow consistent with a representational theory of mind and a telementation model (Harris, 1981) of (at least) linguistic communication. This is false; Hartley’s (1928) and Shannon’s (1948) respective mathematical analyses of communication-as-selection are substantively unrelated to any psychological theories and/or philosophical theses. However, before his “Elimination of psychological factors,” Hartley (1928, p. 538) invites confusion by uncritically assumes that human conversation involves the selection of mental symbols (i.e., representations) by a speaker whose speech transmits *information* (in an undefined, apparently metaphysical sense) that elicits the same selection of mental symbols in the mind of those who hear and understand the speech (Harris, 1981; see also Ogden & Richards, 1923).

In the following section, George Miller's (1951, 1953, 1956) celebrated attempts to incorporate Shannon's (1948) information theory into psychological theorizing are investigated. The clarifications above provide an essential context in which Miller's apparently overlooked misapprehensions can be diagnosed and remediated.

Chapter 3.

George Miller and Information-processing Psychology

George Miller is often lionized as a pioneer of information-processing psychology. For example, in one of Miller's recent obituaries, Stephen Pinker said that "George Miller, more than anyone else, deserves credit for the existence of the modern science of mind…He was certainly among the most influential experimental psychologists of the 20th century" (as quoted by Vitello, 2012). Certain of Miller's influential books and papers from the decade between 1948 and 1958 (Frick & Miller, 1951; Miller, 1951, 1953, 1956; Miller & Frick, 1949) are investigated below. Collectively, these works can be heuristically viewed as exemplifying three phases in the emergence of information-processing psychology: (1) an initial phase during which Shannon's work with Markoff processes is construed as a methodological innovation, allowing the investigation of serially-dependent relations within sequences of behavior, for example, a series of responses in a maze-learning study; (2) a second phase during which Shannon's elaboration of Hartley's (1928) statistical concept of 'amount of 'information'' is viewed as supplying experimental psychology with *a new subject matter*, in particular, experimental participants capacities to *transmit information*; and (3) a final phase, which persists to this day, during which the Hartley/Shannon statistical concept of 'information' is regarded as inadequate for the study of biological agents' epistemic and agentive abilities, although *the idea* of those agents' minds and/or brains 'constituting information-processing systems' in some undefined sense persists.[25] To be clear, I do not argue that

[25] Students of linguistic precision will notice that this situation is subtly difficult to describe clearly. Because the meaning of 'information' involved in characterizing epistemic agents as 'information-processing systems' is unclear, it is also not clear that 'the idea of epistemic agents constituting information-processing systems' actually counts *as an idea.*

these are distinct, chronological phases through which the discipline of psychology in general passed. Rather, they are superimposed on the decade of 1948-1958 by this writer for the purposes of demonstrating how Shannon's (1948) information-theoretic concepts were misapplied by pioneering information-processing psychologists.

Documenting Distortions

It is widely and correctly recognized that Shannon's information theory was an essential stimulus to the development of "research on human information-processing, as the cognitive movement was called early on" (as cited in Mandler, 2002, p. 342). In an autobiographical history of what he calls the "cognitive (r)evolution" Mandler recalls that:

> One of the most direction-giving occasions was the "Special Group on Information Theory" of the Institute of Electrical and Electronics Engineers which met at MIT in 1956…At that meeting Noam Chomsky, George Miller, and Alan Newell and Herbert Simon presented the initial papers of a trend that would be defining in the next decade. (p. 346)

Shannon's (1948) Math versus Wiener's (1948, 1950) Metaphor

How did information theory give direction to these developments? It is widely held that information theory involves a *metaphor* that is ultimately more valuable to psychologists than are Shannon's (1948) mathematics. For example, in a commentary on Miller's (1956) famous *Magical Number Seven…* paper, Baddelely (1994) writes:

> In the late 1950s and early 1960s, information theory seemed likely to transform experimental psychology and form an essential component of any psychologist's education, and yet it is now rarely mentioned. Why should that be? The information-processing approach taken by Miller had two components; the first involved the general concept of the organism as an information-processing system, whereas the second comprised a specific mathematical theory of 'information' that allowed the capacity of the system to be accurately measured. Although the information-processing metaphor has been enormously influential in developing the field that became known as cognitive psychology, the precise measures of 'information'-processing capacity have proved to be much less valuable. (p. 354)

Baddeley (1994) is correct in identifying "two components" of Miller's approach: "the general concept of the organism as an information-processing system" (p. 354) and "a specific mathematical theory…" (p. 354). In Chapter 2, it was demonstrated that the "concept of the organism as an information-processing system" (p. 354) involves Wiener's very hazy 1948/1950 concept of 'information' whereas the "mathematical theory…[of 'information' measurement]" (as cited in Baddeley, 1994, p. 354) involves Shannon's (1948), precisely and quantitatively defined, *statistical* concept of 'information', *H*, as in $H_i = \Sigma p_i \log_2(1/p_i)$, for which p_i is the probability of message/event *i* being selected from a set of alternatives. These two senses of 'information' are logically distinct.[26] However, this is not to say that Shannon's (1948) work is *incompatible* with Wiener's (1950) concept of the organism as an information-processing system; Shannon is neither compatible nor incompatible with Wiener, in this regard. Rather, the essential point is that Shannon's (1948) and Wiener's (1948) respective concepts of 'information', being fundamentally distinct, are just *not-interchangable*. That is, if one is discussing *H* values, then one is not discussing Wiener's (1950) concept of 'information' as "a name for the content of what is exchanged with the outer world as we adjust to it, and make our adjustment felt upon it," a name for what a organism "receives" through sensory and perceptual process and which is later "co-ordinated through his brain and nervous system until, after the proper process of storage, collation, and selection, it emerges through…his muscles" (p. 17). Although Wiener's concept of 'information' is anything but clear, it is absolutely certain that his statements pertaining to organisms as information-processors do not amount to claims that organisms process *H* values.

Baddeley (1994) is correct to distinguish between these "two components" of Miller's approach (p. 354). This distinction can be further specified by identifying these two components as reflective of Shannon's (1948) and Wiener's (1948, 1950) respective contributions. It is demonstrated below that Miller (1951, 1953, 1956) chronically failed to observe the very distinction that Baddeley (1994) draws in identifying the "two

[26] They are historically related in the sense that Shannon and Wiener are reported to have found one another's work mutually inspiring. For example, Weaver (1949) reports that "Shannon…emphasized that communication theory owes a great debt to Professor Norbert Wiener for much of its basic philosophy" (p. 1).

components" of Miller's approach (p. 354). So it might be surprising that Baddeley (1994) praises Miller's (1956) *Magical Number…* paper for what Baddeley judges to be "...a beautifully clear exposition of Claude Shannon's mathematical theory of 'information'..." (p. 353). In light of the fact that Shannon's statistical sense of 'information' is literally *defined* by a set of equations typified by $H_i = \Sigma p_i \log_2(1/p_i)$, the critical reader might be suspicious of Baddeley (1994) praising Miller (1956) for explaining Shannon (1948) "totally without recourse to mathematics" (Baddeley, 1994, p. 353). And, given that Shannon's engineering concerns are highly specialized and technical, it might also be concerning that the *Magical Number…* paper describes Shannon (1948) "in terms that" Baddeley (1994) judges to be " immediately comprehensible to the novice…" (p. 353).

A clue as to why Baddeley's (1994) assessment of Miller is so glowing is found in his comment that Miller "*demonstrat[ed] the need to go beyond information measures*… (p. 353*,* emphasis added). What kind of "going beyond" is alluded to here? The suggestion appears to be that in walking in Wiener's (1948, 1950) shoes by conceiving of organisms as information-processors (in some, undefined sense), Miller (1956) thereby strides *ahead* of Shannon's (1948) mathematical theory of communication and/or his theory of signal transmission (see Ritchie, 1986, for more on this helpful distinction between these two aspects of Shannon's work). But this is incorrect. The arguments below will demonstrate, *pace* Baddeley (1994) that (a) Miller's (1951, 1953, 1956) pioneering works involved fundamental distortions of Shannon's information-theoretic concepts, and (b) involved an unjustified and unidentified assumption of substantive continuity between Shannon's (1948) mathematical work and Wiener's (1948, 1950) metaphysical speculations. This misperception persists. For example, it is apparent in Baddeley's (1994) claim that "The *idea* of 'information' as abstract, but nevertheless measurable...and the general utility of the concept of limited channel capacity have been enormously influential…and continue to be valuable" (p. 353, emphasis added). The strangeness of this argument becomes apparent when its logical form is rendered: Measuring *X*-processing capacity has proven to be much less valuable than the *idea* of *X* as measurable (see also Lachman et al., 1979, pp. 74-75).

The arguments below will demonstrate that such a *continuity interpretation* of the relation between Shannon's (1948) information theory and Miller's (1951, 1953, 1956) information-processing psychology is grossly distorted. It is demonstrated below that Miller's seminal works in information-processing psychology involve a variety of specific errors including:

(a) an overgeneralized conception of the sense in which Shannon's theorems relate to 'uncertainty reduction';

(b) confusing Shannon's stipulations regarding his model of communication with natural laws of communication;

(c) confusing the degree to which a statement is informative with the amount of 'information', in the statistical sense, that would be associated with the *message* that corresponds to said statement in the context of a Shannonian communication system;

(d) confusing the *signal sequence* that corresponds to a message with the *amount of 'information'*, in the statistical sense, associated with that message, resulting in the mistaken idea that *bits* of 'information', in the statistical sense, are *transmitted*;

(e) confusing the everyday epistemic, and statistical senses, respectively, of 'information';

(f) confusing the possible states of a system with the set of messages that *describes* those system-states;

(g) failing to observe the distinction between symbols and signals that is essential to Shannon's model;

(h) confusing the event of a particular random sequence with a particular sequence of random events;

(i) confusing 'number of binary digits' with 'amount of 'information', in the statistical sense' due to the fact of *bits* being the *unit* in which information, in the statistical sense, is measured;

(j) confusing a language with a code; and thereby

(k) mistaking (certain cases of) thinking for recoding.

I will demonstrate that Shannon's (1948) analysis of *signal* sequences as Markoff processes was misinterpreted in terms of existent assumptions regarding the *symbolic,* or representational nature of mind. In Chapter 2, I documented Hartley's (1928) particular contribution to confusion between mathematical theories of signal transmission and representational theories of mind. In this section, Miller's (1951, 1953, 1956) elaboration of Hartley's (1928) error is investigated.

Miller's (1951, 1953, 1956) Misunderstanding of Shannon (1948)

In his book, *Language and Communication,* Miller (1951) claims under an apparently Shannon-inspired heading, *The Idealized Communication System,* that "communication means that information is passed from one place to another" (p. 6). How are we to understand the concept of 'information' that is invovled this claim? Miller's elaboration makes it clear that he has Shannon's information theory in mind:

> Whenever communication occurs, we say that the component parts involved with the transfer of information comprise a communication system. Although the specific character of these parts changes from one system to another, there are general functions that the components must perform if the communication is to succeed…Every communication must have a *source* and a *destination* for the information that is transferred….Between the source and the destination there must be some…*channel*…In order that the information can pass over the channel, it is necessary to operate on it in such a way that it is suitable for transmission...At the destination there must be a *receiver* that converts the transmitted information into its original form. These five components—source, transmitter, channel, receiver, and destination —…are present in every kind of communication. In most communication systems the source of the information is a human being. From his past experience and present needs and perceptions this source has information to pass along to others. The transmitter…is the human speech machinery. This machinery operated upon the information and changes it into a pattern of sound waves that is carried through the air.
> (pp. 6-7)

Before considering how this passage fares as an information-theoretic analysis, it will be helpful to notice that, with respect to the everyday epistemic sense of 'information', Miller's initial claim that "communication means that information is passed from one place to another" (p. 6) appears to be roughly true or, perhaps more precisely, to be a truism, a statement that might be useful for explaining what either (but not both) 'information' or 'communication' mean in their respective everyday senses. However, it is false to claim that information, in Shannon's statistical sense, is *transmitted*. But this is difficult to understand due to the misleading practice of misconstruing *signal transmission* as *information transmission,* identified as error (d) above. For example, Shannon (1948) himself misleadingly states that "Teletype and telegraphy are two simple examples of a...channel for *transmitting information…"* (p. 381, emphasis added; see also Hartley, 1928). But there is a difference between the *amount of information*, in the statistical sense, associated with a particular message and the *signal sequence* into which that message must be encoded for transmission. It is signal sequences, not values of the *H* statistic, that are transmitted. Signal transmission is the means by which the transmitting and receiving ends of a communication system affect one another, respectively. And signals can take many different forms (e.g., radio frequencies, patterns of a light beam, etc.). However, it is not possible for a signal to take the form of information, in the statistical sense, because an 'amount of information' in the statistical sense is simply a numerical value, akin to a particular value of a Pearson product moment correlation, for example.

Furthermore, there is an important difference between correctly claiming on the one hand that certain cases of conversations among human beings can be given an information-theoretic analysis, and, on the other hand, falsely claiming that Shannon's restricted definition of a 'communication system' denotes "general functions that…components must perform if…communication is to succeed" (Miller, 1951, p. 6). As Ritchie (1986) observes, Shannon's work does not:

> Constitute…a *communication theory* as students of human communication understand the term, but a general theory of *signal transmission*…[and] his theory was designed specifically for the system he described…[Shannon did not] prove that the formula $H = \Sigma p_i \log_2(1/p_i)$ is the "best" measure for information, or even that it *is* a measure for information. Rather, he stated a set of criteria to describe a measure that would serve the requirements of his signal transmission theory, and demonstrated that the entropy formula meets those criteria. There is nothing in Shannon's paper to justify use of this formula under any other circumstances. (p. 281)

Ritchie's (1986) conservative impulse is understandable and admirable in light of the widespread misunderstanding of Shannon's (1948) work. However, for the purposes of understanding *how* and *why* Shannon's work was misunderstood by Miller (1951, 1953, 1956) in particular and information-processing psychologists more generally (e.g., Broadbent, 1958), it will be helpful to adopt a more empathic approach and consider that psychologists' experimental situations vary in their degree of correspondence to Shannon's criteria and, accordingly, in the degree to which they can coherently described in information-theoretic terms. Below, Miller's (1956) treatments of experiments regarding absolute judgment capacity (e.g., distinguishing among different auditory tones) on the one hand and those regarding short term recall (e.g., for a random sequence of binary digits) on the other are contrasted with respect to their degree amenability to information-theoretic analyses.

Nevertheless, it is perfectly clear that Shannon's (1948) statistical analyses of signal sequences are not concerned with the aspects of communication relevant to Miller's (1951) observation that "from his past experience and present needs and perceptions [the human being] has information to pass along to others" (p. 7). In contrast, Shannon (1948) wrote:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. (p. 379)

Putting Shannon (1948) and Miller (1951) side by side, it is clear that, *pace* Miller, the type of information that the "[human being] has…to pass along to others" is "irrelevant" (p. 7) to information-theoretic analyses. Miller's error is identified as (e) above: confusing the everyday epistemic, and statistical senses, respectively, of 'information'.

Perception and Sensation as Shannonian Communication

Beyond conflating the everyday epistemic and statistical senses of 'information', respectively, Miller (1951) emphasizes the continuity between communication and perception, positing a representational character to both phenomena:

> It is common to begin a discussion of communication by pointing out that words are *signs* that conveniently replace the objects or ideas they represent. It would be misleading to imply, however, that this representative character of words distinguishes them sharply from all other stimuli…The word 'chair' is clearly not the chair itself, but a symbol for the chair. Similarly, the light reflected from the object is not the chair itself. In either case, the response is made to something that represents the chair. (p. 4)

Miller assumes that because both the word 'chair' and the perception *of* a chair are distinct from chairs themselves, this establishes that both the word 'chair' and the perception of a chair constitute representations. Certain problems associated with conceiving of (sensations and) perceptions as representations are identified briefly in **Chapter 4.**[27] However, the task at hand is merely to establish the vastness of the conceptual gap between Miller's representational information-processing psychology and Shannon's (1948) explicitly non-psychological information theory. Consider, for example, Miller's (1951) claim that:

> In an ever-changing world, the state of the organism and the state of the environment must be able to mold and direct the organism's behavior. *The necessary information* is supplied by specialized cells called receptors and neurons. Because of these cells the central nervous system is affected *indirectly* by the changes going on in and around the organism…. As far as the brain is concerned, the activity of the sensory cells stands for, or represents, the stimuli. (p. 4, emphasis added)

[27] Most succinctly, the fundamental problem with representational theories of mind is that they presuppose that mental and/or neural representations have intrinsic meanings. For example if knowing what a chair is is causally dependent upon possessing a mental representation of chairs, then such chair-representations must intrinsically represent chairs. But, as Putnam (1989) puts it "none of the methods of representation that we know about – speech, writing, painting, carving in stone, etcetera – has the magical property that there cannot be different representations with the same meaning. None of the methods of representation that we know about has the property that the representations intrinsically refer to whatever it is that they are used to refer to. All of the representations that we know about have an association with their referent that is contingent, and capable of changing as the culture changes or as the world changes. This by itself should be enough to make one highly suspicious of theories that postulate a realm of ''representations'' with such unlikely properties" (pp. 21-22).

In light of the clarifications in Chapter 2, it is clear that Miller (1951) has Wiener's (1948, 1950) rather than Shannon's (1948) work in mind, although Miller appears to be unaware of it.[28] Is Miller (1951) unique in assuming equivalence between this *representational* type of information and Shannon's statistical concept of 'information'? Hardly. The outlines of this picture are present in contemporary information-processing psychology as well. Gazzaniga, Irvy, and Mangun (1998) write that "Encoding refers to processing information to be stored…[and] retrieval utilizes stored information to create a conscious representation or to execute a learned behavior…" (p. 247, footnote). And, on the topic of Shannon's influence on philosophers of mind, Adams (1991) suggests that:

> If one thinks of the mind as receiving information from the environment, storing and coding that information, and then causally guiding behavior in virtue of the stored representational content, it is not too hard to see why information theory would apply to these elements. (pp. 472-473)

However, if one both accepts this representational view of mind, and also understands Shannon's (1948) work, then it *is very* "hard to see why" information theory should be so attractive (Adam, 1991, p. 473). For, as Adams also correctly observes, "a treatment of the *amounts* of information," in Shannon's statistical sense "and the mathematics of the time series and oscillations that may be applied to the nervous system [does] not [concern] the *contents* of such information signals being processed" (p. 473). But to conceive of "the mind as receiving information from the environment, storing and coding that information, and then causally guiding behavior in virtue of the stored representational content" *is* to assume that the contents of the signals it processes are relevant to behaviors produced (p. 472). For example: NN visually perceives a chair which he then treats as a chair. Clearly, if NN's chair-related perception and behavior is said to involve the processing of visual information, then the posited (yet undefined) information must be information *of* or *about* a chair. For if the information processed in seeing a chair is not distinct from the information involved in, say, seeing a table, then it

[28] However, Miller (1951) departs from Wiener (1950) in positing a *representational* quality to the information that organisms are hypothesized to process.

is not clear how an appeal to 'visual information-processing' can be viewed as potentially explanatory of the perceptual behavior of NN's treating a chair as a chair.

How is this apparent internal inconsistency supposedly resolved? By erroneously implying that information, in the statistical sense, is a type of information that *has* content, although such informational contents are said to be irrelevant to purely quantitative information-theoretic analyses. Miller (1953) offers the following deeply misleading advice: "Only the *amount* of 'information' is measured—the amount does not specify the content, value, truthfulness, exclusiveness, history, or purpose of the information" (p. 3). But this is false. Shannon's (1948) work does not constitute a technique for analyzing quantitative aspects of communication and/or information, in the respective everyday senses of those terms. The labyrinth of errors ((a)-(k) above) running through this misinterpretation of Shannon's (1948) information theory and into information-processing psychology's contemporary *impressionistic* (Dennett, 2005) and putatively *implicitly defined* (Sloman, 2011) sense of 'information' is surveyed below.

Markoff Processes and the Statistical Structure of Behavior

Prior to the emergence of information-processing psychology, pioneering psychologists noticed certain similarities between the concerns of communication engineers and those of behavioral psychologists. Accordingly, Miller and Frick (1949) "attempt[ed]…to formulate certain psychological problems in such a way as to take advantage of the techniques developed by" researchers in disparate fields (p. 324). In particular, Miller and Frick find inspiration in Shannon's (1948) analysis of communication a Markoff processes. Such Markoffian analyses involve a highly restricted definition of 'communication' as reproducing a sequence of serially dependent selections from a pre-defined set of alternative messages, for example, the 32 possible messages of the Morse Code, each of which is correlated with a particular "dot/dash" signal. Statistical regularities in the employment of symbols can be exploited such that systems are designed to anticipate that, say, E, is more likely to follow the sequence SAF than is X, on the assumption that the signal sequence corresponding to the English word *safe* is more likely to be transmitted that is the signal sequence corresponding to the non-word *safx*.

Miller and Frick (1949) observe that many phenomena which interest experimental psychologists also involve serial-dependencies, for example, the sequences of responses emitted by an experimental subject that knows its way about a maze. Clearly, for a skillful maze runner, a, say, correct left-turn response, R_n, is dependent upon the sequence of responses that were previously emitted (R_{n-1}, R_{n-2}, etc.). It is a logical truth that to know one's way about a maze *is* for one's responses to be serially-dependent; a sequence of *random* responses that led to successful completion of maze would constitute good luck rather than genuine knowledge. The choice of turning one way or the other is, loosely speaking, guided by the previous choices of that maze-running expert, not unlike successive message selections are guided by previous selections. However, such 'guidance' must be qualified. It is essential to bear in mind that Shannon's (1948) information theory is not a theory of agents' communication *choices*. In Shannon's (1948) restricted sense of 'communication' as a statistical phenomena, the normative, volitional, semantic, and epistemic aspects are explicitly and intentionally left out. So it is may be clearer to say that Shannon's concern with frequency relations among messages (or, message components, for example, the message HI consist in two components, H and I, respectively) pertains to *what is likely to happen* in the process of message-reproducing signal exchange rather than what communicating agents are *likely to do.* In Chapter 2, this distinction was described as that between *selecting message M* and *message M being selected.*

Serially Dependent Actions versus Serially Dependent Events

Miller and Frick (1949) observe that construing certain behavior sequences as Markoff processes allows psychologists to investigate previously-ignored aspects of their data. The authors understandably lament that their contemporaries are lacking "procedures that can be used to analyze serial dependencies in chains of responses" (p. 311). So, for example, in a maze-learning situation, for which patterns among sequences of serial responses are of great interest, the authors note that psychologists "largely neglected…sequential information…" in their classification and analysis of subjects' responses. But "ignoring the sequential information is equivalent to assuming that successive responses are independent" and such an assumption of response independence is clearly problematic (p. 313).[29] The order of responses is crucial; not just any combination of *n* right turns and *n* left turns will do. For an organism to have learned its way about a maze *is* for its responses to be non-independent. Accordingly, Miller and Frick are "surprised…to discover how few psychological studies have reported the…sequential information necessary for" analyses of serial dependencies among sequential behaviors (p. 321). They argue that "the concept of the course of action seems to us better adapted to most psychological problems than the concept of isolated response-units or attributes of such units" (Frick & Miller, 1951, p. 36). This concept would also expand the scope of inquiry for experimental psychology.

> If interest is confined to individual responses and no attempt is made to describe and discuss sequences of responses, experimental psychology may have difficulty in meeting many problems posed in the clinical and social areas. Some such concept as the course of action seems inevitable. (Miller & Frick, 1949, p. 322)

In contrast to Shannon's analysis of signal transmission, for which the agentive aspects of communication are intentionally left out, Miller and Frick (1949) emphasize that sequences of maze-running behaviors are more accurately construed as a course of *actions* than as a series of *events*. However, the authors are careful to distinguish their *methodological* argument regarding the advantages of *representing learning as* a Markoff process from a *substantive* claim regarding *the nature* of learning:

> We have too long classified types of learning according to the particular experimental procedure involved in measuring the learning. *Whether or not all learning is reducible to a single basic process*, it seems possible that the results of all learning experiments can be described in a uniform

[29] The authors' reference is clearly to the everyday epistemic sense of 'information'; they are concerned with correcting invalid assumptions and improving our scientific *knowledge* of behavior.

manner and the results of one experiment compared with the results of others. (Miller & Frick, 1949, p. 324, emphasis added)

Miller and Frick (1949) describe a "point of view which [they] propose to call *statistical behavioristics,* [a] point of view [that] represents a methodological bias that [they] believe is accepted by many experimental psychologists," without, however, having ever "received explicit formulation nor exploit[ation]" (p. 324). The influence of American behaviorism, but also cautious ambivalence regarding its more strident claims, is apparent in their choice "to call this approach *behavioristics"* which they justify by stating that "behavior seems to be the observable datum of psychology" (p. 324). It is difficult to imagine a psychologist of any ideological and/or methodological stripe who would disagree that behavior is "the observable datum" of psychology. What *was* a contentious issue for psychologists of the early 20th Century in particular is the extent to which the "observable data" of behavior constitute valid premises from which to draw conclusions concerning putatively unobservable mental phenomena.[30]

Serial-dependence as Learning

Not until the end of their paper do Miller and Frick (1949) hint that Shannon's work might be relevant to the vexed issue of mentalism. Viewed retrospectively in light of Miller's influence on the emergence of information-processing psychology, it is easy to appreciate that it *was not predictive precision per se that interested Miller and Frick.*[31] Rather, the predictability of an experimental participants' behavior is of particular interest because it indicates the degree to which the participants' sequential responses are serially dependent. And the extent to which responses are serially dependent, in turn,

[30] See Bennett and Hacker (2003, pp. 68-108) for an analysis that dissolves the familiar problem of unobservability with respect to mental phenomena.

[31] This is an unabashed example of interpreting the past in light of present knowledge. To be clear, I am not claiming that Miller and Frick (1949) were consciously concealing an interest in mental phenomena for political reasons. Rather, my argument is that the seeds of the information-processing revolution yet to come are visible to the historian who enjoys perspicacity of hindsight. However, for an analysis that helpfully foregrounds the continuity between the experimental work across both the behaviorist and information-processing eras, see Mandler (2002).

sheds light on certain of the participant's *mental processes*. For example, an analysis finding a greater degree of behavioral stereotypy in the serial responses of a group of human female participants than in those of a (single) rat who completed the same task is interpreted in terms of what those participants "remember[ed]" and "discovered." Subsequently, Miller and Frick state that this "experiment shows that girls have a greater capacity for *symbolic processes* than do rats" (p. 322, emphasis added).

Now, Miller's and Frick's (1949) explicit intention is to use Shannon's work on Markoff processes to *enrich* rather than *transform* psychology's then-traditional focus on observable behavior. They offer "a quantitative index of predictability, or…an *index of behavioral stereotypy*" (p. 317): "C_s = 1—(relative U_s), for which U_s indicates 'uncertainty.'" The authors explain that "a value of *C* of 0.50, for example, is equivalent to the statement that half of the responses are determined and that half are maximally uncertain" (p. 320). In this statement, 'determined' means 'predictable' whereas 'maximally uncertain' means 'unpredictable.' The index of behavioral stereotypy can range from 0, indicating that prediction is no better than chance, to 1, indicating that the agent's responses are perfectly predictable. In this context, 'stereotyped' means 'predictable' and greater predictability is interpreted as evidence of *learning*. For example, a rat that has learned to find its way through a maze would be a rat that emits highly predictable, or stereotyped, behavior.

Information-theoretic versus Cognitive Uncertainty

Miller and Frick (1949) treat 'predictability' and 'certainty' as interchangeable.

> If there are only a few alternative classes a prediction is more certain than if there are many alternatives. 'Anything can happen' is a common expression of uncertainty. In other words, the uncertainty, *U,* is a monotonic increasing function of the number of alternatives. (p. 317)

The passage quoted above evinces the error identified as (a) above: an overgeneralized conception of the sense in which Shannon's (1948) work relates to 'uncertainty reduction.' For if Shannon's concepts are to be applied, then it is false to claim that *anything* can happen. In to satisfy Shannon's (1948) most fundamental criterion, each "actual message [must be] *selected from a set* of possible messages" (p. 379). And if it is *events* rather than *messages* that the information-theorist is interested in, then it is necessary for the set of possible events to be explicitly delimited so that the pattern of selections from that set can be represented in information-theoretic terms. So, although "anything can happen" may be a common expression of uncertainty, it does not apply to the type of uncertainty with which information theory is concerned (Miller & Frick, 1949). Information-theoretic uncertainty is a "monotonic increasing function of the number of alternatives," but there are very many cases of uncertainty for which this definition would be nonsense. For example, it is possible to begin a walk feeling certain that Broadway is east of 5th Avenue at 20th Street in New York City, but to become increasingly uncertain about this as one walks uptown. Clearly, such an increase in uncertainty would not involve a monotonic increase in the number of alternatives; Broadway is either east or west of 5th, and that's that. The very restricted sense in which 'uncertainty' figures in information theory is exemplified in the following passage, in which Miller and Frick (1949) characterize "a logarithmic function of the number of alternatives" as constituting a desired "additive measure of uncertainty" (p. 317).

> Suppose a choice is made from among 10 equally probable alternatives, and that, having made this first choice, a second choice must also be made from among 10 equally probable alternatives. Two successive choices from 10 alternatives make possible 100 alternative pairs of choices—when the alternatives are equally probable and successive choices are independent, two choices from 10 alternatives are equivalent to one choice from 100 alternatives. If the uncertainty in these two situations is represented by a linear function of the number of alternatives, the first choice from 10 alternatives represents 10 units of uncertainty, and the second choice adds 90 more. It is not obvious why the second event should be assigned nine times the uncertainty of the first, and it would be more satisfactory if a choice from 10 alternatives always represented the same amount of uncertainty. (p. 317)

Strictly speaking, one might object that although 'un/certainty' admits of degree—for example, a person can be more or less certain of one thing or another—it is not clear that such varying degrees of un/certainty can be divided into units. Nevertheless, provided that 'uncertainty' is understood as a term of art with a restricted definition, Miller and Frick's (1949) argument is reasonably clear. And it is easy to appreciate their motivation to promote a statistical tool that designed for analyses of serially-dependant sequences. They succinctly summarize as follows.

> It has been assumed that a behavior sequence can often be represented by a process with a definite number of states. If the process is in any given state there are several possible events that can follow. Consequently, there is a degree of uncertainty associated with each state. In one state the next event may be very certain, while in another state the next event may be relatively uncertain. (pp. 319-320)

It is essential to note that the "uncertainty associated with each state" does *not* refer to the *behaving subject's* un/certainty about which response it will produce next, or about which response is the *correct* response. 'Uncertainty' in the restricted, information-theoretic sense, pertains exclusively to *non-eliminated alternatives* such that as alternatives are eliminated, uncertainty is said to be reduced, with the proportion of alternatives eliminated by a message and/or event expressed in terms of *bits* of information, in Shannon' (1948) statistical sense of the term. The information-theoretic sense of 'uncertainty' is distinct from what Ritchie (1986) calls the *cognitive uncertainty* that either an experimenter and/or a research participant might experience:

> Shannon's use of such terms as *uncertainty*…can only be understood in the context of a statistical discussion. He was not addressing *cognitive* uncertainty of a human individual trying to figure out, for example, "What could Aunt Mary have meant by that telegram?" (p. 283)

But it is not just the type of *interpretive* uncertainty described above that we must distinguish from information-theoretic 'uncertainty.' For with respect to an information-theoretic analysis of Miller and Frick's (1949) maze-learning situation—and this is a difficult point—there is a risk of conceiving of information-theoretic 'uncertainty' as identical with an *experimenter's* predictive uncertainty. Rapoport (1956), for example, erroneously personalizes the information-theoretic sense of 'uncertainty' by writing that "the receipt of the message transmitted without error 'destroys' the uncertainty of the recipient, with regard to which message will be chosen" (p. 304). But, strictly speaking (and it is crucial to be strict here), information-theoretic uncertainty isn't the uncertainty *of* a message recipient, or of any other agent. For, in contrast to the everyday sense of uncertainty ("the state of being uncertain," see "Uncertainty," 2013) and/or the typically pluralized sense of 'uncertainties' (e.g., "something that is uncertain or that causes one to feel uncertain: [for example] *financial uncertainties,*" see "Uncertainty," 2013), information-theoretic uncertainty does not pertain to any agent's state of mind and/or feeling. Rather, the information-theoretic sense of 'uncertainty' was created to satisfy certain quantitative criteria, described by Miller and Frick (1949) as follows:

> We would like to have an additive measure of uncertainty. This is realized if uncertainty is a logarithmic function of the number of alternatives. Then the uncertainty of the first event is log 10 and the uncertainty of the second event is log 10; the total uncertainty for two successive events is 2 log 10, which is equal to log 100, the uncertainty of one choice from 100 alternatives. (p. 317)[32]

Shannon (1948) and later, Miller (1953, 1956) preferred to use the log $base_2$ form in calculating *H* values in the unit of the *bit*. However, irrespective of the logarithmic base utilized, how certain or uncertain experimenters might *feel* and/or the degree of *confidence* they might express in a particular probability estimate are distinct from the information-theoretic sense of 'uncertainty' as a "logarithmic function of [a] number of alternatives" (Shannon, 1948, p. 379). It is the predictability of a system, and not anyone's degree of confidence in their predictions that constitutes information-theoretic 'uncertainty.' This sharply distinguishes such predictability from the everyday sense of 'uncertainty' that pertains to states of mind. The importance of this distinction will become increasingly apparent as this argument proceeds and attention turns to Miller's (1953, 1956) subsequent attempts to apply Shannon's (1948) statistical sense of 'information' to psychological phenomena.

[32] The authors neglect to mention that this equation assumes that each of the alternatives are equiprobable: "If all the p_i are equal, $p_i = 1/n$, then [uncertainty] should be a monotonic increasing function of *n*. With equally likely events there is more choice, or uncertainty, when there are more possible events…If a choice be broken down into two successive choices, the original *H* should be the weighted sum of the individual values" (Shannon, 1948, p. 389).

Engineers' vs. Behavioral Psychologists' use of *Markoff Processes*

Despite the applicability of Shannon's work on Markoff processes to certain psychological experiments, there are a variety of important differences between the situation of the communication engineer and that of the experimental psychologist. Communication engineers are interested in the statistical properties of signal transmission patterns such that they can design communication systems that are custom-made for the signal sequences that they are expected to transmit. Shannon (1948) explains:

> the main point at issue is the effect of statistical knowledge about the source in reducing the required capacity of the channel, by the use of proper encoding of [messages][33]. In telegraphy, for example, the messages to be transmitted consist of sequences of letters. These sequences, however, are not completely random. In general, they form sentences and have the statistical structure of, say, English...The existence of this structure allows one to make a saving in time (or channel capacity) by properly encoding the message sequences into signal sequences. (pp. 383-384)

So, it is the behavior of signal-transmitting machines rather than the actions of message-composers that concerns Shannon. However, confusingly, it is also true that knowledge regarding message-composers' behavior is relevant to calculating amounts of information, in the statistical sense. For example, consider an information-theoretic analysis of an English language text messaging system. The symbol sequence LOL (laugh out loud) might constitute a high-probability message due to the fact that the system's users are expected to select the message, LOL, with some relatively high degree of frequency. So it would be misleading to state that the behavior of message selectors is *irrelevant* to information-theoretic analyses because, in certain cases, estimating the frequency with which a particular signal sequence will be transmitted is identical with estimating the frequency with which the message that corresponds to said signal sequence will be selected. Nevertheless, it is also mistaken to conceive of information-theoretic analyses as analyses of message-selecting behavior; it is *selections* and not the act of *selecting* that is relevant.

Hartley (1928) provides a helpful reminder that, for an information-theoretic analysis, a message that is the "result of a series of conscious selections" (p. 537) is no different from a "sequence of arbitrarily chosen symbols…sent by an automatic mechanism…in accordance with a series of chance operations such as a ball rolling into one of three pockets" (p. 537). Grammatically, the essential distinction is that between *selections* (a noun) and *selecting* (a verb).

In contrast to the engineering concerns described above, Miller and Frick (1949) are concerned with detecting statistical regularities *in the behavior of research participants* for the purposes of drawing inferences regarding what those behaving subjects have *learned*; the maze-running participant plays the part of the message source, and the index of behavioral stereotypy quantifies the degree to which the participants behavior is predictable, with increasing values regarded as evidence of *learning*. Of course, in the context of a psychological experiment, there is no question of reproducing "either exactly or approximately" any selected messages (Shannon, 1948, p. 379). Rather, in predicting future behavior on the basis of previously observed behaviors, it is assumed that previously observed behaviors are identical with previously emitted behaviors. So, the engineer's "problem of communication" (p. 379) is very different from the psychologist's problem of predicting a "course of action" (Frick & Miller, 1951, p. 36) even though both situations can be represented as Markoff processes.[34] Shannon (1948) is concerned with evaluating the performance of a machine whereas Miller and Frick (1949) investigate the skilfulness of a learner. However, to observe these

[33] Note that in the quoted passage I have replaced Shannon's original reference to "proper encoding of the information" with "proper encoding of [messages]" because, as is correctly implied in Shannon's very next sentence ("In telegraphy, for example, the messages to be transmitted consist of sequences of letters…") it is *messages* and not *information*, in the statistical sense, that are encoded into transmittable signal form.

differences is not to establish that Shannon's theorems are misapplied by Miller and Frick. Despite the contrast between the engineers' and the behavioural psychologists' respective interest in serially-dependent sequences, Miller and Frick's adaptation of Shannon's mathematics for analysis of serially dependent behavior sequences is particularly commendable in light of their insightful critique of (mis)representing sequential maze-running responses as series of discrete trials.

Mentalistic Interpretations of Behavioral Stereotypy

Although Miller and Frick (1949) do not refer to measuring amounts of information, the non-behavioristic character of information-processing psychology is anticipated in their interpretation of *their index of behavioral stereotypy* as representing the degree to which a subject is *strategic* and/or *thoughtful* in its responses (with values ranging from 0, indicating that responses are random to 1, indicating that responses are perfectly predictable). Behavioral predictability is interpreted as evidence of strategizing and/or thoughtfulness. In turn, strategizing and thinking are assumed to be *symbolic*—that is *representational*—processes. This is manifest in Miller and Frick's (1949) analysis of a historical data set. The experimental context to which this historical data set pertains is described prior to examining Miller and Frick's (1949) mentalistic interests and their symbolic (or representational) gloss.

> In 1916 G. V. Hamilton…used an apparently unsolvable multiple-choice problem that required the subjects to find their way out of an enclosure from which there were four possible exits. The correct exit varied from trial to trial, so it was useless to try again the exit that had proved successful on the preceding trial. No other clue was available. This apparatus was supposed to prevent the formation of a specifically adaptive habit and to provide an extensive sample of trial and-error behavior…The question to ask of these data is how unpredictable or how stereotyped is trial-and-error behavior in this situation? Does a subject systematically vary his responses to produce a random sequence of exit

[34] Another important difference between the *C, index of behavioral stereotypy* (*C*) statistic and Shannon's 'amount of 'information'' (*H*) statistic is that *C* ranges from 0 to 1 whereas *H* values are in principle unlimited.

> attempts, or does he settle down to a relatively routine and stereotyped pattern that is adequate though not always rewarded? (p. 321)

Hamilton's experiment is cleverly designed to distinguish the type of learning that results from reinforcement from that which would appear to require thoughtful deliberation (as cited in Miller & Frick, 1949). Miller and Frick note that it is exceptional for including complete data pertaining to participants' sequences of responses. Thus, it is possible to calculate indices of behavioral stereotypy for each of the participants. Especially interesting is that it includes data for one rat, as well as for a number of 7-year-old female children. Miller and Frick use a series of comparisons between the girls' and the rat's responses to illustrate the benefits of analyzing serial dependences in behavior sequences.

> The first approximation for the behavior of the five girls shows the proportion of responses to… [be] almost uniform, with only a slight tendency for the girls to favor the middle two doors. When there are four alternatives, the maximum value of U is $\log_2 4$, or 2 bits. The calculated value of U_1 for these data is 1.96. The relative uncertainty is, therefore, 1.96/2, or 0.98. The index of stereotypy for the first approximation is $C_1 =$ 0.02. With no knowledge of preceding responses any prediction of the next exit to be tried is very uncertain. (p. 321)

In other words, if each serial response is treated as independent, the trial-and-error behavior of these experimental subjects appears to be almost random. In this respect, the behavior of these human subjects is mostly indistinguishable from that of a rat.

> Inspection of the tabulations of pairs of responses shows a striking similarity between girls and rat. The same six pairs of responses that the girls favoured were favoured by the rat, and comprised 78 per cent of its responses. For the rat C_1 was 0.06, and C_2 was 0.26 On the basis of a second-order approximation it is difficult to discover any difference between the animal's and the girls' adjustment to the situation. This in spite of the fact that it took the rat approximately twice as many responses to complete 100 successful exits as it took the average girl…it is necessary to consider the third-order approximation before the difference becomes clear. The sequences of three responses show that the girls tended to make a third response different from either of the two preceding responses. There was no need, they *discovered,* to try again a response that was unsuccessful the time before last. This tendency appears in the form of greater stereotypy in their behavior…The rat, on

> the other hand, was not as proficient in *remembering* what he had done on the trial before the last, and often fell into the unproductive pattern of which 3, 2, 3, 2, 3, 2, etc., is an example. (p. 322, emphasis added)

Although their paper is couched primarily in terms that are superficially consistent with the anti-mentalism of American behaviorism, it is clear that Miller and Frick (1949) are interested in *mentalistic* interpretations of this maze-running behavior. They write of what Hamilton's female subjects "*discovered*" and which subsequently "appear[ed] in the form of greater stereotypy in their behavior" (p. 322). The rat, on the other hand, "was not as proficient in *remembering* what he had done on the trial before last" (p. 322, emphasis added). These descriptions might appear to be almost unavoidable, unless one is dogmatically committed to eschewing the mentalistic vocabulary that is endemic to non-technical, English-language explanations of behavior. However, Miller and Frick go far beyond presupposing the validity of such every day, mind-related language. They go on to write that:

> If one agrees…that a response influenced by the penultimate of the preceding responses indicates a symbolic process, then Hamilton's experiment shows that girls have *a greater capacity for symbolic processes* than do rats. This is not a startling conclusion, but it is comforting to find that the computations do not contradict the obvious. (p. 322, emphasis added)

The claims above do not merely presuppose the concepts of 'discovering' and 'remembering.' By implying that remembering is a "symbolic process," Miller and Frick (1949) offer an unelaborated theory of *how remembering works*. That they consider this to be an "obvious" conclusion demonstrates the degree to which a representational theory of mind is uncritically assumed even by psychologists who are (temporarily, at least) invested in avoiding mentalistic concepts altogether (p. 322).[35]

[35] The anti-mentalism of radical behaviorism (e.g., Skinner, 1953) and/or eliminative materialism (e.g., Churchland, 1981) are typically viewed as the only alternative to a representational theory of mind.

Information Measurement as a Research Method in Psychology

In a 1953 article, George Miller invites psychologists to consider not just what Shannon's (1948) statistical analyses of Markoff processes might offer them, but how Shannon's statistical concept of 'information' might transform the discipline. Miller (1953) responds to the question that constitutes the title of his paper, *What is Information Measurement?* with considerable rhetorical skill and attempts to clarify what he characterizes as a buzz among certain experimental psychologists: "In recent years a few psychologists, whose business throws them together with communication engineers, have been making considerable fuss over something called 'information theory'" (p. 3). Miller's article is clearly intended to encourage psychologists' excitement about information theory. However, he also issues a commendable warning against an overzealously reductive information-theoretic approach to psychology: "With the development of 'information' theory we can expect to hear that animals are nothing but communication systems. If we profit from history, we can mistrust the 'nothing but' in this claim" (p. 3). Thus, Miller appears to be suggesting that information-theoretic analyses might supplement rather than replacing other data analytic methods.

Shannon's Stipulations versus Natural Laws

Just a few lines down, however, Miller (1953) suggests that Shannon's information theory involves certain fundamental truths about psychological subjects: "Insofar as living organisms perform the functions of a communication system, they must obey the laws that govern all such systems" (p. 3). This is a much stronger claim about the relevance of information theory to experimental psychology. Although Miller does not claim that psychologists *must* construe their subject matter in information-theoretic terms, he claims that psychological subjects "must obey the laws that govern" communication systems, as defined by information theory (p. 3). This 'must' is worth considering carefully. Is the sense in which any particular communication system "must obey the laws that govern all such systems" akin to claiming that physical objects must obey the laws of gravity? Not at all. For we expect that the phenomenon of gravity affected the behavior of physical objects long before the invention of a concept of 'gravity.' Thus, although the concept 'gravity' is clearly a human invention, it is obvious that the *phenomenon* of gravity existed prior to its being *discovered* by physicists. Such gravity-related phenomena as oceanic tides and apples falling from trees to earth do not depend on any person having discovered any underlying principles that apply to either or both of these apparently disparate occurrences.

In contrast, the fundamental concepts of information theory apply exclusively to systems that *correspond to Shannon's idealized model of a* "communication system" (see Figure 2.1). Of course, correspondence comes in varying degrees so, for example, it is possible to use the example of a coin toss to explain the meaning of '1 *bit* of information' (although other fundamental information-theoretic concepts such as 'channel capacity' would not apply in any obvious manner to this situation). In Chapter 2, I emphasized that Shannon's "model" of a communication system is also a definition of what *counts as* a 'communication system' in the context of his information theory. And so, it would be misleading to say that Shannon *discovered* that his theorems applied to such systems. Rather, Shannon *invented* a sharply-bounded concept of a 'communication system' precisely because the behavior of such systems could be analyzed in terms of Markoff processes (see also Ritchie, 1986). So the sense in which communication systems "must" obey the "laws that govern all such systems" is akin to the way in which a bachelor must be an unmarried man (Miller, 1953, p. 3). Therefore, to claim that "insofar as living organisms perform the functions of a communication system, they must obey the laws that govern all such systems" is just to claim that certain of living organisms' behavior sequences *can be represented in information-theoretic terms*

(p. 3). Miller's error here is identified as (b) above: confusing Shannon's stipulations regarding his model of communication with natural laws of communication.[36]

A natural question to ask is, "*Why should* certain of organisms' behavior sequences be represented information-theoretically?" Miller (1953) notes that pioneering psychologists of his day:

> …drop words like "noise," "redundancy," or "channel capacity" into surprising contexts and act like they had a new slant on some of the oldest problems in experimental psychology. Little wonder that their colleagues are asking, 'What is this information you talk about measuring?' and 'What does all this have to do with the general body of psychological theory?' (p. 3)

Miller (1953) responds to his rhetorical question:

> The reason for the fuss is that information theory provides a yardstick for measuring organization. The argument runs like this. A well-organized system is predictable—you know almost what it is going to do before it happens. When a well-organized system does something, you *learn* little that you didn't already *know—you acquire little information*. A perfectly organized system is completely predictable and its behavior provides no information at all. The more disorganized and unpredictable a system is, the more *information you can get* by watching it.
> (p. 3, emphasis added)

[36] However, there *is* a connection between Shannon's *H* statistic and a natural law. Shannon himself explicated that the form of his *H* statistic measuring amount of 'information', in the statistical sense, related to the concept of 'entropy,' which the OED defines as "a thermodynamic quantity representing the unavailability of a system's thermal energy for conversion into mechanical work, often interpreted as the degree of disorder or randomness in the system: the second law of thermodynamics says that entropy always increases with time…" (OED 3rd Ed., 2012, online). Shannon (1948) writes that "form $H = - \Sigma\, p_i \log p_i$ of *H…*will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell *i* of its phase space. *H* is then, for example, the *H* in Boltzmann's famous *H* theorem" (p. 379). So, assuming the second law of thermodynamics, it would make sense to say that the entropy of a system *must* increase over time. Nevertheless, it *only* makes sense to say that "living organisms…must obey [Shannon's] laws" (Miller, 1953, p. 3) if this is understood as expressing that *information-theoretic descriptions* of living organisms are possible.

‘Amount of Information’ vs. ‘Degree of Informativeness’

In the passage quoted above, Miller (1953) ironically obscures the very distinction he intends to clarify. The most glaring of the errors involved in this passage is that Miller implicitly equates the statistical concept of ‘amount of information’ with ‘degree of informativeness,’ in the everyday epistemic sense (e.g., *The talk was more informative than I expected*). He writes that “you learn little that you didn't already *know—you acquire little information*” (p. 3) from observations of expected behavior. This is identified as error (c) above: confusing the degree to which a statement is informative with the amount of information, in the statistical sense, that would be associated with the *message* that corresponds to said statement in the context of a Shannonian communication system.[37]

This misidentification of ‘amount of ‘information’,’ in the statistical sense, with ‘degree of informativeness,’ in the everyday epistemic sense, runs parallel to Miller and Frick’s (1949) overgeneralized misstatement of the manner in which ‘uncertainty reduction’ pertains to Shannon’s (1948) theorems (error (a) above). The sense in which H values can be said to measure *reductions in uncertainty* is highly restricted. Not all cases of uncertainty pertain to selections from sets of alternatives. One can be uncertain of, that is, not know, very many different kinds of things, such as what one wants to eat for lunch, the answer to 678,945 ÷ 297, and/or the year in which the French revolution occurred. Furthermore, information-theoretic descriptions pertain to potential states of an observed system and not actual states of an observer’s mind.[38]

[37] Fifty years after Miller (1953), the confusion persists. For example, Adams (1991) begins by correctly observing that with respect to an information-theoretic analysis, “[t]he more likely an event, the less information it generates — while the less likely the event, the more information it generates.” However, he erroneously continues to say that “For example, on any random day, telling you truly that it is going to rain today is *more informative* in Phoenix than Seattle” (p. 476, emphasis added).

[38] Although it is possible to do, it is also Procrustean to construe these cases as a processes of *predicting* what one is going to eat, and/or of *predicting* the value of the variables which constitutes the true answers to the questions *678,945 ÷ 297 = ?* and *In which year did the French Revolution occur?*

With respect to information-theoretic analyses, that an event is of low probability *entails* that its occurrence will be associated with a greater ‘amount of information’ than will a higher-probability event. So, to oversimplify a bit (pun irresistible), the lower probability of the event and the greater H value associated with its occurrence are the same thing. The reason that high-probability events are associated with lesser amounts of information (H) is *not* because observers learn little from observations of high probability events (the falsity of ‘observers learn little from observations of high probability events’ is demonstrated below). To suppose otherwise is to commit a variation on error (c) above: confusing the degree to which a statement is informative with the amount of ‘information’, in the statistical sense, that is associated with the *message* which corresponds to said statement. The variation is confusing the degree to which *observing an event* is informative with the amount of ‘information’, in the statistical sense, which is associated with the event’s occurrence. The reason that higher probability events and/or messages are associated with lower amounts of ‘information’, in the statistical sense, is just because the ‘amount of ‘information’,’ in the statistical sense, associated with particular events and/or messages is an inversely proportional function of the frequency with which they are expected to occur. Who can learn how much from what is irrelevant.

It will be helpful to relate these arguments to an equation that *defines* the ‘amount of information’ associated with message/event i: $H_i = -\log_2 p_i$, for which p_i indicates the probability with which i is expected to occur. “If the message probabilities are $p_1, p_2 \ldots p_k$, then the amounts of ‘information’ associated with each message are $-\log_2 p_1$, ,$\log_2 p_2 \ldots -\log_2 p_k$ (Miller, 1953, p. 4). As in Chapter 2, consider the comparison of two message sources, so to speak: a coin toss and the roll of a die, respectively. The probability of a heads-side-up outcome in a fair coin toss is $p = 0.5$ whereas the probability of rolling a, say, six is approximately $p \approx 0.17$. Accordingly, the amount of ‘information’, in the statistical sense, associated with a fair coin toss is exactly 1 *bit.*

$H_i = -\log_2 p_i$; $p_i = 0.5$

$-\log_2 (0.5) = \log_2 (1/.05) = \log_2 (2) = 1.$

The less-likely event of rolling a six is associated with the greater amount of approximately 2.6 *bits*:

$$H_i = -\log_2 p_i;\ p \approx 0.17$$

$$\log_2 (0.17) \approx \log_2 (1/0.17) \approx \log_2 (5.9) \approx 2.6.$$

These calculations demonstrate that lower-probability messages and/or events are associated with greater amounts of 'information', in the statistical sense, by definition, and not because, the receiver/observers of such messages/events "learn little that [they] didn't already know" and therefore "acquire little information" (Miller, 1953, p. 3).

Furthermore, it is false that observers necessarily learn little from observations of high probability events. For example, from observing the highly predictable behavior of a simple motor, an observer may learn much about how the motor works. Broadly speaking, observing empirical regularities would appear to play a foundational role in accumulating knowledge about the world. And it is because the behavior of properly functioning digital calculators is highly predictable that such devices are routinely relied upon to provide such information as, say, the answer to 256 x 75. The predictability of a digital calculator would appear to be *essential* to its role in providing such mathematical information. Clearly, the conditional frequency of an event and the degree to which observing it may be informative (i.e., *how much you can learn from it)* are logically distinct. And recognition of this distinction can be found even in foundational works of information theory. For example, as Hartley (1928) points out, "in communication as viewed from the psychological standpoint…the single word 'yes'…when coming at the end of a protracted discussion, may have an extraordinarily great significance" (p. 540). Davies and Gardner (2010) report that 'yes' is estimated to be the 259th most frequently occurring word in the English language. 'Yuppie,' by way of contrast, is estimated to be the 18,739th most frequent word. Unless 'yuppie' is playful for 'yes', *Yes* is a more informative response to, say, *Is it raining?* than is *Yuppie,* demonstrating that the extent to which an utterance is informative is not identical with the expected frequency of its occurrence.

On the Idea of a Statistical Process

How are Shannon's (1948) engineering concerns with Markoff processes imagined to relate to communication in general? Miller (1951) writes that:

> The very nature of communication makes it clear that it is a variable, *statistical kind of process*. If we could predict in advance exactly what a talker was going to say, then he would not need to say it. The very fact that communication occurs implies that the accuracy of predictions of communicative behavior is limited. (p. 8, emphasis added)

This passage exemplifies what Wittgenstein (1953) identifies as "A main cause of philosophical disease—a one-sided diet: one nourishes one's thinking with only one kind of example" (para. 593). For this line of reasoning neglects to account for the very many cases of linguistic communication involving highly predictable behavior in the form of conventions, violations of which would appear to be quite noticeable and interpretable, for example, A*: Hi, how are you?* B: *Fine, thanks, and you?* versus A: *Hi, how are you?* B: *[silence]* (Sacks & Jefferson, 1992; Turnbull, 2004). Second, it is invalid to infer from the fact that linguistic communication is not completely predictable that *the phenomena* of linguistic communication are a "statistical kind of process." For one thing, it is particular forms of representation that can be coherently characterized as 'statistical,' and not the phenomena to which such statistical descriptions might refer. The distinction that Miller appears to have in mind might be more clearly expressed by saying that if linguistic communication is to be represented quantitatively, the associated data must be analyzed probabilistically rather than deterministically.

To be clear, it is undoubtedly true that there are cases for which Miller would be correct in asserting that "If we could predict in advance exactly what a talker was going to say, then he would not need to say it." For example, if A knows in advance that B's response to a particular question will be *No*, then there is a sense in which B "would not need to say it." In contrast, however, if A and B are lovers, whoever fails to say *Happy birthday* and/or respond with *I love you too* might have some explaining to do.

What Kind of Information can be Acquired?

A grammatical consequence of the *restricted range* of cases to which the statistical concept of 'information' can be validly applied is that *it is only the everyday epistemic sense of 'information' that can serve as the direct object of 'acquire'* as in, *I acquired the information from a website.* In very many cases, *knowing that p* is synonymous with *possessing the knowledge that p*, and, analogously, *becoming informed that p* is synonymous with *acquiring the information that p*. But it does not make sense to speak of an amount of information in the statistical, *H*-value sense as being acquired because 'to acquire' means to 'come to possess' and one cannot come to posses the amount of information (*H*) associated with a particular message and/or event. *Pace* Miller (1953), the amount of information expressed by an *H* value of, say, 2.4 *bits* is *not* that type of information that "you can get by watching" something (p. 3).

It is helpful to bear in mind that *H* is a *statistic* such as a Pearson *r*, for example. The *H* value associated with a particular message and/or event represents the degree to which its occurrence eliminates members of a pre-specified set of alternatives, and is inversely proportional to the probability of the event's occurrence; less probable events are associated with higher *H* values. Accordingly, it does not make sense to say that an agent *acquired* an *H* value, just as it is self-evidently clear that, say, $r = .07$ cannot be acquired. Neither a correlation nor an amount of information (*H*) can be acquired, for the same reason that colorless green ideas cannot sleep furiously (Chomsky, 1957). That is, such statements don't make sense because their meanings cannot be stated (Baker & Hacker, 1984a). Of course, it is possible to acquire the information, *in the everyday sense*, *that* the amount of information, *in the statistical sense*, associated with a particular event is, say, 2.4 *bits*. But the information that is thereby acquired is not identical with the 2.4 *bits*.

This distinction is easier to appreciate if we substitute an *r* value for an *H* value, because we are less likely to be confused by the term 'correlation' than by 'information.' N might acquire *the information that X* and *Y* are correlated at, say $r_{x,y} = .07$. But it is clear that in acquiring this information, NN is not acquiring $r_{x,y}$. Rather, NN is acquiring *the information that…*And the only type of information that can be coherently followed by ‘that’ is every day, epistemic information. In other words, there is no such thing as *N bits that* such-and-such is the case. Unfortunately, the longstanding practice of treating ‘signal transmission’ and ‘information transmission’ (e.g., Hartley, 1928) as synonymous makes this point quite difficult to grasp. That is, although an agent may become informed by receiving and decoding a signal sequence that has been transmitted to him, it is nonsensical to conclude that the agent has thereby acquired information, in the statistical (*H*) sense. To draw this erroneous conclusion is identified as error (d) above: confusing the *signal sequence* that corresponds to a message with the *amount of information*, in the statistical sense, associated with that message, resulting in the mistaken idea that *bits* of information, in the statistical sense, are *transmitted*.

The Statistical Sense of ‘information’ Is Not a Quantified Version of Its Everyday Epistemic Sense

Miller (1953) also wrongly suggests that Shannon’s (1948) statistical concept of ‘information’’ is continuous with the everyday epistemic sense of ‘information’. This is identified as error (e) above. Miller (1953) claims that “Information, organization, predictability, and their synonyms are not rare concepts in psychology,” and that, thanks to Shannon’s information theory, “each place [these terms] occur now seems to be *enriched by the possibility of quantification*” (p. 3, emphasis added). This statement implies that the statistical concept of ‘information’ developed by Hartley (1928) and Shannon (1949) constitutes a quantified version of the everyday epistemic sense of the term. But this is patently false. Ironically, Miller (1953) fails at heeding his own wise warning that:

> Most of the careless claims for the importance of ‘information’ theory arise from overly free associations to the word information. This term occurs in the theory in a careful and particular way. It is not synonomous [*sic*] with ‘meaning.’ Only the *amount* of information is measured—the amount does not specify the content, value, truthfulness, exclusiveness, history, or purpose of the information. The definition does not exclude other definitions and certainly does not include all the meanings implied by the colloquial usages of the word. (p. 3)

Yet, by way of explaining the relevance of the statistical concept of 'information' to psychological phenomena, he asks his reader to:

> imagine a child who is told that a piece of candy is under one of 16 boxes. If he lifts the right box, he can have the candy. The event—lifting one of the boxes—has 16 possible outcomes. In order to pick the right box, the child needs *information*. (p. 4, emphasis added)

Which sense of 'information' is employed here? Miller (1953) is correct in stating that "content, value, truthfulness, exclusiveness, history, [and/or] purpose" (p. 3) are irrelevant to the *statistical* sense of 'information'. Accordingly, we should expect such properties to be irrelevant to the example that he provides. However, it is clear that most of these properties *are* relevant. Miller (1953) continues with his example as follows.

> If we say, 'The candy is under the white box,' we give [the child] all the information he needs—we reduce the 16 alternatives to the one he wants.[39] The amount of 'information' in such statements is a measure of how much they reduce the number of possible outcomes. Nothing is said about whether the information is true, valuable, understood, or believed—we are talking only about *how much* information there is. (p. 4)

Miller (1953) goes on to write that "anything we tell" this hypothetical, candy-searching child "that reduces the number of boxes from which he must choose will provide some of the information he needs. If we say, 'The candy is not under the red box,' we give him just enough information to reduce the number of alternatives from 16 to 15" (p. 4).[40]

An Information-theoretic Analysis of a 16-state System

Before investigating Miller's confusing treatment of this 16-boxes example, it will be helpful to complete a clear information-theoretic analysis of this situation. Fortunately, Miller (1953) provides two very helpful and simple methods for calculating amounts of

39 Assuming that there is only one white box.

40 Assuming that there is only one red box.

information, in the statistical sense, one of which foregrounds the *probabilistic* aspect of the *H* statistic ($H_i = -\log_2 p_i$, as above) and the other of which concerns the *proportion of alternatives eliminated* by a particular message and/or event: "the amount of information [associated with] a message that reduces *k* to *k/x* is $\log_2 x$ bits" (p. 4).[41] The 16 boxes should be understood as constituting the "set of predefined alternatives" that constitute an essential component of Shannon's (1948) definition of a 'communication system.' However, it is also important that these 16 possible locations of the candy constitute 16 possible *states of a system*, not 16 alternative *messages*.

Consider first the equation that pertains to the *proportion of alternatives eliminated* by a particular message and/or event, and for which the equiprobability of each alternative is assumed: "the amount of information [associated with] a message that reduces *k* to *k/x* is $\log_2 x$ bits" (Miller, 1953, p. 4). To calculate the amount of information, in the statistical sense, associated with the elimination of one of these 16 alternatives, we set $k = 16$ and proceed as follows: $16/x = 15$, so $16 = 15x;\ x = 16/15 \approx 1.067$. $\text{Log}_2 (1.067) \approx .093$. Therefore, we can conclude that an event that eliminates one of 16 equiprobable alternatives is associated with approximately 0.093 *bits* of information, in the statistical sense.

To calculate Shannon's (1948) *H* on the basis of *probabilistic* knowledge rather than knowledge of the proportion of alternatives eliminated by an event, proceed as follows. Assuming that each of the 16 boxes are equally likely to be eliminated at the point of any one observation, we can use the equation $H_i = -\log_2 p_i$ to calculate the amount of information, in the statistical sense, associated with the event of any one box being eliminated. However, we must substitute $1\text{-}p_i$ for p_i. This substitution is necessary because the probability of the candy *not being* under one of the 16 boxes is equal to the probability of it *being* under one of the remaining 15. Thus, to observe that "the candy is not under the red box" is to observe that it *is* under one of the remaining 15 alternatives (assuming that there is only one red box, which qualification Miller, 1953, neglects to

41 Miller neglects to mention that this latter equation requires the assumption that each of the *k* alternatives is equiprobable.

include). So, assuming that each of the 16 alternatives is equiprobable, we can set p_i = 1/16 = 0.0625. Therefore, 1-p_i = 1-0.0625 = 0.9375. Substituting known values into H_i = -$\log_2 p_i$ we get:

$$H_i = -\log_2 (0.9375)$$
$$= \log_2 (1/0.9375)$$
$$\approx \log_2 (1.067).$$
$$= 0.09356$$

Therefore, we conclude, as above, that in the context of a system with 16 equiprobable possible states, observing that it is *not* in one of those 16 states is associated with approximately 0.09 *bits* of information, in the statistical sense.

In light of the foregoing, clarifying calculations, consider Miller's treatment of his 16-box-system example. The relevant event or message, so to speak, is the experimenter's statement. Temporarily setting aside the crucial distinctions among statements, messages, and events, is it correct to suppose that the "content, value…[and] purpose" of the information that is supplied to the child by the experimenter's statement is irrelevant, and/or that it matters not whether the child understands and/or believes that the experimenter's statements are truthful? Such suppositions are clearly wrong. With respect to *content*, it is clear that the type of information that this child "needs" in "order pick the right box" is information *about* the location of the candy. Therefore, the content of such statements as, *The candy is not under the red box* is *essential* to Miller's (1953) example. And although, strictly speaking, Miller (1953) is correct that, in his example, "Nothing is *said* about whether the information is true, valuable, understood, or believed," the example nevertheless involves very many *assumptions* regarding the truth, value, comprehension, and credibility of the statements that he erroneously presents as examples of information, in the statistical sense (p. 4, emphasis added). For *The candy is not under the red box* must be *understood* if the experimenter's statement is to *count as* information regarding the location of the candy. If the candy-seeker does not know what *The candy is not under the red box* means, then the event of the experimenter's uttering it has no *value* with respect to the child "reduc[ing] the number of alternatives" which he must search.

Statements versus Messages versus Events

There is a world of difference between the *event* of one or more boxes being *eliminated*, on the one hand, and *the utterance* of a statement that one or more of the boxes *ought* to be *ignored,* on the other. If it is the experimenter's statements that are supposed to yield the relevant reduction in alternatives, then it is clear that the content, value, and purpose of those statements are relevant. For, in order for the child in Miller's (1953) example to regard the experimenter's statements *as information* (rather than as misinformation or disinformation), the child would have to trust that the experimenter is not misleading him, because, of course, *The candy is under the white box* provides the child with "all the information he needs" if and only if the statement is true (p. 4). If it is *false* that *The candy is under the white box* then the statement still reduces the alternatives that the candy-searching child must choose among, provided that he *understands* what that statement *means* and appreciates its *value* within a deductive reasoning process. So clearly, the child's *beliefs* about the speaker's *purpose* and the *truth* of what he says are relevant.

It is potentially confusing, however, that Shannon's (1948) statistical concept of 'information' is said to be internally related to the physics concept of 'entropy;' 'amount of 'information',' in the statistical sense is just negative entropy, of which relation Shannon writes that the "form $H = -\Sigma\ p_i \log p_i$ of *H*…will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell *i* of its phase space" (p. 379). In the context of research in thermodynamics, the relevant predictions do not concern the reproduction of signal patterns that correspond to selected messages but rather concern *states of a system*, for example, the location of a particular particle within a chamber. The amount of information (in the statistical sense) associated with a particular observation expresses the degree to which that observation deviates from its expected value, that is, the expected location of the particle (Tolman, 1938).

It is crucial that the physicist's message source, so to speak, *is the system* whose states he is investigating. And, of course, it is metaphorical to construe an experimenter's (E) *observation* of a system (*S*) in a particular state (*i)* as '*S* transmitting message *i* to E.' But in Miller's example, the message source is *the experimenter*, and the relations among the experimenter's possible messages and the possible states of the system are determined by the content of the experimenter's statements. There is a difference between *observations of* entities in particular locations on the one hand and the receipt of *messages about* the location of entities, on the other. But in Miller's example, the distinctions among messages, statements, and states-of-a-system are lost. In particular, the experimenter's statements are presented as if they constituted the *states* of the candy-box system that they *describe*. But the statement, *The candy is under the red box* is not identical with the *fact* of the candy being under the red box. It is the *fact* of the candy being under the red box that eliminates the possibility of it being in any of the alternative locations, not the utterance of a statement to this effect. To suppose otherwise is identified as error (f) above: confusing the possible states of a system with the set of messages that *describe* those system-states.

Is it valid to regard the statements of Miller's (1953) example experimenter as messages, in Shannon's (1948) information-theoretic sense? Miller quite wrongly assumes that it is and claims that calculating *H* values constitutes a "way to *measure the amount of information in such statements"* as *The candy is under the red box* (p. 4). But an amount of information, in the statistical sense, is not a property of a statement; there is simply no such thing as an amount of information, that is, a value of *H*, in the statement, *The candy is under the red box*. However, provided that (a) THE CANDY IS UNDER THE RED BOX figured among the possible alternative messages in the context of a Shannonian communication system, and that (b) the probability of this message being selected for encoding-and-transmission was known, then there *would be* an amount of information (*H*), in the statistical sense, associated with the *message* THE CANDY IS UNDER THE RED BOX being selected for encoding and transmission. But Miller's example does not meet these requirements. Of course, there is an amount of *everyday epistemic information* in the (true) statement, *The candy is under the red box*. However, because everyday epistemic information is not measurable, we are forced to say something vague about the amount of it, such as that the amount of information in the statement depends on what being apprised of it is worth to its reader. Miller's breezy conflation of messages, in the information-theoretic sense, with statements constitutes yet another manifestation of error (e) above: confusing the everyday epistemic and statistical senses of ‘information’.

Miller’s (1953) candy box example also involves confusion regarding *which alternatives* are relevant to an information-theoretic analysis of this situation. Upon reflection, it is clear that each of the 16 possible states of the candy-box system could be described in an *infinite* number of ways, for example, THIS ONE, THAT ONE, NONE OF THESE, etcetera. So there is no limiting relation between the 16 possible locations of the candy and the set of predefined alternative messages that would accurately describe the location of the sweets (provided that the statements corresponding to said messages were true). Miller’s example confounds alternative messages with alternatives states-of-affairs, exhibiting what is identified as error (f) above: confusing the possible states of a system with the set of messages that *describe* those system-states.

On the Idea of Isolating Amounts of Information

Supposedly by way of establishing the very distinction that eludes him, Miller (1953) writes that in calculating *H* values, “only the *amount* of information is measured—the amount does not specify the content, value, truthfulness, exclusiveness, history, or purpose of the information” (p. 3). However, this implies that the statistical sense of ‘information’ refers to a type of information that *has* content, purpose, and history and that may or may not be true, valuable, and/or exclusive. But this is also false. Particular *messages* have contents and *if* a message consists in a sequence of symbols, then *what is expressed by* said sequence of symbols may or may not be true, valuable, and/or exclusive, etcetera. And particular messages that figure among a set of predefined alternatives within the context of a Shannonian communication system are associated with *H*-values, that is, with definite amounts of information, in the statistical sense. But it is nonsensical to conceive of the “content, value, truthfulness…” that characterize the *content* of messages as “[un]measured” (p. 3) qualities of the ‘amounts of ‘information’’ with which those messages are associated. Miller (1953) inadvertently blurs the very distinction that he purports to sharpen.

Statements versus Messages in Shannon's (1948) Information-theoretic Sense

Miller uncritically assumes that the descriptive statements of the experimenter in his example can be regarded as *messages*, in the information-theoretic sense, that is, sequences of particular symbols selected from a pre-specified set of alternatives. This restricted sense of 'message' corresponds to the restricted sense of 'communication' with which information theory is concerned: "reproducing at one point either exactly or approximately a message selected at another point" (Shannon, 1948, p. 379). Yet, it is obvious that the candy-seeking child's orientation to the experimenter's statements is not one involving this "fundamental problem of communication" (p. 379). Recall that, in Shannon's (1948) words, the "semantic aspects of communication are irrelevant to the engineering problem and that the "significant aspect is that the actual message is one *selected from a set* of possible messages…" (p. 379). But the child in Miller's (1953) example is *not* interested in *reconstructing* one of 16 alternative encodable-and-transmittable messages (p. 379). Rather, the child in Miller's example is concerned with *understanding* the experimenter's statements, drawing inferences from them, and using them as reasons to behave in one way or another. Thus, the child's orientation towards the experimenter's statements is that of an epistemic agent. The child will (1) *know* what to (2) *do* on the basis of what the child (3) *understands* the experimenter's statements to (4) *mean*, and the information that the child seeks is internally related to the epistemic, agentive, and semantic phenomena denoted by verbs (1)-(4). What Miller presents as a purely quantitative, information-theoretic description of an intellectual task turns out to be riddled with the everyday epistemic sense of 'information', with all of its ramifying relations to such non-quantitative, epistemic notions as 'meaning,' 'truth,' 'understanding,' and 'belief.'

Miller (1953) is plainly incorrect in supposing that a message *whose content licenses an inference that eliminates* half of a pre-specified number of alternatives, is a message associated with 1*bit* of information. The amount of information, in the statistical sense, that is associated with a particular message is logically independent from what the symbol sequence consists in might mean or entail, for example, that the candy is under the red box. When Miller writes that "some messages that the source selects involves [*sic*] more information than others," this should be understood as expressing that certain messages, qua symbol sequences, are less likely than others to be encoded into signal sequences and transmitted than are other messages, *not* that certain messages are more informative to their recipient(s) than are others (p. 4).[42] So, with respect to calculating the amount of information, in the statistical sense, that is associated with a particular message, it is *completely irrelevant* that its content may license an inference eliminating half of a pre-specified number of alternatives. The errors involved in Miller's mistakenly identifying the inferential value of a statement with the amount of information, in the statistical sense, that is associated with it include those identified above as (c) confusing the degree to which a statement is informative with the amount of information, in the statistical sense, that would be associated with the *message* that corresponds to said statement in the context of a Shannonian communication system; and (e) overestimating the continuity between the everyday epistemic, and statistical senses, of 'information'.

Unfortunately, these crucial conceptual distinctions are neglected by Miller (1953). As will be demonstrated by reviewing Baddley's (1994) laudatory assessment of Miller's (1956) landmark *Magical Number…* article, it appears that these logical facts remain obscure to many information-processing psychologists, who follow Miller (1951, 1953, 1956) in positing a chimerical continuity between Shannon's information theory and information-processing psychology. Accordingly, a superficially similar, supposedly technical sense of 'information' appears, a concept whose "use in telecommunications and computer technology [gave] it a tough, brittle, technical sound" while it remains "spongy, plastic, and amorphous enough to be serviceable in cognitive and semantic studies" (Dretske, 1981, p. viii). This is the concept of 'information' that figures in 'information–processing psychology' and which has yet to be defined.

[42] To frame this same point in terms of 'observing a system's behavior,' we could say that observing the system in the state of transmitting an improbable signal sequence constitutes an observation associated with an amount of 'information' (in the statistical sense) that is inversely proportional to the improbability of the observed transmission sequence. The amount of 'information' (in the statistical sense) associated with an observation increases as the probability of the observed behavior decreases.

From Bits to Chunks: Miller's (1956) Magic

Baddeley (1994) writes that he was once "intrigued to discover that [a] 1956 volume of Psychological Review opened virtually automatically at [Miller's] article, which in contrast to the pristine state of the rest of the journal was distinctly dog-eared, as if nibbled by generations of hungry [students]" (p. 353). He continues to comment and wonder, respectively: "There is, I think, little doubt that The Magic Number Seven is alive and well; but why?" (p. 353).

Baddeley (1994) on Miller (1956)

Baddeley (1994) attributes the apparent quantity of Miller's (1956) readers to the quality of Miller's work. In particular, Baddeley (1994) states that in Miller's *Magical Number…* article:

> The concept of information is introduced and related to a range of more familiar concepts including both news value and variance, and its use is elegantly illustrated. Miller made it clear that the importance of information theory comes, first of all, from the general concept of the brain as an information-processing machine, a concept that has come to dominate …psychology since that time. (p. 353)

This assessment is puzzling; neither 'brain' nor 'news value' appear even once in Miller's article (although he does relate the statistical concept of 'information' to that of 'variance'). And nowhere in the 1956 paper does Miller write explicitly about the "general concept of the brain as an information-processing machine," although this Wienerian (1948, 1950) idea is clearly foundational for Miller's (1951) earlier book. It appears that Baddeley's (1994) interpretation of Miller's (1956) *Magical Number…* paper is every bit as distorted as Miller's (1956) version of Shannon's (1948) work.

What does Miller (1956) actually say? He reviews psychological experiments in which participants are "considered to be...communication channel[s]" (p. 82). In doing so, Miller applies the concept of 'channel capacity' to the research participant as a whole, *not* to her brain (for more on part-whole, or mereological relations and information-processing psychology, see Bennett & Hacker, 2003). It is true that the (very hazy) idea of "the brain as an information-processing machine" is a fundamental assumption of information-processing psychology, and that the *Magical Number*...paper was an influential stimulus to the discipline's development in the information-processing direction (Baddeley, 1994, p. 353). However, it is important to remain clear on what Miller's (1956) paper actually involves. There is a profound difference between claiming that a research participant can be construed as a communication channel for the purposes of an experiment on the one hand (which is what Miller claims in the 1956 paper) and claiming that a *part of* a participant (i.e., her brain) *is* an information-processing machine on the other (Baddeley's, 1994, contention).[43]

Information Theory and Information-Processing Psychology: The Continuity Interpretation

In certain places, Baddeley (1994) correctly emphasizes the *discontinuity* between Miller's (1956) work and Shannon's (1948) information theory:

> ...the reason that [Miller's] article continues to be influential at a time when information theory is largely ignored within psychology stems from the insights that allowed Miller to go beyond *the restrictions of the theory itself.* In emphasizing the importance of recoding, Miller pointed the way ahead for the information-processing approach to cognition.... (Baddeley, 1994, p. 356, emphasis added)

It is demonstrated below that what Baddeley describes as "go[ing] beyond the restrictions of" information theory also involves transgressing the bounds of sense (p. 356; Bennett &Hacker, 2003).

[43] To be fair, Baddeley's comments do appear to accurately characterize Miller's thinking. For example, towards the end of his career, Miller (2003) mused that the "dream of a unified science that would discover the representational and computational capacities of the human mind and their structural and functional realization in the human brain still has an appeal that [he] cannot resist" (p. 144). However, Baddeley (1994) is undoubtedly inaccurate in ascribing such neurological speculation to Miller's (1956) paper in particular.

What might account for the venerated status that Miller's (1956) paper enjoys? In a similarly laudatory assessment of the *Magical Number...* article, Shiffrin and Nosofsky (1994) highlight the significance of Miller's prose:

> Why has this article been following us around through the literature all these years? There can be no question that writing style has a great deal to do with this. Of course, Miller was as capable of writing impenetrable academic prose as any of the rest of us when he wanted to...[but this] article was of a different sort... It contains all those qualities that make an article a delight to read, qualities that almost always preclude publication: *elimination of inessential detail*, clarity of expression combined with a casual and anecdotal style, superposition of fundamental and important ideas without elaborate attempts at justification, and, of course, a good dose of humour. (p. 360, emphasis added)

Although they may be right to praise Miller's (1956) paper for its readability, Shiffrin and Nosofsky (1994) are wrong to characterize the details of Shannon's (1948) work as inessential to the coherent application of his information-theoretic concepts. After investigating the influential 1956 paper in detail, Miller's (1989) own less enthusiastic and perhaps surprising reflections on his celebrated paper will be reviewed. It will be demonstrated that Miller's (1956) *Magical Number…* paper is free of "elaborate attempts at justification" of its supposedly "fundamental and important ideas" (Shiffrin & Nosofsky, 1994, p. 360) precisely because those inchoate would-be ideas are fundamentally unjustifiable.

Miller's (1956) Verbal Innovations

Miller's (1956) widely-celebrated (2,178 citations as of September 10, 2012) "Magical Number..." paper summarizes paradigmatic, early applications of information *theory* by experimental psychologists, and illustrates how Shannon's mathematical work stimulated the *imaginations* of information-processing psychologists, whose *verbal,* as opposed to *mathematical,* innovations unwittingly buried equivocation errors deep in the conceptual foundation of contemporary psychology. In particular, the arguments below will demonstrate that what Baddeley (1994) celebrates as "the heart of" (p. 354) Miller's article—Miller's novel concept of '*chunking*'—constitutes a massive confusion between the everyday epistemic and statistical senses of 'information', respectively, diagnosed as error (e) above. Moreover, Miller's (1956) hazy concept of a 'chunk of information' (he admits that he is "not very definite about what constitutes a chunk of information") also involves another, subtler equivocation error that appears to be both a *symptom of* confusion regarding information as well as a *stimulus to* the maintenance of this mistake (p .93). Miller misidentifies 'number of binary digits,' (e.g., '1 0 0' constitutes three binary digits) with Shannon's concept of a *bit* (e.g., the event of a fair coin toss involves 1 *bit* of information, as either/both outcomes eliminate one-half of a set of equiprobable alternatives, that is, heads-or-tails) (p. 354). This is identified as error (h) confusing 'number of binary digits' with 'amount of information, in the statistical sense, due to the fact of *bits* being the *unit* in which information, in the statistical sense, is measured. It does not help matters that Shannon (1948) spoke loosely and metaphorically in claiming that "a device with two stable positions, such as a relay or a flip-flop circuit,"—or, of course a binary digit—"can *store* one bit of information" (p. 379, emphasis added). The general distinction between 'binary digits' and 'bits,' in Shannon's (1948) information-theoretic sense is explored below, and Miller's confusion in this regard is demonstrated to be endemic to his inchoate idea of a 'chunk of information.'

Information, in the Statistical Sense, as *Variance*

It is clear from his 1953 paper that Miller confuses the statistical and the everyday epistemic senses of 'information'. But, beyond recognizing that error, it is difficult to determine just what Miller (1956) believes about the statistical concept of 'information'. He begins this celebrated article by claiming that "'amount of information' is *exactly the same concept* that [psychologists] have talked about for years under the name of 'variance'" (p. 81, emphasis added). This is patently false. As Miller (1956) correctly observes, the average 'amount of 'information'' per message from a particular source is:

$$H(x) = = \Sigma p_i \, (- \log_2 p_i)$$

for which p_i = the probability of the i^{th} message being selected from a set of *i…j* alternatives (Miller, 1953*,* p. 4). 'Variance,' on the other hand, is conventionally defined as

$$\text{Var}(X) = \text{E}[(X-\mu)^2]$$

(for a random variable X that has the expected value (mean) $\mu = \text{E}[X]$).

These concepts are related, but are not identical. 'Variance' is a measure of statistical dispersion of a random variate about its mean. *H*-values pertain to probabilities (the amount of information associated with a message/event is partially determined by the probability of its selection/occurrence) and proportions (of possibilities eliminated by the selection of a particular message and/or the occurrence of a particular event). Of course, a true proposition expressing the average amount of information (in the statistical sense) per message from a particular source could be reasonably described as providing information, in the everyday epistemic sense, *about* the variability of that message source. But it is the *statement as a whole* that could be described as *about* the variability, not the *H*-value that figures in it.

Apparently contradicting his prior claim that 'amount of information', in the statistical sense, and 'variance' are "exactly the same concept," Miller (1956) goes on to describe *differences* between 'amount of information' and 'variance':

> Variance is always stated in terms of the unit of measurement—inches, pounds, volts, etc.—whereas the amount of information is a dimensionless quantity...[so it] enables us to compare results obtained in quite different experimental situations where it would be meaningless to compare variances based on different metrics…Since *the information in a discrete statistical distribution* does not depend upon the unit of measurement, we can extend the concept to situations where we have no metric and we would not ordinarily think of using the variance. (pp. 81-82, emphasis added)

But this too is incorrect. The information statistic, *H*, is *not* a unit-less measure of any property as, for example, a Pearson *r* could be fairly said to be a unit-less expression of the degree of probabilistic co-dependence among phenomena (Miller & Frick's 1949 *index of behavioral stereotypy* would also appear to be unit-less). Far from being unit-less, Shannon's (1948) particular concept of 'information' is *internally related* to the unit of a *bit*. Shannon (1948) arbitrarily stipulated that an event that eliminates one-half of a set of possible alternatives is associated with 1 *bit* of 'information': "The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits…*" (p. 379). This could be set elsewhere (say, at an event that eliminates 10% of the alternatives), *but it has to be set somewhere*. That is, a definition of 'information,' in the statistical sense *must* include an anchor point that establishes the quantitative relation between some proportion of eliminated alternatives, on the one hand, and some unit in which the amount of information will be expressed, on the other.

Miller is clearly incorrect in characterizing the statistical sense of 'information' as logically dissociable from the concept of a metric. Of course, it is dissociable from any *particular* metric, in the same sense that length can be measured in inches, centimetres, paces, or by reference to a sample (e.g., "as long as *this*"). But if one is asked, *how long* some object is, any valid answer must contain reference to a standard. And just as it is impossible to measure length without reference to a standard or unit of some kind, so is it impossible to measure information in the statistical sense without such reference.

On* Information *that Is* in *Things

Of course, there is a unit-less type of information that comes in varying amounts—everyday epistemic information. And this is the sense of 'information' that Miller (1956) unwittingly employs when he writes that "*the information in a discrete statistical distribution* does not depend upon the unit of measurement…" (p. 81, emphasis added). It is clear that it is the everyday epistemic sense of 'information' that Miller has in mind because that is the type of information that can be *in* things. For example, there is geographical information *in* geography textbooks; it is commonplace to say that there is information *in* those things by which epistemic agents can become informed (see Slaney & Maraun, 2005 for more on the container metaphor). When Miller (1956) writes of the "information *in* a discrete statistical distribution" (p. 81, emphasis added), he appears to be referring to, roughly, the knowledge that can be derived from analyzing that distribution. For example, Miller (1956) comments that among the advantages of calculating *H*-values rather than variances is that comparing amounts of information (in the statistical sense) “enables us to compare results obtained in quite different experimental situations where it would be meaningless to compare variances based on different metrics” (p. 82). Presumably, the goal of making such cross-metric comparisons is to learn something, that is, to acquire knowledge.

‘Information’ and ‘Correlation’

Persisting with his already-confusing misidentification of the statistical concept of ‘information’ with ‘variance,’ Miller (1956) goes on to write that “The input and the output [of communication systems] can...be described in terms of their variance (or their information).” But then he proceeds to re-define *H-values* as ‘a measure of correlation.’ For a:

> communication system…there must be some systematic relation between what goes in and what comes out. That is to say, the output will depend upon the input, or will be correlated with the input. If we measure this correlation, then we can say how much of the output variance is attributable to the input and how much is due to random fluctuations or ‘noise’ introduced by the system during transmission. So we see that *the measure of transmitted information is simply a measure of the input-output correlation*. (p. 82, emphasis added)

The careful reader will have become very confused by Miller’s identifying ‘amount of ‘information’’ with both ‘variance,’ and ‘correlation,’ for not only are each of these (mis)identifications erroneous, they are also mutually incompatible. That is, if Miller conceives of ‘information’’ in the statistical sense as more-or-less synonymous with ‘variance,’ then it is not clear how he can also understand it to be more-or-less synonymous with ‘correlation.’ Miller’s imprecision here marks the emergence of what Dretske (1981) identifies as the “spongy, plastic, and amorphous” quality of ‘information’ in the psychological sciences (p. viii).

Research Participants as Communication Channels

After this thoroughly confusing, supposedly general introduction to the statistical concept of ‘information,’ Miller describes certain of its paradigmatic applications in the psychology of the early 1950s. The experiments to which Miller refers concern human research participants' capacities to form absolute judgements regarding one and/or two stimulus dimensions. For example, Pollack (1952) investigated participants' ability to discriminate distinct tones, Garner (1953) investigated participants' abilities to discriminate different loudness levels of the same tone, and Beebe-Center, Rogers, and O'Connell (1955) investigated participants' skill at discriminating among different concentrations of a salt solution. With respect to absolute judgments regarding two dimensions, Klemmer and Frick (1953) investigated participants' abilities to distinguish among different positions of a dot inside a square figure, and Pollack (1953b) asked participants to distinguish among a set of pure tones that differed in both loudness and pitch. It would be unfair to suppose that Miller (1956) should be regarded as speaking for all of the individual psychologists whose work he reviews. However, because Miller's influence is so widely recognized, the manner in which he (mis)construes these findings is especially worthy of investigation:

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

Careful examination of this passage reveals a variety of logical problems. Their most glaring grammatical manifestations include the presence of the prepositions 'in' ("the amount of information *in* the stimuli" and "the amount of information *in* [participants'] responses," p. 82) and 'about' ("…the *channel capacity…*represents the greatest amount of information that [the participant] can give us *about* the stimulus…" p. 82) indicates that the passage involves equivocating between the statistical and everyday, epistemic senses of 'information.' There is no such thing as N *bits* of information *in A* or *B* and/or *about X* or *Y*. The type of information that can be both *in* things (e.g., information in a library) and about things (e.g., information about libraries) is everyday epistemic information. Accordingly, Miller confuses the *already-confusing* concept of 'transmitting information' (because it is signals and not information, in the statistical sense, that is transmitted within Shannonian communication systems; see Chapter 2) with, roughly, 'knowing.' For it is experimental participants' correct responses that *count as* their having 'transmitted the information in the stimulus' to the experimenter in the form of their responses.

Epistemic Agents versus Mere Media

Miller (1956) writes that "In the experiments on absolute judgment, the observer is considered to be a communication channel" (p. 82). This is a radical departure from Shannon's (1948) model, for which he defined 'communication channel' as "*merely the medium* used to transmit the signal from transmitter to receiver. It may be a pair of wires, a coaxial cable, a band of radio frequencies, a beam of light, etc." (p. 380, emphasis added). It is clear that the participant-as-communication-channel is not regarded as a *mere* medium. On the contrary, it is (certain of) the participants' perceptual and epistemic capacities that constitute the psychologist's objects of investigation. So there are important differences between the role that 'channel capacity' plays in Shannon's engineering on the one hand, and in Miller's psychology, on the other. For example, in an engineering context, 'channel capacity' refers to a more-or-less stable property of a particular communication channel. It would be absurd to suppose that a coaxial cable might increase its channel capacity through diligent study habits. This alone does *not* establish that it is nonsensical to construe a human research participant as a communication channel. However, it demonstrates that in doing so, the channel-capacity-measuring psychologist is implicitly characterizing research participants as bereft of the very epistemic potentialities (to learn, to become confused, etc.) that constitute the objects of investigation. Miller (1956) writes that "confusions will appear near the point that we are calling [the participant's] 'channel capacity'" (p. 83). And although it is *impossible* for “a pair of wires, a coaxial cable, a band of radio frequencies, [or] a beam of light” to become confused, a research participant *must* become confused if the psychologist is to measure his channel-capacity (Shannon, 1948, p 380).

Furthermore, there is individual variation among participant-communication-channels of a kind that is not found among, say, coaxial cables. For example, Miller (1956) observes that “there is evidence that a musically sophisticated person with absolute pitch can identify accurately any one of 50 or 60 different pitches” (p. 84). Not only are certain individuals unusually talented in this domain, but improving pitch discrimination abilities in the non-naturally-talented is a routine aspect of musical education. But of course, a, say, beam of light cannot learn to beam better.

Channel Capacities: Descriptive or Explanatory*?*

With engaging candor, Miller writes:

> Fortunately, I do not have time to discuss these remarkable exceptions. I say it is fortunate because I do not know how to *explain* their superior performance. So I shall stick to the more pedestrian fact that most of us can identify about one out of only five or six pitches before we begin to get confused. (p. 84, emphasis added)[44]

If we follow Miller’s lead and restrict focus to the non-exceptional cases, does it make sense to say that the channel-capacity of a particular participant *explains* his performance in, say, discriminating different pitches? It does not, at least, not in the way that determining channel capacity is explanatory in the context of an engineer’s investigation. In the context of investigating the properties of a particular electronic communication system, calculating the transmission capacity of a particular channel is an aspect of developing an analytic and/or functional understanding of that communication system, that is, an understanding of the overall system’s functioning in

[44] However, this is overgeneralized. In Pollack’s (1952) research “The tones were different with respect to frequency, and covered the range from 100 to 8000 cps in equal logarithmic steps” (Miller, 1956, p. 83). Restricting the range to 100 to 101 cps would be expected to affect participants’ respective channel capacities, that is, their ability to identify discrete pitches.

terms of the interrelated behaviors of component parts.[45] The communications engineer wishes to measure the capacity of a particular channel to transmit reliably distinguishable signals. The capacity of a particular channel (CCh) to *more or less quickly* transmit reliably distinguishable signals is *one* among the various factors that the communication engineer takes into account in designing a system (see Figure 2.1). Accordingly, in the case of an electronic communication system, channel-capacity can be considered a sub-capacity of the system's overall capacity to communicate (CCo), that is, reproduce series of selections from a set of alternatives. Capacities CCo and CCh, respectively, are logically distinct; the transmission capacity of a particular channel (CCc) is *not* identical with the capacity (CCo) of the communication system (of which it is a part) to communicate. CCh is a *sub-capacity* of CCo. CCo and CCh are also empirically distinguishable, for example, CCo is also affected by the efficiency of a communication system's encoding/decoding capacities. If reliably distinguishable signals are transmitted more rapidly than they can be accurately decoded into message form, then CCo suffers, but not for a deficit in sub-capacity CCh. Determining the transmission capacity of a particular channel, then, can be reasonably viewed as (partially) *explaining* the communication capacity of the system as a whole.

In contrast, measuring the channel capacities of research-participants-as-communication-channels (CChRP) cannot be reasonably regarded as explaining those participants' absolute judgment capacities (CAJ) to reliably distinguish different stimuli-as-signals, *because the only way to calculate the channel capacity for a particular participant is to record his capacity to reliably distinguish different stimuli-as-signals*! However, CChRP and CAJ, respectively, are not quite identical; an amount of *bits* is not the same as a count of number of items, although, as is demonstrated below, Miller was confused about this. Nonetheless, although CChRP and CAJ are logically non-identical, they are *internally related* such that they are *empirically indistinguishable*. Measuring CChRP consists in nothing but performing a mathematical operation on a particular value for CAJ. Miller (1956) makes this clear in his explanation of "data…on the amount

[45] To be clear, this is *not* to characterize a system's functioning as a material or physical understanding.

of information that is transmitted by listeners who make absolute judgments of auditory pitch" (p. 83). In particular, findings that "as the amount of input information is increased by increasing from 2 to 14 the number of different pitches to be judged, the amount of transmitted information approaches as its upper limit a channel capacity of about 2.5 *bits per judgment*" (p. 83, emphasis added). Miller subsequently comments:

> So now we have the number 2.5 bits. What does it mean? First, note that 2.5 bits corresponds to about six equally likely alternatives. The result means that we cannot pick more than six different pitches that the listener will never confuse. Or, stated slightly differently, no matter how many alternative tones we ask him to judge, the best we can expect him to do is to assign them to about six different classes without error. Or, again, if we know that there were *N* alternative stimuli, then his judgment enables us to narrow down the particular stimulus to one out of *N/6.* (p. 84)[46]

Given that the measurement of CChRP presupposes that of CAJ, there are no grounds for concluding, as Miller (1956) does, that "the span of….absolute judgment is *limited by the amount of information*" that research-participants-as-communication-channels can transmit (p. 92, emphasis added). It does not make sense to claim that CAJ is *limited* by CChRP, unless it also makes sense to say that, say, a man's height is limited by the $\log_2$ of the number of inches tall that he stands. It is clear that regarding CChRP as *explaining* CAJ amounts to something in the genre of an aesthetic preference for CChRP over CAJ, once again exemplifying Dretske's (1981) observation that information-theoretic concepts have "a tough, brittle, technical sound" while remaining "spongy, plastic, and amorphous enough to be serviceable in cognitive and semantic studies" (p. viii).

In contrast, with respect to investigating the properties of an electronic communication system, it does make sense to conceive of the transmission capacity of a

46 Consider the example of a fair die toss, as above. The probability of rolling, say, a six is approximately $p \approx 0.17$. To calculate the amount of 'information', in the statistical sense, associated with this reduction of six equiprobable alternatives to one: $H_i = -\log_2 p_i$; setting p_i to 0.17, we get $\log_2(0.17) \approx \log_2(1/0.17) \approx \log_2(5.9) \approx 2.6$ *bits* of 'information', in the statistical sense, associated with the reduction of six equiprobabale alternatives to one.

communication channel CCh as placing a limitation on the capacity of the system to communicate (CCo); in particular, CCh indicates the *rate* at which reliably distinguishable signals can be transmitted, expressed as an amount of information, that is, number of *bits*, per unit of time.[47]

Per Unit Time vs per Unit Judgement

Another important difference between CCh and CChRP is that CCh is a rate. Clumsily blurring the distinction between symbols and signals that is essential to his own model, Shannon (1948) clarifies "If the system *transmits n symbols per second* it is natural to say that the channel has a capacity of 5*n* bits per second (Shannon, 1948, p. 382, emphasis added). Strictly speaking, however, it is signals and not symbols that are transmitted. To suppose otherwise is identified as error (g) above: failing to observe the distinction between symbols and signals that is essential to Shannon's model. However, what is relevant here is that a particular CCh is an amount-of-information (in the statistical sense) *per unit of time*. In contrast, a particular CChRP is an amount of 'information' (in the statistical sense) per...what? "As the amount of input information is increased by increasing from 2 to 14 the number of different pitches to be judged, the amount of transmitted information approaches as its upper limit a channel capacity of about 2.5 *bits per judgment*" (p. 92, emphasis added). Is CChRP, then, a rate? This depends on whether or not a judgement can be considered a duration. To be fair, we should restrict focus to judgements of the type that are relevant to the research that Miller summarizes and not, for example, judgements that, say, one vegetable is more delicious than another. In the context of absolute judgment experiments, each particular judgment that a participant makes *has a duration*, typically referred to as 'response time.' So if a judgment of this type *is* a duration, then it is a duration that also *has* a duration, which makes as much sense as saying that every height has a height, that is, not much

[47] Interpretation of Miller's argument is further complicated by the fact that he switches from discussing investigations of the respective abilities of research-participants-as-communication-channels to *transmit* information to a discussion of "limitations that are imposed on our ability to *process information* (p. 92, emphasis added). The concept of 'processing information' is introduced without explanation. In Chapter 4, Sloman's (2011)

sense at all. It is preferable to say that particular values of CChRP express a relation between a number of *bits* and *an event*, rather than between a number of *bits* and a *duration*. There is nothing unusual about relating a quantities and event in this manner, so it is easy to identify cases analogous to CChRP in this respect, for example, that a particular nation spent, on average, *N* dollars per war in the 20th Century.

To be clear, the above criticisms pertain not to the value of the empirical findings that Miller discusses but rather to the coherence of psychological researchers' Procrustean efforts to construe their investigations in information-theoretic terms. Baddeley (1994) offers a familiar version of this history when he credits Miller's *Magical Number…* with "demonstrating the need to go beyond information measures…" (p. 353). Having demonstrated above that psychologists such as Miller (1953) grossly distorted the statistical concept of 'information' in applying it to psychological investigations, the manner in which Miller (1956) putatively went "beyond information measures" is investigated below. Particular cases of errors (following (a) –(g) described above) that will be encountered are:

(h) confusing the event of a particular random sequence with a particular sequence of random events;

(i) confusing 'number of binary digits' with 'amount of 'information', in the statistical sense' due to the fact of *bits* being the *units* in which information, in the statistical sense, is measured.

(j) confusing a language with a code, and thereby

(k) mistaking (certain cases of) thinking for recoding.

From Bits to Chunks

The *Magic Number...* (1956) paper is widely celebrated for Miller's novel concept of a 'chunk' which he contrasts with that of a 'bit.' Baddeley (1994) especially praises Miller's paper for:

claim that the now-ubiquitous concept of 'processing information' may be *implicitly defined* is investigated.

> ...demonstrat[ing] the crucial difference between the limitations on span and on absolute judgment, with judgment being limited by the amount *of information,* measurable in bits, whereas immediate memory span is determined by the *number* of items, or to be more accurate, the number *of chunks*. (p. 354)

However, the arguments below demonstrate that Miller's concept of a 'chunk' presupposes confusion between the everyday epistemic and statistical senses, of 'information'. Miller (1956) jokes that "[his] problem is that [he has] been persecuted by an integer" (p. 81). In what follows, it will become apparent that he is also persecuted by a word, information.

More- and Less-valid Extensions of Shannon's (1948) Information Theory

Although both involve extending the statistical concept of 'information' beyond its intended range of application, there are important differences between psychologists' information-theoretic investigations of absolute judgment on the one hand, and Miller's *pseudo*-information-theoretic gloss on experimental results regarding immediate memory capacity on the other. In construing research participants as communication channels, the number of items among which research participants can reliably distinguish *per response* (or *per judgment*) is compared to the number of distinguishable signals that a communication channel can transmit *per unit of time*. The differences between these two uses of 'channel capacity' are discussed above. Despite these differences, the analogy between communication channels and research participants' absolute judgement capacities remains comprehensible. Somewhat crudely, distinguishing among different stimuli values (e.g., different pitches and/or loudness levels of auditory tones, different salt concentrations, etc.) is enough like distinguishing among alternative signals as to make the comparison credible, if imperfect. In contrast, the immediate memory experiments that Miller reviews do not resemble Shannonian communication activity in any relevant respect. Therefore, there is no substantive continuity whatsoever between Shannonian information-theoretic analyses on the one hand, and, on the other, Miller's (1956) information-processing psychology.

Consider Miller's (1956) stated motivation to supplement *bits* with a new unit, *chunks*. He begins by observing that:

> There is a clear and definite limit to the accuracy with which we can identify absolutely the magnitude of a unidimensional stimulus variable. I would propose to call this limit the *span of absolute judgment,* and I maintain that for unidimensional judgments this span is usually somewhere in the neighborhood of seven. (p. 90)

He then remarks that "everybody knows that there is a finite span of immediate memory and that for a lot of different kinds of test materials this span is about seven items in length." Putting 3.5 and 3.5 together, Miller (1956) reasons that:

> if immediate memory is like absolute judgment, then it should follow that the invariant feature in the span of immediate memory is also *the amount of information that an observer can retain*. If the amount of information *in the span of immediate memory* is a constant, then the span should be short when the individual *items contain a lot of information* and the span should be long when the items *contain little information*.
> (p. 91, emphasis added)

Miller (1956) argues that if this were true, "then we should be able to remember only two or three words chosen at random" (p. 91). However, as "everybody knows that…this span is about seven items in length" it cannot be the amount of information, as measured by the *number of bits* that determines the limits of our immediate memory capacity (p. 91). Rather, Miller concludes, it must be a *different* amount of information that limits the span of immediate memory:

> In order to capture this distinction in somewhat picturesque terms, I have fallen into the custom of distinguishing between *bits* of information and *chunks* of information. Then I can say that the number of bits of information is constant for absolute judgment and the number of chunks of information is constant for immediate memory. The span of immediate memory seems to be almost independent of the number of bits per chunk, at least over the range that has been examined to date.
> (pp. 92-93)

‘Chunking’ Clarified

Miller (1956) speculates that "the process of memorization may be simply the formation of chunks, or groups of items that go together, until there are few enough chunks so that we can recall all the items" (p. 94). So, ‘chunking’ essentially means ‘grouping,’ for example, construing a sequence of 14 decimal digits as two 7-digit phone numbers. Candidly, Miller (1956) admits that he is “not very definite about what constitutes a chunk of information” (p. 93). It will now be demonstrated that this confusion regarding what a chunk is results from a logically prior confusion regarding ‘information’ and ‘*bit’* in Shannon’s (1948) statistical senses. First, consider that, as observed above, it is only the everyday epistemic sense of ‘information’ that can serve as the direct object of the verbs ‘contain’ and/or ‘retain’ (e.g., this geography textbook *contains* geographical information that must be *retained* by students if they are to pass their geography tests). So, it is clear that the premise of Miller’s (1956) *chunk* argument—that “If the amount of information *in the span of immediate memory* is a constant, then the span should be short when the individual *items contain a lot of ‘information’* and the span should be long when the items *contain little information*”—suffers from serious equivocation errors (p. 91, emphasis added). And this grammatical criticism is not reducible to a mere terminological preference. Rather, it betrays fundamental misunderstandings of the statistical concept of ‘information.’ Consider how Miller (1956) introduces the (pseudo) information-theoretic investigation of immediate memory.

> Up to this point we have presented a single stimulus and asked the observer to name it immediately thereafter. We can extend this procedure by requiring the observer to withhold his response until we have given him several stimuli in succession. At the end of the sequence of stimuli he then makes his response. We still have the *same sort of input-output situation* that is required for the measurement of transmitted information. (p. 91, emphasis added)

In this passage, Shannon’s (1948) statistical concept of ‘amount of ‘information’’ is distorted beyond all recognition. Miller (1956) writes that experimental investigations of immediate memory involve “the same sort of input-output situation that is required for the measurement of transmitted information" (p. 92) but offers no justification for this claim, which crumbles when subjected to critical scrutiny.

Whereas in information-theoretic investigations of absolute judgment, the amount of information that a research participant is considered to transmit is internally related to (and empirically indistinguishable from) the number of items among which he can reliably distinguish, Miller (1956) now introduces the idea that there are varying amounts of information *per item*. In supposing that different items and/or stimuli involve varying amounts of information, in the statistical (*H*) sense, Miller very subtly, but consequentially, changes the basis of comparison between research participants and Shannonian communication systems. In particular, the analogy is no longer between the research participant and a communication channel (or any other component of a Shannonian communication system) but between the research participant and *the communication engineer*. That is, in supposing that the items presented to immediate-memory-research participants (e.g., a sequence of random words, digits, or phonemes) involve amounts of information (i.e., values of *H*), Miller implicitly compares said items to the amounts of information associated with different messages in a Shannonian communication system.

Superficially, this comparison appears acceptable. After all, provided that the to-be-recalled items that an experimenter presents consist in symbols and/or sequences of symbols (e.g., 7, T, wug, apple), it is easy to imagine a Shannonian communication system for which such items are among the possible messages associated with transmittable signal sequences. The problem with this argument is that that amount of information (*H*) associated *with* a particular message is *not* a property of that message per se, but rather, is a property of that message *in the context of a particular Shannonian communication system.* And it is the communication engineer who knows the statistical structure of the signal sequences that his communication system is anticipated to transmit. So, whereas 'channel capacity' is a property of a communication system, the 'amount of information per message' is a property of what is more accurately described as 'Shannonian *communication*' or 'information-theoretic *communication*'.

Information-theoretic Analyses of Natural Languages

If a sample of verbal communication (broadly construed, including conversation and writing), is regarded as Shannonian, information-theoretic communication (i.e., as involving selections from a set of alternative messages), then one can investigate the statistical properties of that particular sample of verbal communication by computing values for the amount of information associated with different alphanumeric characters, words, phrases, etcetera. For example, Shannon (1948) writes:

> artificial languages are useful in constructing simple problems and examples to illustrate various possibilities. We can also approximate to a natural language by means of a series of simple artificial languages. The zero-order approximation is obtained by choosing all letters with the same probability and independently [*sic*]. The first-order approximation is obtained by choosing successive letters independently but each letter having the same probability that it has in the natural language. Thus, in the first-order approximation to English, E is chosen with probability .12 (its frequency in normal English) and W with probability .02, but there is no influence between adjacent letters and no tendency to form the preferred digrams such as TH, ED, etc. In the second-order approximation, digram structure is introduced. After a letter is chosen, the next one is chosen in accordance with the frequencies with which the various letters follow the first one. This requires a table of digram frequencies $p_i(j)$. In the third-order approximation, trigram structure is introduced. Each letter is chosen with probabilities which depend on the preceding two letters. (pp. 385-386)

So, the communication engineer develops an information-theoretic analysis of natural language for a particular purpose. First, it is assumed that the frequency of the signal-sequences selected for transmission in the system that he designs will bear some degree of resemblance to the frequency with which the verbal messages associated with those signal sequences were found to appear in a sample of verbal communication. Accordingly, the engineer can design a communication system to exploit the fact that, for example, THE may be a more probable message than QWE. Shannon (1948) makes this perfectly clear:

> In order to obtain the maximum power transfer from a generator to a load, a transformer must in general be introduced so that the generator as seen from the load has the load resistance. The situation here is

> roughly analogous. The transducer which does the encoding should match the source to the channel in a statistical sense. (pp. 396-397)

How does the engineer consider statistical influences? Shannon (1948) cites Pratt's 1939 estimation of letter, digram (e.g., TH), and trigram (e.g., SHO) frequency and Dewey's 1923 estimations of word frequency. By citing these authors, it should be understood that Shannon's (1948) purpose is illustrative and didactic; these estimates are used to demonstrate how knowledge regarding the statistical properties of messages sources can be used to inform the design and/or otherwise evaluate the performance of electronic communication systems that transmit signal sequences:

> The main point at issue is the effect of statistical knowledge about the source in reducing the required capacity of the channel, by the use of [optimal] encoding....In telegraphy, for example, the messages to be transmitted consist of sequences of letters. These sequences, however, are not completely random. In general, they form sentences and have the statistical structure of, say, English. The letter E occurs more frequently than Q, the sequence TH more frequently than XP, etc. The existence of this structure allows one to make a saving in time (or channel capacity) by properly encoding the message sequences into signal sequences. This is already done to a limited extent in telegraphy by using the shortest channel symbol, a dot, for the most common English letter E; while the infrequent letters, Q, X, Z are represented by longer sequences of dots and dashes. (pp. 383-384)

Ritchie (1986) elaborates:

> For the *engineer's* purpose, the English language can be viewed as a generalized transmission source. A description of the statistical properties of grammatical English provides a useful heuristic for developing and explaining the function of redundancy in a signal transmission system. It also provides a basis for specifying maximum signal transmission rates for given levels of noise or, conversely, for specifying the maximum level of noise allowable for a given desired transmission rate. (p. 281)

So, the communication engineer computes the *H*-values that are associated with various symbols and/or symbol sequences for a particular, heuristic purpose. The engineer is not discovering the information-theoretic properties of English-language communication *in general*. Rather, a sample of verbal communication is construed in information-theoretic terms as part of an effort to "consider the influence of the statistics of [each possible] message" as "the one which will actually be chosen…is unknown at the time of design" (Shannon, 1948, p. 383). An analogy helps to clarify this distinction: any sample of spoken or written language of the form five syllables/seven syllables/five syllables *can be viewed as* a haiku, for example, *Many samples of/language make a haiku/if you see them so.* However, it would be absurd to suppose that it is possible to study haiku by analyzing *any* sample of language that takes this form. Similarly, it is absurd to suppose that any sample of verbal communication possesses the information-theoretic properties that the sample would have *if* that sample constituted some portion of possible messages in a Shannonian communication system.

Apparently, however, Miller (1956) wrongly regards information-theoretic analyses of samples of verbal communication as revealing intrinsic properties of verbal communication. This is identified as error (b) above: confusing Shannon's (1948) stipulations regarding his model of communication with natural laws of communication. For example, Miller (1956) writes that "isolated English words are worth about 10 bits apiece" and that "decimal digits are worth 3.3 bits apiece," and (p. 91). Miller does not provide the calculations that would justify such claims, but it is easy to detect certain of his assumptions. For example, if 0…9 constitute 10 equiprobable alternative messages, then the probability (*p)* of selecting each of 0…9 is 0.1. The amount of 'information', in the statistical sense (H), associated with a *particular* message/event (i) is defined as:

> $H_i = -\log_2 p_i$, for which p represents the probability of said message's occurrence.
> Substituting 0.1 for p, we get $H = -\log_2 0.1$.
> We eliminate the negative sign to yield $H = \log_2 (1/0.1)$.
> This reduces to: $H = \log_2 (10) \approx 3.3$ *bits*.

So Miller's (1956) claim that "decimal digits are worth 3.3 bits apiece" is true, provided that the 0…9 constitute 10 equiprobable alternative messages (p. 91). For it is only because the probability of each 0…9 being transmitted is known to equal $p = 0.1$ that we were able to calculate an amount of bits. But if, *and only if,* it makes sense to construe 0…9 as 10 equally-likely alternative messages can we attach the idea of *any* amounts of *bits* to these integers. Now, it is easy to imagine situations that meet the requirements for an information-theoretic analysis. For example, if A's job is to utter a random decimal digit and B's job is to correctly record which digit A named, then each of A's utterances is associated with approximately 3.3 *bits* of information. That is, A-and-B's situation is enough like a Shannonian communication system to make the application of 'amount of information,' in the statistical (*H*) sense, comprehensible. The probability of the signal, loosely speaking (that is, the acoustic energy associated with the utterance of each of 0…9), being transmitted is known. Also essential is that A-and-B's communication can be construed as a matching of messages selected from a predefined set of alternatives, with 'communication' being defined as a matching of the decimal digit that A states aloud at T_1 with the decimal digit that B records at T_2. Of course, an information-theoretic description of A-and-B's game constitutes one among many coherent descriptions of A-and-B's game. For example, their game could be understood as an informal test of B's capacity for sustained attention. And there are no amounts of bits associated with A's utterances under this description, because '*bits*' only pertain to the information-theoretic aspects of A-and-B's game. That is, there is no justification for thinking that, in B's *attending* to which decimal digit A names, there are any amounts of information, in the statistical sense, (i.e., amounts of *bits*) involved. Shannon (1948) clarified that the "semantic aspects are irrelevant to the engineering problem," with which his information theory is concerned (p. 378). Similarly, the psychometric and/or attentional aspects of A-and-B's game are also irrelevant to that same game's information-theoretic description.

Imagine that A and B grow tired of the digit-recording game and start playing a memory game in which A is to read aloud successively longer sequences of digits, each of which he asks B to recall before moving on to the next sequence. In this case too there is no justification for applying Shannon's (1948) concept of *bits* to the read-aloud-and-recalled digits. For the short-term memory-testing aspects of this situation are "irrelevant to the engineering problem" with which information-theoretic concepts are concerned (p. 379). Recall that unit information, 1 *bit*, is defined as the amount of information, in the statistical sense, associated with a message, and/or event, the occurrence of which eliminates one-half of a predefined set of alternatives. In the case of A-and-B's digit-recording game, B's hearing A utter, say, *three* can be reasonably described as A's eliminating the possibilities of A having said *one, two, four, five…*and so on. So there is a logical relationship between A-and-B's digit-recording game and Shannon's (1948) information-theoretic concept of a '*bit*.' But there is no relationship between the length of the longest digit sequence that a person can recall and that concept of a 'bit.' Predicating amounts of such *bits* to A-and-B's memory-game responses —for example, Miller's (1956) claim that "we can recall about seven [decimal digits] for a total of 23 bits of 'information'"—is unjustifiable (p. 91). The 3.3 *bits* associated with each decimal digit *in the context of A-and-B's prediction game* do not apply to the appearance of decimal digits in other contexts, for example, A-B's memory game, and/or, say, A writing down B's phone number while B recites it aloud. Picturesquely, it is not as if there are 3.3 *bits* of something *attached* to or *built in* to the symbols 0…9.

To be precise, it is not *false* to claim that a person who correctly recalls a sequence of seven decimal digits thereby recalls 23 bits of information, in the statistical sense of 'bits of information.' Rather, it is *nonsensical*, for there is just no such thing as 'recalling *N* bits of information' because amounts of information, *in the statistical sense*, cannot serve as the direct object of the verb 'recall' (Bennett & Hacker, 2003). However, the un-recall-ability of information, in the statistical sense, is easily overlooked because the concept of 'recalling information' (in the everyday epistemic sense) is so familiar. For example, it is common to recall information *that* such-and-such is the case, and/or to recall information *about* one or another topic.

For the purposes of illustration, let us consider carefully the case of a phone number and liberally characterize A's reciting his phone number aloud as his transmitting a message to B. First, A-and-B's situation as a whole must conform to—or at least resemble—a Shannonian communication system if any amount of information, in the statistical sense, can be predicated to A's utterance. Most fundamentally, for Shannon's (1948) restricted sense of 'communication' each "message is one *selected from a set* of possible messages…" (p. 379). Accordingly, for an information-theoretic analysis A's and B's situation, A's utterance of, say, 8675309, must be evaluated against the background of the alternative messages that could have been selected.[48] But what would constitute the set of possible messages from which 8675309 was selected? The set of all possible utterances that A can produce would appear to be infinite, and there is no such thing as a proportion of infinity, so it does not appear that 'all possible utterances' constitutes a 'set of possible messages' with respect to information-theoretic concerns. Perhaps the relevant set is that of all possible phone numbers within a given area code. Alternatively, we might consider the set of all possible seven decimal digit sequences to constitute the relevant set of alternative utterances. If the probability of each of the 7! = 5040 possible sequences are known, then it is possible to compute the amount of information, in the statistical sense, that is associated with each particular sequence. For example, if all sequences are equiprobable, then the amount of information (*H*), in the statistical sense, that is associated with each particular sequence (H_i) is computed as follows:

$H_i = -\log_2 p_i$; setting $p_i = 1/5040 \approx 0.0002$;

$H_i = -\log_2(0.0002)$

$H_i = \log_2(1/.0002)$

$H_i = \log_2(5000) \approx 12.3$ *bits*

But is this the right way to attach a probability to A's uttering his phone number? If B is entirely ignorant of A's phone number, but A knows his own phone number by heart, how are we to estimate the probability of A's uttering *seven* after saying, *eight, six*, given that his phone number is 867-5309?

Miller (1956) claims that "isolated English words are worth about 10 bits apiece" (p. 91). However, (messages corresponding to) particular words are "worth" different amounts of information, in the statistical sense, in different contexts. Furthermore, it is far

[48] This example involves some loose speaking in that (1) it is not clear that an utterance can be reasonably characterized as a 'symbol sequence', although an utterance can be *represented* as a symbol sequence by expressing what was uttered in written form, and (2) it is not clear that every utterance can be reasonably construed as a message in the Shannonian sense of being selected from a set of possible messages.

from clear that estimations of particular word and/or letter frequencies in spoken and/or written English (or any other language) constitute a valid premise upon which to base an information-theoretic analysis of, say, English language communication *in general*. For, in very many contexts, such as that of A reciting his telephone number to B, there is no clear way to attach any *H*-value to A's utterance. Shannon's (1948) definition of 'message' in the information-theoretic sense is that which "selected from a set of possible messages" (p. 379). However, competent language speakers are capable of producing an infinite variety of utterances (e.g., Chomsky, 1957), and there is no such thing as a proportion of infinity. But, of course, as so impressed Miller (1951), linguistic behavior is predictable to some degree. However, there are myriad ways to assign probabilities to various utterances, and the amounts of information, in the statistical sense, associated with the messages-qua-utterances are arbitrary artefacts of the probabilities that the information-theorist-linguist *assigns* to the utterances, not intrinsic properties of the utterances or the utterings.

The Event of a Sequence versus a Sequence of Events

It appears that Miller (1956) makes a very basic error in his putatively information-theoretic analyses of studies of immediate recall. After justifying this criticism, a reasonable, but nonetheless misleading objection to it will be considered.

It is demonstrated above that there are very many cases of communication that do not conform to (or even resemble) communication in the restricted, information-theoretic sense, for example, the very ordinary event of one person reciting a seven-digit phone number to another. However, the studies of immediate recall that Miller (1956) reviews involve *random* sequences of decimal digits. Therefore, it *is* possible to determine the probability of any particular sequence being uttered. The amount of information (*H*), in the statistical sense, associated with the utterance of a 7-item sequence of random decimal digits is a function of the probability of *that particular sequence* being uttered, as well as of the number of alternative utterances that are eliminated by the utterance-qua-selection. That *H* value is calculated above and yields the value of approximately 12.3 *bits*. So, in a context for which the decimal digits 0...9 constitute an exhaustive set of equally-likely alternatives, Miller's simple multiplication of

3.3 (the *H*-value associated with each individual digit) by seven (the number of symbols in the sequence) to conclude that a 7-item random sequence of decimal digits involves "about 23 *bits* of information" appears to be invalid. The error involved, identified as error (h) above, is that of failing to distinguish between *the event of a sequence* and a *sequence of events*. More precisely, the neglected distinction is that between *the event of a particular random sequence* and *a sequence of particular random events*. That is, on the one hand there is the occurrence of a particular random sequence of seven equiprobable decimal digits, associated with the approximately 12.3 *bits*. On the other hand, there is a sequence of seven discrete trials of equiprobable selections from the set of 0…9, with which 23 *bits* of information, in the statistical sense, would be associated, if Miller's (1956) additive attitude towards the discrete amounts of information (*H*) associated with discrete events is regarded as legitimate. Miller's multiplication of 3.3 *bits* by seven items might be lampooned by comparing it to claiming that the probability of observing heads-heads-tails in a series of three fair coin tosses is 1.5, as it would involve three consecutive events whose probability *as discrete trials* equals 0.5. Of course, the error is easier to spot in the latter case, as '$p = 1.5$' is a nonsense; there is no such thing as a *p*-value greater than 1, whereas there is such a thing as 'approximately 23 *bits* of information.'

However, the validity of this criticism hinges upon whether the utterance of a 7-item random sequence of decimal digits, is considered to constitute *one, particular 7-itemed message* or, on the other hand, if such a sequence is considered as a *series of seven, particular messages.* Given that what experimental participants are instructed to recall in the studies that Miller reviews is *a particular, N-itemed sequence*, there does not appear to be a clear justification for construing such sequences as *N*-itemed *series* of *particular* messages. However, it is also easy to appreciate why Miller might suppose that the whole of such sequences is identical with the sums of their item-parts. For it is clear that random sequences of decimal digits *do not* involve the types of serial dependencies that so interested Miller and Frick (1949) and Shannon (1948). That is, quite unlike successive responses in a maze-running trial and/or the successive alphanumeric characters of written English (for example) language messages, successive items of an *N*-itemed random sequence of decimal digits constitute discrete trials. Nevertheless, even for cases in which Ds constitute discrete trials, the probability of the sequence D_1, D_2, D_3, D_4 [p (D_1, D_2, D_3, D_4)] is not equal to the sum of the individual probabilities of $D_1 \ldots D_4$.

What constitutes the relevant selection from a set of specified alternatives? Is it *the sequence* of decimal digits that the experimenter actually utters? Or is it also possible to coherently consider *each particular item* that the experimenter utters to constitute the relevant selection? If each successive item that is uttered constitutes a particular selection of a random decimal, Miller's claim that each item involves 3.3 *bits* of information, in the statistical sense, is correct. There does not seem to be a definitive distinction between right and wrong here. Nevertheless, it is essential to distinguish between a *particular sequence of random values* on the one hand, and a *series of particular random values*, on the other.

On Containing Bits of Information

Miller highlights that (persons') short-term memory capacity is partially dependent upon the manner in which the recalling party organizes the to-be-recalled items (e.g., it is easier to remember three, seven-digit phone numbers than 21 individual digits). I take it as self-evident that this constitutes an important fact about persons' short-term memory capacities. However, Miller's (1956) effort to construe such findings in information-theoretic terms inadvertently introduces conceptual confusion. He writes that "since the memory span is a fixed number of chunks, we can *increase the number of bits of information* that it contains simply by building larger and larger chunks" (p. 93, emphasis added). However, there is no such thing as a memory span containing an amount of *bits* of information, in the statistical sense. Values of Shannon's (1948) *H* are akin to values of a Pearson *r* and/or a particular variance value; that is, they are values of a *statistic*. However, what is recalled in the experiments that Miller (1956) summarizes are sequences of verbally presented items, in particular, random sequences of decimal digits. And the amount of information, in the statistical sense, that is associated with a 7-itemed random sequence of decimal digits —whether that amount is held to be 12.3 or 23 *bits*—is not what is recalled in the event of a participant successfully recalling the administered sequence. It is the sequence of decimal digits that is recalled. A sequence of items, *S,* is non-identical with the amount of information that is associated with that sequence under a particular information-theoretic description (H_s).

The arguments above may be difficult to accept, in part because a span of immediate memory may very frequently involve 'bits of information,' *in the everyday epistemic sense*, for example, the items on a shopping list that a shopper rehearses to himself en route the store. Immediate memory spans might even appear to intrinsically involve information, in the everyday epistemic sense, *about* what is recalled. For example, an experimental participant who correctly recalls a sequence of seven random decimal digits can be coherently described as recalling information (in the everyday epistemic sense) *about* which sequence of decimal digits the experimenter had previously uttered. So, predicating amounts of information, in the everyday sense, to memory-related phenomena is very familiar. These habits of speech make it difficult to detect the incoherence of predicating *bits* of information, in the statistical sense, to an immediate memory span, *especially* when certain aspects of the immediate recall study can be described in information-theoretic terms.

It does not help that there is a rich history of imprecision here. For example, Shannon (1948) claimed that "a device with two stable positions, such as a relay or a flip-flop circuit, can *store* one bit of information" (p. 379, emphasis added). But, more precisely, it is the event of a flip-flop circuit being in *one of two equiprobable* possible positions, that is associated with 1 *bit* of information, in the statistical sense. For if it is more probable that the flip-flop circuit is in position_1 than in position_2, then there is less information, in the statistical sense, associated with its being in the more probable position_1 than there is associated with it being in the less probable position_2. The concept of "stor[ing] one bit of 'information'" (p. 379) is metaphorical. A critic might observe that this is akin to claiming that peanut butter stores its correlation with jam. Of course, the storage metaphor applies more naturally to the case of the flip-flop circuit, perhaps because it is so common to speak of storing information, in the everyday epistemic sense. Miller's (1956) intuitively appealing predication of *bits* to memory spans is also metaphorical, but the comparison is invalid, because there is no discernable relationship between immediate-recall capacity and the statistical concept of a '*bit*.' That is, it is coherent to construe $position_1$ and $position_2$ of a flip-flop circuit as jointly constituting an exhaustive set of equiprobable alternatives. Of course, this simple circuit does not meet the requirements of a 'communication system ' as defined by Shannon (1948) and, accordingly, there are very many information-theoretic concepts that cannot be coherently applied to it, for example 'channel capacity.' Nevertheless, the flip-flop circuit does suffice as an example of the information-theoretic concept of a '*bit*'; the event of the flop-flop circuit being in one or another position provides the simplest example of an event involving any amount of *bits* information, and the example can be used to define the statistical sense of '*bit.'*[49] But the memory-related aspects of a situation are as irrelevant as its meaning-related aspects with respect to information-theoretic analyses are concerned (Shannon, 1948, p. 379).

Binary Digits versus Bits

It appears that Miller's (1956) analysis suffers from a word-based confusion, regarding 'bits.' Shannon (1948) writes that "the choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used, the resulting units may be called binary digits, or more briefly *bits,* a word suggested by J. W. Tukey" (p. 379). So, the unit of information measurement, *bit*, is derived from the noun phrase 'binary digits.' 'Binary digits' and 'bits' are entomologically related, but logically distinct. To suppose otherwise is identified as error (i) above: confusing 'number of binary digits' with 'amount of *bits,'* in Shannon's sense. However, most unfortunately, 'binary digits' is often shortened to 'bits' such that there are two distinct senses of 'bit.' On the one hand, there are *bits* of information, in the statistical sense. This concept of a '*bit*' is a unit of measure that expresses the proportion of a set of alternatives that are eliminated by the occurrence of a particular message and/or event. Miller (1953) succinctly defines this sense of 'bit' by example: "if one message reduces *k* to *k/x,* it contains one bit less information than does a message that reduces *k* to *k/2x"* although it

49 The oversight involved in neglecting to explicate the assumption of 'equiprobability' may result from conflating the *arbitrariness* of a particular binary value (e.g., either 0 or 1 can be used to represent 'true') with the idea of binary values being *random* and thereby equiprobable.

is essential to note that this quantitative definition of a bit presupposes that each of the *k* alternatives are equiprobable (p. 4).

On the other hand, there is the sense of 'bit' that is synonymous with 'binary digits.' *This* concept of 'binary digit' refers to a set that contains two members: 0 and 1. For example, the sequence, 0 1 1, is constituted by three binary digits, and this is frequently called 'three bits'. But this case does not involve the sense of 'bits' that figures in, say, '3.3 *bits* of 'information'.' Obviously, there is no such thing as 3.3 binary digits, because there is no such thing as 0.3 of a 0 or of a 1. But is there an amount of *bits* associated with the sequence 0 1 1? That is, is there an amount of information, in the statistical sense, associated with the sequence 0 1 1? Provided that an assumption about the probability of the sequence 0 1 1 is accepted, it is possible to calculate an *H*-value that is (or, would be) associated with the sequence. For example, consider the "message-source" of a series of fair coin tosses for which *heads* = 1. That is, if:

(a) 0 1 1 is generated by a series of discrete random trials, and
(b) 0 and 1 jointly exhaust the set of
(c) equally-likely alternatives (i.e., 0 and 1 are equiprobable, the probability of each being 0.5),

then the sequence 0 1 1 is associated with approximately 1.4 *bits* of 'information', in the statistical sense.[50] So, provided that we observe the requirements of Shannon's (1948) model, and stipulate the probability of observing 0 1 1, it is possible to calculate an

[50] In order to compute the amount of information, in the statistical sense, associated with the sequence 0 1 1 given assumptions (a), (b) and (c), first calculate the binomial probability of the sequence. ${}^{n}C_{r} \cdot p^{r} \cdot q^{n-r}$ for which n = number of trials, r = number of successes, p = the probability that success will occur, and q = the probability that success will not occur. Setting 'success' = 1, we get $({}_{3}c_{2}) \cdot (0.5)^{2} \cdot (0.5)^{3-2}$ = (6)·(.25)·(.5) = 0.375. To calculate the amount of information, in the statistical sense, associated with this sequence we use the formula $H_i = -\log_2 p_i$. So $H_{(011)} = -\log_2 0.375 = \log_2 (1/0.375) \approx \log_2 (2.67) \approx 1.4$ *bits*.

amount of information, in the statistical sense, associated with the sequence 0 1 1. These computations yield a value of 1.4 *bits* of information, in the statistical sense.[51]

But, of course, there is no such thing as a sequence of binary digits that is 1.4 items in length, so it is clear that the sense of 'bits' that figures in expressions of *H*-values, is logically distinct from the sense of 'bit' that is simply a contraction for 'binary digits.' However, it is terribly confusing that a sequence of binary values such as 0 1 1 might be described as 'representing three bits' in two different senses: (1) in the sense of 0 1 1 being *constituted* by three binary values, and (2) the metaphor of one binary value "storing one bit" of information, in the statistical sense, which might be multiplicatively transformed into a claim that three binary values store 3 *bits.*

So, there are two distinct senses of 'bit,' one of which is a contraction of the term 'binary digit,' the other of which pertains to *H*-values. But *is there such a thing as a 'binary digit of information,'* in the statistical sense? No, but it is very easy to get this wrong. It is potentially confusing that the example of a binary digit, and/or a two-position, or "flip-flop" circuit, described in binary terms, is often used to define Shannon's measurement concept of a 'bit.' For example, as quoted above, Shannon (1948) writes that a "device with two stable positions, such as a relay or a flip-flop circuit, can *store* one bit of information" (p. 379, emphasis added). Shannon's reference to *storage* in explaining that 1 *bit* of information is associated with a flop-flop circuit being in one or another of its two positions involves a container metaphor, which is a very common figure of speech (Slaney & Maraun, 2005). However, for the purposes of avoiding confusion, it is preferable to observe the following distinctions. First, if we accept that one binary digit can store one *bit* of information, in the statistical sense, we are thereby assuming that that the concepts of 'binary digit' and 'bit' are logically distinct, for it is

[51] This simple example of a discrete trial is chosen for illustrative purposes. Its didactic power comes at a price, however. The example of a series of discrete trials obscures the fact that one of the most powerful aspects of Shannon's (1948) work is its capacity to account for serial dependencies, for example, the fact that the sequence TH is more likely than TX for English-language messages, such that the receipt of a signal corresponding to T increases the probability that the signal for H will be received next whereas the probability of a signal for X being transmitted is decreased. So, for the transmission of English language messages, the selections of particular signals for transmission *do not* constitute discrete trials.

impossible for something to store itself. Second, because Shannon's concept of a *'bit* of information' can take decimal values, it is clear that '*amounts* of information,' in the statistical sense, are not identical with *counts* of binary digits. So, the best answer to the question *Is there such a thing as a binary digit of information?* is, *No.*

On the Idea of Recalling Bits of Information

The arguments above demonstrate that the concept of a 'bit,' in the information-theoretic sense, is distinct from the concept of a 'binary digit.' So there is a difference between (a) 'recalling a particular sequence of binary digits' and (b) 'recalling an amount of *bits*' on the other. For one thing, there is such a thing as (a) but (b) is a nonsense, because it is not possible to recall an amount of information (*H*), in the statistical sense. Of course, it is possible to recall *that* a particular message and/or event is associated with *N* bits of information, in the statistical sense. But what is thereby recalled is a proposition expressing a fact, not an amount of *N bits*. And, to make matters more complicated, obviously it is possible to recall varying amounts of information, in the everyday epistemic sense, for example, a student who recalls the capitals of all 50 U.S. states recalls more information than a student who recalls only 25.[52] This might be reasonably described as the first student recalling twice as much information than the first, provided it is understood that information is understood in its everyday epistemic sense, and that the unit of information that allows the proportion 'twice as much' to apply here is simply a count of correct items.

.52 If A can correctly recall random sequences of binary digits of up to six items in length whereas B can only correctly recall sequences of up to three items in length, does A recall more information *in the everyday epistemic sense*¸ than B? There does not appear to be a definite answer here. On the one hand, a random sequence of digits does not appear to be well-characterized as a 'statement object of event from which knowledge might be derived.' However, on the other hand, if A can recall twice as many items as can B, A's recalling more might be, somewhat loosely, but understandably, be interpreted as A's *knowing* more about what B stated. In any case, the blurriness of this boundary presents no challenge to the above argument demonstrating the invalidity of Miller's misidentifying increases in memory capacity as involving increases in *bits* of information.

Miller's (1956) Concept of *Recoding*

It is easy to appreciate why Miller (1956) is impressed with Smith's 1954 report that experimental subjects can recall sequences of binary digits of up to 40 items in length. Less clear is why Miller is "convinced that [the] process" of, for example, encoding sequences of random binary digits by "group[ing] by pairs…is a very general and important one for psychology" (p. 93). He explains that, as participants used increasingly longer grouping-units, or to follow Miller, *chunks*, to encode sequences of random binary digits, they were able to recall increasingly longer sequences. At first:

> four possible pairs can occur: 00 is renamed 0, 01 is renamed 1, 10 is renamed 2, and 11 is renamed 3… That is to say, we recode from a base-two arithmetic to a base four arithmetic. In the recoded sequence there are now just nine digits to remember…next….the same sequence of binary digits is regrouped into chunks of three. There are eight possible sequences of three, so we give each sequence a new name between 0 and 7. Now we have receded from a sequence of 18 binary digits into a sequence of 6 octal digits…last…the binary digits are grouped by fours and by fives and are given decimal-digit names from 0 to l5 and from 0 to 3. (pp. 93-94)

Miller (1956) incorrectly characterizes this activity as "recoding" rather than *encoding* (p. 93). He claims that in "the jargon of communication theory" the process of "organiz[ing]" stimuli such as sequences of binary digits "into patterns…would be called *recoding*" (p. 93). However, 'recoding' does not appear in Shannon's (1948) famous paper, (nor in the works of Wiener, 1950, or Weaver, 1949). So the jargon of 'recoding' belongs to Miller (1956), who is concerned that readers might "think of [Smith's (1954) experiment as] merely a mnemonic trick for extending the memory span" (p. 95) and:

> …miss the more important point that is implicit in nearly all such mnemonic devices. The point is that receding is an extremely powerful weapon for increasing the amount of 'information' that we can deal with. In one form or another we use recoding constantly in our daily behavior. (pp. 94-95)

Miller (1956) draws a very general conclusion regarding the supposed constant use of recoding. This is especially puzzling in light of his report that it was difficult for experimental subjects to learn the so called "recoding" (that is, encoding) procedures.

For example, he explains a decline in performance as indicating that "the few minutes the subjects had spent learning the recoding schemes had not been sufficient. Apparently the translation from one code to the other must be almost automatic…" (p. 94). What might justify the claim that persons are *constantly* employing such encoding procedures, even though it takes persons some time to learn how to use them when instructed to do so by an experimenter? The answer to this question is found in Miller's massively expanded definition of 'coding.' Without justification, Miller (1956) assumes that *all* verbal description entails coding:

> In my opinion the most customary kind of recoding that we do all the time is to *translate into a verbal code*. When there is a story or an argument or an idea that we want to remember, we usually try to rephrase it 'in our own words.' When we witness some event we want to remember, we make a *verbal description* of the event and then remember our verbalization. (p. 95, emphasis added)

But does it make sense to conceive of a verbal description of an episodic memory as a code? And even if we accept that verbal descriptions constitute a code of some sort, why would generating a verbal description of an episodic memory be said to involve *re*coding rather than *en*coding?

A Language versus a Code

What is a code? Bennett and Hacker (2003) claim that, "a code is a method of encrypting linguistic expression (or any other form of representation) according to conventional rules" (p. 167). This is consistent with Shannon's (1948) information-theoretic sense of 'encoding,' for example, he writes that "in telegraphy we have an *encoding* operation which produces a sequence of dots, dashes and spaces on the channel corresponding to the message" (p. 380, emphasis added).[53] That is, elements of linguistic representation (in particular, alphanumeric characters) are correlated with sequences of signals (in particular, the signals corresponding to the 'sounds of dots-and dashes') in a rule-guided manner. Accordingly, as Miller (1956) observes, it is possible to learn how to *de*code telegraphic code:

> A man just beginning to learn radiotelegraphic code hears each dit and dah as a separate chunk. Soon he is able to organize these sounds into letters and then he can deal with the letters as chunks. Then the letters organize themselves as words, which are still larger chunks, and he begins to hear whole phrases. (p. 93)

But Miller (1956) claims that such learning involves *re*coding and elaborates that he is "simply pointing to the obvious fact that the dits and dahs are organized by learning into patterns and that as these larger chunks emerge the amount of message that the operator can remember increases correspondingly (p. 93)." This appears to involve an illicit identification of *learning to decode* with *learning by recoding.* That is, from the fact that a telegraphic operator can learn *to* decode telegraphic messages, one cannot validly infer that he is learning *by* recoding them.

Consider Miller's (1956) claim that a "kind of recoding that we do all the time is to *translate into a verbal code*. When there is a story or an argument or an idea that we want to remember, we usually try to rephrase it 'in our own words" (p. 95, emphasis added). On what grounds should describing and/or paraphrasing be considered *recoding*? This entails that all verbal description is a form of coding. However, this is

[53] However, Shannon also applies the concept of 'encoding' to the *mechanical*, rather than *conventional* aspects of communication. For example, he writes that "[i[n telephony [the encoding] operation consists merely of changing sound pressure into a proportional electrical current" (p. 380). And it is this application of 'encode' that has inspired information-processing psychologists to characterize perceptual processes in quasi-information-theoretic terms (e.g., Marr, 1982). This form of encoding, pertaining to continuous rather than discrete categorical variables (e.g., alphanumeric characters) is every bit as rule governed as the conventional case, only differently so. Continuous distributions are chopped up, so to speak, into discrete values on the model of a continuous volume control knob marked 1-10. However, it is important to note that even cases of encoding that do not involve the conventions associated with linguistic expression do involve rule guided operations upon representations that *constitute* particular encoding procedures.

problematic, because the relevant concept of a ‘code’ *presupposes* that of a language.[54] In other words, without a non-encoded system of representation, there cannot be an encoded one. This is because what a code *consists in* is a system of rules for transforming from one system of representation to another. Obviously, it is possible to translate from code to code, for example, the Morse code could be translated into telegraphic code. However, the existence of the Morse code and telegraphic code *qua codes* is that their relationship to non-encoded alphanumeric characters can be specified.

But Miller (1956) construes verbal description as ‘recoding’ not ‘encoding’. This implies that language *itself* is a code as well as that *what* we “rephrase” when we encounter “a story or an argument or an idea that we want to remember” is *also* a code. In his earlier book, Miller (1951) explains that his information-processing psychology involves a very inclusive concept of a code:

> We usually think of codes in terms of secrets and international intrigue, but here we shall speak of codes in a much more general sense. Any system of symbols that, by prior agreement between the source and the destination, is used to represent and convey information will be called a code. (p. 7)

Miller’s (1956) notion of ‘code’ is entirely distinct from Shannon’s (1948). For, the “semantic aspects of communication are irrelevant” to information-theoretic analyses (p 379). But semantic concerns are front and center for a “system of symbols that, by prior agreement between the source and the destination, is used to represent and convey

[54] To be clear, there are other senses of ‘coding’ such as in ‘genetic coding’ that are different in that they pertain to causal as opposed to conventional regularities. In other words, elements of the code are correlated with structural and/or functional elements of organisms through biological laws rather than through stipulated rules. Bennett and Hacker (2003) argue that “We all have got [sic] used to the metaphorical use of the term ‘code’ in the phrase ‘the genetic code.’ It is a metaphor that has been more damaging than illuminating” (p. 167). However, the idea of a genetic code is more comprehensible than Miller’s characterization of language as a “verbal code.” What makes the characterization of genetic material as *a code* easy to get used to is that that elements *of* the so-called genetic code are correlated with elements *outside* the code, so to speak. For example, if a particular sequence of proteins is reliably correlated with a particular morphological feature, then it is easy (although perhaps not optimal) to see the genetic material as *coding for* the feature with which it is correlated.

information.” Furthermore, ‘sources’ and ‘destinations’ of a Shannonian communication system are not capable of agreeing upon anything (see Figure 2.1).

Miller (1989) on Miller (1956)

It is relieving to find that Miller himself is largely in agreement with the iconoclastic arguments above. In 1989, Miller wrote an article in which he recalled being:

> Asked to prepare an invited address for the April 1955 meeting of the Eastern Psychological Association…The invitation threw me into high ambivalence…I wrote a long letter to the program chairman explaining that I was working on two totally unrelated projects at the moment: the application of ‘information’ measures to absolute judgments of unidimensional magnitudes, and the use of recoding to extend the span of immediate memory…The problem, I confessed, was that neither project alone was sufficient for a one-hour public lecture…In order to provide an hour's entertainment, I had to report on both lines of work. The stylist in me refused to give two 30-minute talks having nothing to do with one another. So I asked myself whether there was *anything* in common to the two of them. The only thing I could think of was a numerical similarity. The span of immediate memory for digits is about seven. The channel capacities that had been coming out of the studies of absolute judgments ran around 2.5 to 3 bits of ‘information’. When I suddenly realized that 2.5 bits is six alternatives, I saw how the two might be linked together. It was a superficial similarity, but it enabled me to accept the EPA's invitation. I chose a humorous title for the talk, "The Magical Number Seven, Plus or Minus Two," thinking to make it obvious that I knew this shotgun wedding of absolute judgment and immediate memory was little more than a joke…I don't really understand why the paper has been so widely cited. It has some good ideas in it, but other papers I have written with equally good ideas sank from sight without a ripple. Its central message is that the human mind is limited, which may please some people for reasons of their own. (pp. 400-402)

Miller (1989) distances himself from his much celebrated previous (1956) work. But it is clear that Miller’s lukewarm self-assessment is an outlier. For, Baddeley (1994) explains that:

> In emphasizing the importance of the *recoding* of information and developing the concept of chunking, Miller set the agenda for the next phase of cognitive psychology in which information-processing concepts *went beyond the confines of information theory.*”
> (p. 353, emphasis added)

In other words, Miller (1956) can be read as heralding the era during which the term 'information' is employed "*somewhat* impressionistically" (as cited in Dennett, 2005, p. 15) as:

> few cognitive psychologists measure "information" in its technical sense nowadays, [although] many still use the term…[at times] nearly as a synonym for knowledge. Contemporary research *in the information-theoretic tradition* is no longer concerned with quantity of information, but with the nature of psychological information and its structure.
> (Lachman et al., 1979, pp. 74-75, emphasis added)

Shannon's (1948) Information Theory versus Miller's (1951, 1953, 1956) Information-processing Psychology

Shannon's (1948) information theory is typically understood as an essential stimulus to the now ubiquitous description of the mind (or, of the organism with a mind) as an information-processing system.[55] Miller (1951) was an early adapter of Shannon's work. But it has been demonstrated above that construing information-processing psychology as "research in the information-theoretic tradition" (Lachman et al., 1979, pp. 74) is entirely unjustified. Questions regarding "the nature of psychological information and its structure" (p. 75) are just not information-theoretic concerns. For, as the same authors observe, "Precise measurement of…information [in the statistical sense] requires conditions that do not frequently obtain in the cognitive life of human beings…So cognitive psychologists have mostly abandoned the formalisms of information theory" (pp. 75). And, of course, to abandon the formalisms of information theory is just to abandon any substantive connection between information-processing psychology and Shannon's (1948) information theory.

What does Miller have to say about this? He (Miller, 2003) writes that:

> The Markov processes on which Shannon's analysis of language was based had the virtue of being compatible with the stimulus–response

[55] And/or the brain, or, the mindbrain, mind/brain. The concept of a 'mind/brain' confuses a (set of) *potentialitie*s with an *actuality* through erroneously identifying *an organ* (the brain) with the set of *capacities* (the mind) that causally depend upon that organ (Bennett & Hacker, 2003; Kenny; 1993).

> analysis favored by behaviorists. But information measurement is based on probabilities and increasingly the probabilities seemed more interesting than their logarithmic values, and neither the probabilities nor their logarithms shed much light on the psychological processes that were responsible for them. I was therefore ready for Chomsky's alternative to Markov processes. Once I understood that Shannon's Markov processes could not converge on natural language, *I began to accept syntactic theory as a better account of the cognitive processes responsible for the structural aspects of human language.* (p. 141)

It is striking that Miller (2003) himself does not refer to going beyond Shannon's (1948) information-theoretic concepts, but to abandoning them. Miller's self-assessment of his *Magical Number…* paper is thereby consistent with the critical investigation of that paper above, and inconsistent with what appears to be widely assumed misconceptions regarding some form of continuity between information theory and information-processing psychology (e.g., Adams, 1991; Baddeley, 1994; Lachman et al., 1979; Seow, 2005; Shiffrin & Nosofsky, 1994).

It is *not* the case that information-processing psychologists remained interested in quantifying the degree to which an event eliminated some proportion of predefined alternatives. Rather, psychologists began employing a putatively technical sense of 'information' to phenomena that do not involve eliminating proportions of predefined alternatives and are thereby not amenable to information-theoretic analysis (for example, Miller's 1956 putatively information-theoretic discussion of immediate recall research). *Pace* the widely-assumed continuity thesis (e.g., Adams, 1991; Baddeley, 1994; Dretske, 1981; Lachman et al., 1979, Shiffrin & Nosofsky, 1994), Miller (2003) *does not* refer to extending and/or elaborating upon Shannon's work. Rather, he refers to *"accept[ing] syntactic theory as a better account of the cognitive processes responsible for the structural aspects of human language"* (p. 141, emphasis added). Miller's (2003) recollection is consistent with my argument that Miller's (1951, 1953, 1956) earlier works misinterpreted Shannon's (1948) theory of *signal transmission* sequences as related to symbolic, that is, representational, theories of mind. In Chapter 2, certain precursors to this confusion were identified (e.g., Hartley, 1928; Weaver, 1949). In Chapter 4, I investigate Sloman's (2011) recent attempt to develop an implicit understanding of the 'information' in 'information-processing psychology.'

Chapter 4.

Is the 'Information' in *Information-Processing Psychology* Implicitly Defined?

The previous Chapter 3 corrected certain misconceptions regarding the history of information-processing psychology. In particular, it was demonstrated that the seminal 1950s works of George Miller (1951, 1953, 1956) involved manifold distortions of Shannon's (1948) information theory. So, as is widely recognized there is *historical* continuity between Shannon's information theory and information-processing psychology. However, the venerated status that Miller's work enjoys among information-processing psychologists involves widespread misperceptions of *substantive* continuity between information theory and information-processing psychology. Investigating the works of Miller (1951, 1953, 1956) revealed a number of specific errors (a) to (k) involved in the application of information-theoretic concepts to psychological phenomena. This chapter can be described as a corrective to conventional but misleading responses to the question "Where does the information in information-processing psychology come from?"

In Chapter 2, I identified certain sources of encouragement to the conceptual confusions exhibited by Miller (e.g., Hartley, 1928). But it was also demonstrated that Miller was anything but unique in misapprehending the significance of Shannon's work (e.g., Weaver, 1949; Wiener, 1950). Furthermore, it was demonstrated that Shannon's (1948) influential paper itself involves instances of misleading language, minor errors that cannot be overlooked if one is to achieve what Bar-Hillel (1955) identified as the "psychologically almost impossible" goal of not confusing Shannon's (1948) statistical concept of 'information', with the everyday epistemic sense of the term (p. 284). So, although information-processing psychology in general and the works of George Miller

(1951, 1953, 1956) in particular were subjected to particular scrutiny, an attempt was made to contextualize the critical arguments that were developed. In other words, I attempted to use the works George Miller (1951, 193, 1956) as examples without treating their author as a scapegoat.

Having clarified where the 'information' in 'information-processing psychology' comes from to some, hopefully helpful extent, I will now ask, *Where is the information in information-processing psychology now?* Sloman (2011) recently suggested that 'information' might be *implicitly defined* by the theories, or, more broadly, the scientific discourse in which the term figures. And his argument is suffused with optimism: even if we don't now know just what the 'information' in 'information-processing' means, we will someday. Sloman supposes that the widespread use of 'information' in contemporary discourse reflects an empirical truth about the nature of the universe:

> Words and phrases referring to information are now used in many scientific and nonscientific academic disciplines and in many forms of engineering. This chapter suggests that this is a result of increasingly wide-spread, though often implicit, acknowledgement that besides matter and energy the universe contains information (including information about matter, energy and information) and many of the things that happen, including especially happenings produced by living organisms, and more recently processes in computers, involve information-processing. (p. 1)[56]

[56] Two decades prior, Dretske (1981) also posited an apparently metaphysical significance to information: "In the beginning was information" (p. vii). Without supposing that each of Sloman's (2011) particular arguments is representative of information-processing psychology in general, his analysis of 'information' is more relevant than is Dretske's (1981) for the following reasons. First, Dretske is committed to extending *Shannon's* (1948) concept of 'information' whereas it is widely agreed among information-processing psychologists that their employment of 'information' is distinct from Shannon's strictly statistical concept. Second, Dretske's (1981) interest in information is that he supposes it might afford a "purely physical" non-normative analysis of 'knowledge' (p. *vii*). This is a strictly philosophical interest. I suspect that most psychologists are neutral with respect to the question of whether or not 'information' might ever be defined in physical terms. If they desire philosophical support for setting aside questions of material reducibility, Fodor's (1974) *disunity of science as a working hypothesis* paper and Putnam's (1975) *machine state functionalism* (later criticized by Putnam, 1989, himself) can provide it.

Sloman's (2011) paper is worth investigating in detail, for the following reasons. First, it is concerned with the same question as the present work: what is the 'information' in 'information-processing psychology'? Second, both Sloman and I assume that the answer to this question must come in the form of a conceptual analysis (although our respective analyses are markedly different). Third, Sloman's paper is recent, and so I can be relatively certain that he and I have access to the same body of research and theorizing.

Although they are critical in character, the investigations below are also consistent with Sloman's (2011) suggestion that "focus[ing] more clearly…on aspects of information processing that are not yet understood… in far more detail with far more specific examples, can help…" (p. 28). But, what *counts as* helpful? The type of help that Sloman hopes to offer is that which will "drive advances that will produce new, deeper, *more general* explanations…" (p. 28, emphasis added) of phenomena in information-processing terms. In contrast, the arguments below assume a very different criterion for 'helpfulness.' In particular, the arguments below are deliberately restricted in their focus on conceptual clarity, that is, on distinguishing what makes sense to say from what does not (Baker & Hacker, 1984). A primary method of distinguishing sensible from non-sensible claims is to enquire whether it is possible to specify what would *count as* confirming and/or disconfirming particular, 'information'-related propositions. Examination of information-processing psychology's 'information' leads to an attitude of suspicion towards seeking a *general* explanation of all the sundry phenomena that are currently described in informational terms.[57]

Sloman's (2011) paper is also worthy of investigation because he employs many different, sometimes contradictory arguments, all of which appear to be motivated by a conviction that the sensory, perceptual, and epistemic powers of organisms are best understood in informational terms. It is also helpful that he compares his views to those

[57] Sloman (2011) may be especially permissive in his predication of 'information'. Although it would be unjustified to regard Sloman's particular claims as representative of information-processing psychology, in general, it is fair to regard Sloman's permissiveness as illustrative of contemporary 'information'-laden habits of thought.

of Dennett (e.g., 1987) whose work is more widely read. Sloman's (2011) paper is provides a collection of specific lines of argument, each of which is intended to demonstrate that some degree of obscurity with respect to the meaning of 'information' and/or 'information-processing' is tolerable, or even preferable. Following Sloman, we are to understand that information is *implicitly defined* by the theories in which it figures and that scientific progress depends on its remaining at least *partially undefined,* so that there is something left for information investigators to investigate. This is incorrect on almost every count.

Sloman's (2011) Methodological Assumptions

I begin by outlining certain similarities between Sloman's (2011) paper and the present one, establishing that I am at least making an effort to avoid what Dennett (2005) describes as:

> ...the pitfall of what we might call conceptual myopia: treating *one's own* (possibly narrow and ill-informed) concepts as binding on others with different agendas and training. How, indeed, does [the critic] establish that he and those whose work he is criticizing are speaking the same language? That is surely an empirical question, and his failure to address it with sufficient care [will lead] him astray. (p. 10)

It *is* important that writers be held to relevant standards. Accordingly, I will review Sloman's overall vision of conceptual analysis before turning to his analysis of 'information'. Doing so establishes that Sloman's work is (2009, 2011) topically consistent with the present work. Therefore, in subjecting Sloman's analyses to scrutiny, I am not evaluating his apples on the basis of my standards for oranges. In particular, it is significant that by way of introducing his approach to philosophy, Sloman makes the same (or, a similar) distinction between logical and empirical investigations that is essential to my own arguments below. Sloman distinguishes between logical and empirical issues using a cartographic metaphor, which nomenclature he attributes to Ryle:

> The outcome of conceptual analysis, called logical geography by Gilbert Ryle, can be seen as a transient/culturally-based patchwork imposed on an enduring terrain. The underlying logical topography is discovered (gradually) by non-philosophical (scientific and technical) advances. Compare (a) the early theories of kinds of stuff (earth, air, water, wood, carbon, stone, etc.—the old logical geography) with (b) what came after discovery of the architecture of matter and the periodic table of the elements, plus chemistry (new logical topography, spawning new logical geographies). (as cited in Sloman, 2009, slide 6)[58]

In this formulation, 'logical topography' refers to empirical phenomena whereas 'logical geography' refers to the network of concepts through which empirical phenomena are described, classified, and investigated.[59] More crudely, the distinction is between reality and language, between what exists and how we talk about it. It is somewhat confusing that both the topography and the geography are described as 'logical' however. It might be clearer to speak of a *conceptual* geography and an *empirical* topography. Nevertheless, Sloman's (2009) version of Ryle illustrates that scientific investigations can be heuristically divided into empirical questions concerning what the world (the 'logical topography') is like on the one hand, and those concerning how our concepts (our 'logical geography') carve it up, on the other.

Sloman (2009) continues:

> A logical geography carves up some region of reality in a particular way, for particular purposes, partly under the influence of a particular cultural history. E.g. an ontology for kinds of matter and kinds of transformations of matter. The advance of science and technology can gradually reveal a richer underlying reality—the logical topography—that can be divided up in various different ways. The process of "uncovering" the underlying structure can continue indefinitely, revealing alternative ways of carving things up: alternative conceptual schemes, logical geographies. E.g. as more and more is learnt about the architecture of matter, the many forms

[58] Note that the quotes from Sloman's (2009) work are from his notes for a presentation. He makes these notes available on his website and refers to them in his more formal 2011 paper. Because they are notes for a spoken presentation rather than a written work, these quotes should be read charitably with respect to style. Also, I have altered the formatting of certain bits of the text that I quote from this work.

[59] Of course, concepts can be used to do other things as well, such as describe imaginary beings and situations, (e.g., unicorns in outer space), write poems, etcetera.

> it can take and the kinds of process that can occur that depend on the structure, we modify and grow our ontologies. Technology/engineering can contribute substantially to showing what is possible, partly by developing devices that expand what we can observe (e.g. microscopes), and partly by creating things that had never previously existed, that challenge old ontologies. (Slide 9)

Sloman (2009) emphasizes that we are obliged to consider whether or not particular aspects of our logical geography are sensible, as knowledge of a particular subject matter evolves. The arguments below aspire to follow this sage advice with respect to the concept of 'information' in psychology.

Criticism versus Conservatism

Sloman (2009, 2011) highlights that critical, conceptual inquiry is not necessarily motivated by conservative impulses to resist changes in linguistic usage. Nor must it involve a presumption that the critic has privileged access to immutable linguistic rules, as Dennett (2005) for example, supposes:

> If [the critic] were able to *show us* the rules [of language], and show us just how the new uses conflict with them, we might be in a position to agree or disagree with him, but he is just making this up. He has no idea what 'the rules'…are. (p. 10)

Sloman (2009) corrects Dennett's (2005) deflationary vision of the conceptual analyst as "an old-fashioned grammarian scolding people for saying 'ain't'" (p. 13), by describing the more modest (and less ridiculous) role of helping "the sciences clarify the logical topographies they uncover, and investigating pros and cons of alternative logical geographies based on those logical topographies" (Sloman, 2009, Slide 8).[60] If the critic has any force, it is due to *arguments* rather than to insight into obscure *rules* of language (which, of course, no one actually claims to have). It is investigating particular applications of concepts that leads to conclusions regarding coherence. Accordingly, I subject Sloman's (2011) investigation of 'information' in the psychological sciences to the standards that he sets for himself.

How does the conceptual critic avoid inadvertently playing the part of an old-fashioned, possibly-irrelevant grammarian? Sloman (2009) recommends sustained attention to scientific discovery: "Doing…conceptual analysis…in ignorance of new discoveries and achievements in science and technology, is a recipe for a sickly inbred species of thinking" (Slide 7). As an example of "risks" Sloman notes that:

> at one point…a [researcher] analysing common concepts would have included whales, dolphins, etc. as types of fish, e.g. because of their form, habitat and behaviors. As a result of evolutionary theory and discovery of more empirical facts they are now classified as mammals, requiring a revision of our concepts of types of animal. (Slide 7)

We should expect that as scientific knowledge evolves, so do the concepts that scientists employ. Elsewhere he warns that conceptual investigators "ignore biology, psychology, neuroscience and developments in computation at their peril" (Slide 7). Psychological investigations are of particular interest in this regard:

> Muddles in our pre-theoretical concepts of mind surface when we try to ask philosophical or scientific questions, e.g.…[w]hat we normally refer to as consciousness involves the exercise of a large, diverse, ill-defined cluster of capabilities…If there is no well-defined subset of capabilities which are necessary or sufficient for consciousness, then some of our apparently meaningful questions, like many questions involving cluster concepts, may be ill-defined. Many mental concepts share this semantic indeterminacy, e.g. 'emotion', 'intelligence', 'understanding', 'pleasure'….
> (Sloman, 1999, p .1)

[60] I have replaced "philosophers of mind" with "conceptual analysts" and/or "researchers" throughout this review of Sloman (2009) because I believe that conceptual concerns are (or, ought to be) pan-disciplinary and, therefore, relevant to non-philosophers as well.

'Information Architectures' as the Referent of Psychological Predicates

Are there any islands of semantic specificity to which psychological researchers might anchor their investigations? Sloman (1999) argues that 'information' might be the bedrock beneath these confusing cluster concepts: "what we normally refer to as consciousness involves the exercise of a large, diverse, ill-defined cluster of capabilities (many of them unconscious!) *supported by our information-processing architectures*" (p. 1, emphasis added).[61] Assuming that "supported by" means roughly, c*asually dependent upon*, it appears that Sloman regards the 'information' in 'information-processing psychology' as markedly different from such "ill-defined…cluster concepts" as the respective concepts of "thoughts, desires, emotions" (p. 1). Before investigating Sloman's arguments in favor of ambiguity with respect to 'information,' I will quickly review the case against thinking in undefined terms.

Ambiguity as a Vice

I assume that not knowing what 'information' means is a problem for information-processing psychology. If it is unclear what *counts as* information, then it is not possible to test empirical hypotheses regarding it being processed in one way or another. It is a logical truth that a phenomenon must be identified in order to be studied, and, therefore, in order to test an empirical hypothesis that the mind and/or brain *processes information* in one or another way, it must first be clear what would count as rendering that hypothesis true or false (Bennett & Hacker, 2003). However, information-processing psychologists can circumvent this problem of unidentifiability because the empirical phenomena of interest—the various sensory, perceptual, epistemic, developmental, social-behavioral, etcetera capacities of organisms (and/or, for some, machines)—can be identified *without* presupposing the coherence of the information-processing vocabulary. For example, to study visual information-processing is, in less theory-laden terms, to study vision. An information-processing theory of vision is supposed to help perceptual psychologists investigate one or another aspect of vision (e.g., Marr, 1982).

In response to criticism regarding ambiguity and equivocation (Bennett & Hacker, 2003), Dennett (2005) has suggested that it is "useful to speak, *somewhat* impressionistically, about…information being processed…[and recommends that] instead of doing what a philosopher might do when challenged about what they meant, namely *defining their terms more exactly*, they instead [should] point to their models" (Dennett, 2005, p. 15). But this will not do as a defense. First, meeting *one* criterion of scientific success, accurate prediction, does not license apathy toward others, such as linguistic precision and semantic clarity. If a particular information-processing "model" of *X or Y* is working in certain respects, then that is all the more reason to seek clarity regarding the 'information' in involved in its description. As Fodor (1975) succinctly puts it, "For better or worse, the ontology of the theories one accepts is *ipso facto* the ontology to which one is committed" (p. 5). And to criticize assumption *A* is not to attack those who assume *A*. To expose impurities in the bathwater is to help the baby.

Ambiguity as a Virtue

In contradistinction to my arguments, Sloman (2011) defends information as a term of art in psychology (and beyond). He even argues that ambiguity is *preferable* to consensus regarding what information means. He contends that "In order to understand

[61] What is an 'architecture'? It appears to be, roughly, a structural theory of the mind. Sloman writes that artificial intelligence researchers (and, I would add, information-processing psychologists) have long "been concerned with algorithms and representations, but we also need to understand how to put various parts together into complete working systems, within an architecture…Explicit or implicit theories of mental architecture are not new….Kant … proposed [an] architecture with powerful innate mechanisms that enable experiences and learning to get off the ground, along with mathematical reasoning and other capabilities. Freud's theories directed attention to a large subconscious component in the architecture…." (Sloman & Scheutz, 2002, p. 1).

how a concept like ‘information’ can be used in science without being definable, we need to understand some general points from philosophy of science” (p. 9). Sloman (2011) then claims that linguistic definitions of ‘information’ are deficient because they inevitably presuppose related concepts:

> …information…cannot be explicitly defined without circularity…Attempts to define “Information” [*sic*] by writing down an explicit definition of the form “Information is...” all presuppose some concept that is closely related (‘meaning’, ‘content’, ‘reference’, ‘description’, etc.). ‘Information is meaning’, ‘information is semantic content’, ‘information is what something is about’ are all inadequate in this sense. (pp. 9-10)

In the fact that information cannot be defined without employing related concepts, Sloman (2011) finds circularity. This criticism would seem to apply to the *double-life* definition of everyday epistemic information developed in Section 1: (1) statements, objects, and/or events from which knowledge might be derived and (2) knowledge potentially derivable from statement, objects, and/or events. Sloman (2011) might observe that such a “circular” definition is only adequate if the meanings of such concepts as ‘statement’ and ‘knowledge’ are known (p. 10). However, he assures his reader that “this kind of indefinability is common in concepts needed for deep scientific theories” and that implicit definitions are not only needed for “information; but also for ‘mass’, ‘energy’ and other deep concepts used in important scientific theories” (p. 10). But it is far from clear why Sloman should characterize the *internal relations* between ‘information’ and other concepts that are “closely related” in meaning as establishing that ‘information’ suffers a “kind of indefinability” (pp. 9-10). The ‘perimeter’ and ‘radius’ of a circle, respectively, are internally related and mutually co-defining; this is different from their being undefined or indefinable (Bickhard, 2003).

Sloman’s (2011) suspicion of verbal definition leads him to suppose that information will be defined *empirically*, by identifying a theoretical model of some information-related phenomena and assessing the degree of correspondence between the model and reality:

> If a theory is expressed logically, and is not logically inconsistent, and its undefined concept labels are treated as variables ranging over

> predicates, relations and functions, then there may be a non-empty set of possible models for the set of statements expressing the theory, where the notion of something being a model is illustrated by lines, points, and relations between them being a model for a set of axioms for Euclidean geometry, and also certain arithmetical entities being a model for the same axioms. The models that satisfy some theory with undefined terms will include possible portions of reality that the theory could describe. (p. 10)

If we strip away the veneer of mathematicity, Sloman (2011) appears to be suggesting that if a model that corresponded to certain empirical phenomena were to be derived from a set of theoretical axioms that involved an undefined term, the correspondence between the model and the "portion of reality that the theory" described would justify the use of the undefined term (p. 10). Sloman argues that:

> for concepts that are implicitly defined by their role in the theory, the evaluation of the concepts as referring to something real or not will go along with the evaluation of the theory…Concepts like 'angel' and 'fairy' are examples of such referentially unsuccessful concepts. (p. 11)

But this notion of 'implicit definition' confuses *nonexistent* with *undefined*. From the (true) proposition that *angels and fairies do not exist*, one can infer that 'angel' and 'fairy,' respectively, do not refer to any constituents of reality, but not that the concepts of 'fairy' and 'angel' are undefined. For regarding the proposition *angels and fairies do not exist* as either true or false and/or treating it as a valid premise in an inferential argument *presupposes* that the meaning of the statement is determinable.[62]

[62] It is also far from clear that 'angel' and 'fairy' should be viewed as aspects of a theory of any kind, especially in light of Sloman's (2011) notion of a theory as involving "undefined concept labels [that] are treated as variables ranging over predicates, relations and functions…" (p.11). While it *might* be fair to regard 'angel' and 'fairy' as aspects of religious and/or supernatural theories, they are certainly very different kinds of theories than those with which Sloman is concerned.

Defining versus Theorizing

Sloman (2011) argues that in order to adequately understand what 'information' means we must develop a:

> deep and complex theory of how parts of the universe that use or interact with information work, for instance entities (information users) that do various things with information: acquiring, manipulating, combining, deriving, storing, retrieving, comparing, analysing, interpreting, explaining, indexing, annotating, communicating, and above all using information for practical purposes. (p. 3)

This interest in theory-construction is consistent with the assumption that "besides matter and energy the universe contains information" (p. 01). That is, Sloman appears to regard 'information' as referring to a class of empirical phenomenon. With Logan (Sloman & Logan, 1998), he even "conjecture[s] that such concepts" as "'believes', 'desires', 'intends' and 'feels'…are grounded in a type of information processing architecture, and not simply in observable behavior" (p. 1).[63] It is not clear what it is for a concept to be *grounded in* one thing rather than another, but presumably it has something to do with establishing its meanings, specifying its class(es) of reference, and/or justifying its ascription. According to this logic, such common psychological predicates as "'believes', 'desires', 'intends' and 'feels'" refer to types of "information-processing architecture[s]" (p. 1), and thusly, by discovering what those specific architectures *are*, we thereby discover the *meaning* of those psychological predicates. That is, if we could identify the "type[s] of 'information'-processing architecture[s]" (p. 1) in which psychological predicates are "grounded" (p. 1), this would allow us to define such predicates ostensively, by pointing to instances, much like one might define 'red' by pointing to examples of red objects.

[63] It would be unfair to regard Sloman's views on this issue as representative of 'information'-processing psychology in general, as it is clearly possible to maintain that the abilities to believe, desire, feel and so on are *causally dependent* upon hypothetical information-processing mechanisms/architectures without also supposing that such psychological predicates *refer to* those hypothetical mechanisms and/or architectures.

Sloman and colleagues (Sloman, 2011; SLoman & Logan, 1998) supposition that a theory-building enterprise will succeed in implicitly defining 'information' as well as other psychological terms will be familiar to most psychologists. His argument is broadly consistent with Cronbach and Meehl's (1955) *construct validation theory*, which plays a foundational role in contemporary psychological methods. Construct validation theory is premised on the assumption that, being putatively unobservable, "psychological processes are elusive" (p. 286) and that "scientifically speaking, to 'make clear what something *is'* means to set forth the laws in which it occurs" (p. 290). For example, it has been suggested "that we cannot say what anxiety, dominance, empathy, memory, or pain is in a suitably scientific way until we have investigated" (Jost & Gustafson, 1998, p. 474). But it is mistaken to regard scientific investigations into *X*-related phenomena as revealing what '*X'* means, because the meaning of '*X'* is what determines what *counts* as *X*-related phenomena (Bennett & Hacker, 2003; Dupre, 1993; Maraun & Peters, 2005).

Information and Inform-able Agents

In Chapter 2, I emphasized that the everyday sense of 'information' is an *epistemic* one. So is Sloman's (2011):

> Information cannot play a role in any process unless there is something that encodes or expresses the information: an 'information bearer' (B), and some user (U) that takes B to express information I (i.e. interprets B). The same bearer B may be interpreted differently by different users, and the same user, U may interpret B differently in different contexts (C). (p. 3)

In its general outlines, this formulation appears to constitute a reminder that the noun form, 'information' presupposes both an inform-able agent ("a user") and a source of 'information' (a "bearer"). Sloman supposes that "We need a theory that explains the different ways in which a bearer B can express information I for U in context C, and what that means" (p. 3). What follows is *not a theory* but rather *an analysis* of what, if anything, Sloman, in particular, and information-processing psychologists more generally mean in construing the epistemic and agentive capacities of organisms in terms of their information-processing.

Mind as Informational

Sloman (2011) regards 'information-processing' as the empirical subject matter of psychological investigation, rather than as a theory-laden description of such subject matter. He writes that "it is arguable that all living organisms acquire and use information, both in constructing themselves and also in controlling behavior, repairing damage, detecting infections, etc." (p. 13). However, *he does not argue this*. Rather, he assumes equivalence between ascribing 'mind' and ascribing 'information-processing,' stating that "there is a sense in which life presupposes mind (informed control)" (p. 14, footnote). He claims (elsewhere) that "microbes, insects, many other invertebrates…all have minds in the minimal sense of being capable of selecting among behavioral alternatives on the basis of available information" (Sloman, 2009, Slide 12), and elaborates:

> Life takes many forms, But they all involve control—[i]nformed control. Control involves selection among sets or ranges of possibilities. Control can be informed by results of various kinds of external and internal sensing....if we construe 'mind' as a label for a collection of 'information' processing capabilities—of any kind, then we can conclude [that] [l]ife requires [m]ind. Simple life forms have very simple minds—but as needs become more complex so do the control systems. (Slides 21-22)

How does this stand up to critical scrutiny? I will paraphrase Sloman's (2009) argument claim-by-claim.

- Even the simplest of organisms have minds because they demonstrate rudimentary epistemic and/or agentive powers;
- The agentive powers of simple organisms are limited and, therefore, their actions can be understood as selections from a range of possibilities, and, in that respect, they exercise *control,* control that is
- *informed* by sensory processes. Accordingly, if we accept that
- the simplest organisms' mind-related abilities involve those organisms exercising *informed control,* then we should also accept that

- the mind consists of a “collection of information processing capabilities,” and, so, Sloman concludes,
- life presupposes information-processing.

Is it valid to regard limit cases of ascribing mind-related predicates as paradigm cases of mindedness? Are the rudimentary epistemic and/or agentive powers of simple organisms well-characterized as *informed control*? Is the ability to *become informed* through sensory process the same as *using information* acquired via the senses? Such questions are considered below.

On What Bases Can Organisms Be Ascribed ‘Minds’?

Sloman (2009) observes that very many organisms “have minds in the minimal sense of being capable of selecting among behavioral alternatives” (Slide 12). He perceives this claim to be vulnerable to criticism for anthropomorphizing and anticipates his critic by stating that he does not “presuppose rationality” in ascribing mind-related predicates to simple organisms (Slide 12). This is an important qualification. Sloman appears to be correct in claiming that it is valid to attribute *rudimentary forms of* mind-related predicates to simple organisms, for example, if a microbe attempts to eat a foreign body, then it is perfectly coherent to say that the microbe *thinks* and/or *believes* that the foreign body is a food source. Such a description is not necessarily overly generous with respect to microbes’ intellectual powers, as it does not entail or even imply that other, more sophisticated thought- and/or belief-related abilities (for example, the ability to think/believe that there is a war in Afghanistan) are thereby also attributed. Because it is possible to explain what it means to claim that *the microbe thinks and/or believes X is a food source*, there is no logical problem raised by this particular case of thought/belief attribution.

Although Sloman’s (2009) claims regarding the broad ascribability of mind-related terms are reasonable, there is a problem with the structure of his argument: it assumes that the most rudimentary forms of behavior which satisfy the criteria for attributing psychological predicates constitute *paradigm cases* which reveal the *essence* of mind-related attributions. But this is an unjustified assumption. It is far from clear that *all* mind-related abilities involve an organism *becoming informed.* For example, the capacity to write a poem is clearly a mind-related capacity, but it is not clear how it involves a poet-organism becoming informed and/or using information.[64]

Becoming Informed versus Using Information

If we agree with Sloman (2009) that microbes have minds, must we also accept that they use information? These waters are murky. On the one hand, it is a logical truth that simply by virtue of sensing objects in its environment, a microbe becomes informed of what is around it. Furthermore, it appears inevitable that in describing the behavior of simple organisms, we depict them as suffused with purposes, intentions, and desires—for example, a microbe that eats might be characterized as wanting to eat, intending to eat, and, by virtue of its eating, as realizing one of the few purposes that it is clearly safe to attribute to such simple organisms (i.e., a purpose to eat). “We see human behavior, mien and expression as *informed* by…thought, feeling, purpose and intention” (Hacker, 2007, p. 14). And microbe behavior is informed by (attenuated) versions of these psychological phenomena. So with respect to ‘information’-related predication, it appears unproblematic to say, for example, that when a microbe senses a potential food source, it has thereby *become informed* of a food source. This might be paraphrased as saying that the hypothetical, hungry microbe has “acquire[d] and use[d] information” from its environment regarding this potential food source (Sloman, 2011, p. 13). On the other hand, one might argue that it only makes sense to say that an agent ‘uses information’ if it is also possible for it to *not use* that information, that one cannot be informed of something about which it is not possible to be uniformed (Bennett & Hacker, 2003; Wittgenstein, 1972). By that logic, it would *not* make sense to say that an organism’s non-voluntary sensing and/or perceiving of objects and forces in its environment involves

[64] Sloman (2011) has other arguments to accommodate this objection, as he assumes that the distinctly non-microbial phenomena of language also involves information:

> “In constructing the question ‘Is that noise outside caused by a lawnmower?’, a speaker can use the same concepts and the same *modes of composition of information* as are used in formulating true beliefs like: ‘Lawnmowers are used to cut grass.’ (p. 9, emphasis added)

No justification for construing constructing a sentence as composing with information is given.

its *using* (sensory and/or perceptual) *information*, because it would be impossible for that organism to *not use* that information (e.g., look at your hand and try not to see it).[65] Should a microbe sense a potential food source that it does not attempt to feed on, this might be said to count as its having acquired information that it has not used. However, this is a *different* criterion of 'using information,' requiring not only sensing but also reacting.

There does not appear to be a clear distinction between right and wrong here. But this much appears certain: by virtue of exercising perceptual and sensory capacities, organisms become informed of certain features of their environments. However, it is not clear that all such cases of becoming informed should count as *using information*. Nevertheless, Sloman's (2009) identification of 'becoming informed' with 'acquiring and/or using information' appears to be representative of information-processing psychology generally. For example, in what might be reasonably described as a founding document of information-processing psychology, Miller (1951) claims that: "in an ever-changing world, the state of the organism and the state of the environment must be able to mold and direct the organism's behavior. *The necessary information* is supplied by specialized cells called receptors and neurons" (p. 4, emphasis added*;* see also Wiener, 1950).

Does Becoming Informed Involve Processing Information?

Even if we charitably decide that it makes sense to say that the environment-sensing microbe is *using* information from its environment, there is still an important difference between saying (1) that a microbe *uses* the information in its environment and (2) that it *processes* information from its environment. The truth of (1) is guaranteed by the fact that 'using information [from the] environment' is defined here as 'sensing objects and forces in the environment.' But Sloman (2011) asserts not just that

[65] It is probably helpful to distinguish between the everyday concept of 'seeing,' with connotations of conscious perception, and 'vision,' which casts a wider net. For example, it is typical to speak of peripheral vision rather than 'peripheral seeing' and we might say that

information is used and/or acquired but that “information is *processed* in perceiving, learning, wanting, planning, remembering, deciding, etc.” (p. 18, emphasis added). What is the nature of this processing, and how does it relate to using and acquiring? This is difficult to determine, because Sloman appears to assume that the meaning of ‘processing’ is self-evident. In Chapter 3, this was also observed in the work of Miller (1956) who breezily switches from discussing research-participants-qua-communication-channels’ abilities to ‘*transmit* information’ (an already-misleading phrase—see Chapter 2) to a discussion of “limitations that are imposed on our ability to *process information”* (p. 92; see also Wiener, 1950). The ‘processing’ in information-processing psychology appears to be every bit as inchoate as the ‘information.’ If we follow Sloman’s (2011) lead and consider ‘processing’ to be *implicitly defined* by the theories in which it figures, what would justify “constru[ing] ‘mind’ as a label for a collection of ‘information’ *processing* capabilities”? But first, according to Sloman, *what is it* for an organism to process information?

> Every living thing *processes* information insofar as it uses (internal or external) sensors to detect states of itself or the environment and uses the results of that detection process either immediately or after further information-processing *to select from a behavioral repertoire,*[66] where the behavior may be externally visible physical behavior or new information-processing. (p. 26, emphasis added)

The justification for ascribing ‘information-processing’ is basically identical with the justification for ascribing ‘mind’: “microbes, insects, many other invertebrates…all

edge detection (e.g., Marr & Hildreth, 1980) is an aspect of vision, but not of seeing, given that we do not typically experience the detecting of edges.

66 Are there any un-explicated connections between ‘selecting from behavioral alternatives’ and ‘information-processing capabilities’? What if we replace ‘selecting from behavioral alternatives’ with ‘selection among sets or ranges of possibilities’? In light of Chapter 2, in which Shannon’s (1948) statistical concept of ‘information’ (or ‘amount of ‘information’’) is explicated, a connection can be discerned. Sloman’s argument might be read as suggesting that the activity of organisms with “two-way powers, i.e. the power to V or not to V” (Hacker, 2004, p.14) can be construed as ‘selecting from a range of possibilities’ in the same manner that information measurement, in Shannon’s (1948) sense, pertains to the degree to which a set of alternatives is reduced. However, Sloman (2011) is clear that the “more recent use of the word ‘information’ in the context of Shannon's ‘information theory’…does not refer to what is normally meant by ‘information’ (the topic of [his] paper)” (p. 5).

have minds in the minimal sense of being capable of selecting among behavioral alternatives on the basis of available information" (Sloman, 2009, Slide 12). But this justification involves certain troubling contradictions. First, the proposition that "every living thing processes information insofar as…" (Slide 12) it senses and "select[s] from a behavioral repertoire" (Slide 12) appears to place a limit on what counts as 'information-processing' in this biological context. This formulation suggests that it is the 2-way volitional powers (to *V* or not to *V*) of organisms that are to be understood in information-processing terms (see Hacker, 2007). In other words, the criteria for 'processing information' includes sensing or becoming informed, as well as *deciding* to *V* or *Z*:

> the information-processing viewpoint—we study: control—selecting among alternatives, according to changing requirements. This involves using information, in discovering options, selecting between options and in carrying out *decisions*. (Sloman, 2009, Slide 19, emphasis added)

But Sloman also asks his reader:

> …not [to] expect a sharp divide between things that do and things that do not process information…[as] [t]here are also many different cases, between objects that merely react to forces by accelerating, though negative feedback control systems (homeostats), and other systems that use quantitative information to control quantitatively varying effector signals—to mechanisms that build and use complex enduring information structures with information about themselves and the environment, past, present and future… (Slide 20)

So, information-processing' is not necessarily about deciding or discovering, but also about "merely reacting to forces" (Slide 20). It is very difficult to identify just what Sloman's argument is. But his overall theme is easy to discern—information-processing, whatever it is, is everywhere. Such diverse activities as:

> growth, repair, digestion, distribution of chemicals (e.g. products of digestion, respiration, and waste materials), conversion of chemical into mechanical energy, [and] defence against attacks of various kinds…all involve bits of the body being involved in (mostly molecular scale) physical and chemical processes concerned with creating, servicing, or controlling bodies and their functions. They all seem to involve some

> form of 'information'-processing, because they all involve control, and control needs to be informed, so as to be sensitive to needs, opportunities, constraints and risks. (Slide 19)

From "merely react[ing] to forces" (Sloman, 2009, Slide 20) to *digesting* and on to *believing*, it seems that there are very few activities that are not describable in terms of 'information-processing,' provided that the claimant is willing to incur the cost of being unable to explain what such claims *mean*. It is difficult to understand how an *implicit* definition of 'information' is supposed to emerge from this tangle of inconsistency, especially in light of the assertion that:

> A great deal of 'information'-processing can also occur outside the body, and often does. For example, humans can reason, remember, and refer with the aid of: external diagrams, devices, and other people; and many animals use scents, markings, landmarks, etc. in controlling behaviors…" (Slide 15)

Sloman (2009) is extremely permissive in predicating 'information-processing,' and, in this respect, it would be unfair to regard his writing as representative of information-processing psychology in general. However, in observing that "humans can reason, remember, and refer with the aid of: external diagrams, devices, and other people" (Slide 15) Sloman describes what appears to be straightforward cases of using information, in the everyday epistemic sense. To review briefly, in Chapter 2, it was demonstrated that the everyday sense of 'information' is an *epistemic* one, and the working definition of everyday epistemic information is described as a *double-life* definition. It casts two wide nets. On the one hand, what counts as information includes (1) statements, objects, and/or *events from which knowledge might be derived.* On the other hand, the (2) *knowledge potentially derivable* from statement, objects, and/or events *also counts as* information in the everyday epistemic sense. Both a *source of* knowledge *that p* (e.g., to learn from information) as well as *what* an agent who knows *that p* knows (e.g., that Vancouver, BC, is north of Seattle, WA) count as information. So the everyday sense of 'information' leads a double-life, designating both sources of knowledge (Dretske, 1981) or belief (Sloman, 2011) (e.g., propositions, maps, plumes of fire-locating smoke, etc.) as well as objects of knowledge and/or belief (e.g., that the cat is on the mat, that the mat is east of the window, that the mat is on fire, etc.). In referring to *reasoning, remembering,* and *referring* with the help of *external diagrams, devices,* and *other people*, Sloman does not appear to be employing an implicitly defined technical sense of 'information'. And if, say, referring to a diagram counts as *processing information outside the body*, Sloman inadvertently raises a very powerful objection to the very paradigm that he defends. How is the ability to use an external diagram supposed to be *explained by* the use of an *internal* one?

Capabilities versus Mechanisms

Sloman's (2009, 2011) arguments are intended to establish that mind consists in a set of 'information'-processing capabilities. If we charitably accept that (1) activities as different from one another as a microbe sensing a food source and a speaker constructing a sonnet all count as *using information,* and that (2) becoming informed and/or using information entails the *processing of information*, then we might accept the identification of 'mind' with 'a set of information-processing capabilities.' But it is *mechanisms* and not *capabilities* that interest Sloman (and information-processing psychologists in general). Sloman (2011) presents his paper as "an outline of a theory about the processes and *mechanisms* various kinds of 'information' can be involved in" (p. 1, emphasis added). He contends that "intelligent agents can use the environment as a store of 'information' or as a source of 'information' or as *part of a mechanism* for reasoning or inferring…" (p. 7, emphasis added) and asserts that "there are many ways in which information media can vary, imposing different demands on the *mechanisms that process* them" (p. 18, emphasis added). But Sloman appears to regard 'capabilities,' 'mechanisms,' and 'architectures' as interchangeable, so it is easy to agree with him that "the information-processing mechanisms and forms of representation required for perceivers to acquire and use…are not yet understood" (p. 16).

There is a difference between construing mind as a collection of *capabilities* on the one hand, and a collection of *mechanisms*, on the other. An important ontological distinction is that *capabilities are potentialities, whereas mechanisms are actualities.* The *Oxford English Dictionary* defines 'mechanism' as employed in the psychological sciences as: "An unconscious, structured set of mental processes underlying a person's behavior or responses" ("Mechanism," 2012).[67] Presumably, 'underlying' means here something in the genre of 'causally explaining.' It appears fair to characterize information-processing psychology as the attempt to identify and describe the information-processing mechanisms that are hypothesized to be causally involved in either/both observable behaviors and/or behavioral capabilities, such as short-term memory capacity (e.g., Baddeley & Hitch, 1974) or the ability to read (e.g., Laberge & Samuels, 1974). The positing of information-processing mechanisms is supposed to *explain* the observed capabilities, so it is essential to distinguish between the *explanandum* of the capabilities and the *explanans* of the mechanisms.

Information-Processing: Descriptive or Explanatory*?*

If 'processing information about a food source' refers to the same set of behavioral phenomena as 'sensing a (potential) food source' and/or 'responding to a (potential) food source,' then the information-processing psychologist is simply employing an ornate vocabulary in describing the data. But it is clear that information-processing psychology is not simply *a way of talking about* behavior. Rather, it is a paradigm of scientific psychological explanation, a conceptual framework for generating explanations of such behaviors (Lachamn et al., 1979). However, in order for a supposed explanation *E*, *processing information about a food source* to (partially or completely) *explain* behavior *B, sensing a food source*, *E* and *B* must be empirically distinguishable. That is, if the criterion for identifying *E* is the identification of *B*, then *E* does not appear to aid in understanding *how* an N manages to *B*. If the goal of positing information-processing mechanisms is to explain observed behaviors, then it is problematic that the putative effects (observed behaviors) cannot be *logically* distinguished from the putative causal-correlates (the hypothesized information-processing mechanisms).

[67] It is preferable to replace "person's" with "organism's."

Bearers of Information versus Sources of Information

Sloman (2009, 2011) presents it as a *logical* truth that life presupposes informed control, and, by implication, information, whatever exactly it is. It is worthwhile to consider what counts as 'available information' for a microbe. It would appear that this refers primarily to what the organism senses and/or perceives. But describing sensible and/or perceptible objects and forces as "available information" simply begs the question. Furthermore, organisms sense objects and forces, not *information about* objects and forces. There is a difference between perceiving a falling rock, on the one hand, and, say, reading a *Danger—Falling Rocks* sign, on the other. But Sloman (2011) does not distinguish between 'bearers of information' and 'sources of 'information''. He contends:

> it is arguable that any object, event, or process is intrinsically a bearer of information about itself (a 'self-documenting' entity), though not all users are equally able to acquire and use the information that is available from the entity. So a twig lying in the forest is a bearer (or potential bearer?) of information about its size, shape, physical composition, location, orientation, history, and relationships to many other things. (p. 16)

But there is a difference between *being* an object that is sensible and/or perceivable, and a sensible and/or perceivable object *bearing information* about itself. Very many if not all sensible and/or perceivable objects, events, or processes are potential *sources* of information, in the trivial sense that, for example they can be seen, touched, smelled, etcetera. But this is different from the sense in which the spine of a book might *bear information* about its identity in the form of a title. Sloman (2011) "uses 'representation' to refer to any kind of information bearer" (p. 03). But there is a difference between *being* and *representing*, and evidence for this distinction can be found within Sloman's own argument. He also claims that "there are all sorts of things you can do with information that you would never do with what it refers to and vice versa. You can eat food, but not information about food" (p. 6). Sloman papers over the gap between these two views by subsequently claiming that that "any object, event, or process is" *not* "*intrinsically* a bearer of information about itself" (p 16) but rather that:

> Self-documenting entities…*potentially* express information for various kinds of information user simply in virtue of their structure, properties and relations. These information bearers do not depend for their existence on users [and] can be contrasted with the sensory signals and other transient and enduring information bearers constructed by information users. (p. 18, emphasis added)

Sloman's (2011) one-concept-fits-all employment of 'information' to characterize all things mind-related leads him into cul-de-sacs of contradiction. On the one hand, "any object, event, or process is *intrinsically* a [representation of] itself" (p. 16, emphasis added) because it can be sensed and/or perceived but on the other, it also enjoys a double life as a thing-in-itself that can be, for example, eaten, and that only *potentially* bears information about itself *to the right kind of user*. In order to investigate what makes for the right kind of user, it would be helpful to know more about what they are hypothesized to be using. Sloman and Logan (1998) recommend that we keep an open mind:

> We make no presumption that information-processing mechanisms must all be computational (whatever that means). Nor is there a commitment regarding *forms* used to encode or express information. They may include logical databases, procedures encoding practical know-how, image structures, neural nets or even direct physical representations, as in thermostats and speed governors. (p. 2)

What is common to each of these examples (except perhaps neural nets) is that they involve representation.[68] As documented in Chapters 2 and 3, information-processing psychology involves a conflation of Shannon's (1948) information theory with representational theories of mind. This misperception persists, for example, Simoncelli and Olshausen (2001) write that:

> more than 40 years ago, motivated by developments in information theory, Attneave (1954) suggested that the goal of visual perception is *to produce an efficient representation* of the incoming signal. In a

[68] However such alternatives to representation are "already classified as representations by broad-minded thinkers" (Sloman, 2011, p.13). It appears that 'representation' is every bit as flexible as information.

> neurobiological context, Barlow (1961) hypothesized that the role of early sensory neurons is to remove statistical redundancy in the sensory input. Variants of this "efficient coding" hypothesis have been formulated by numerous other authors…. (p. 1194, emphasis added)

However, as demonstrated in Chapter 2, the concept of an 'efficient representation' is just not an information-theoretic concept, although the superficially-similar concept of 'efficient coding' is. There appears to be widespread confusion among information-processing psychologists regarding the distinction between Shannon's (1948) statistical analyses of signal sequences on the one hand and information-processing psychology's representational theory of mind on the other. How does Sloman describe the relation between 'information' and 'representation'?

Information and Representation

Sloman (2011) *"…uses representation' to refer to any kind of information bearer.* (p. 3).[69] What do we know about such representations? Sloman claims that:

> What characterizes a form of representation is a collection of primitives, along with ways of modifying them, combining them to form larger structures, transformations that can be applied to the more complex items, mechanisms for storing, matching, searching, and copying them, and particular uses to which instances of the form can be put, e.g. controlling behavior, searching for plans, explaining, forming generalisations, interpreting sensory input, expressing goals, expressing uncertainty, and communication with others. (p. 14)

Sloman helpfully provides some more concrete "example[s] [of] information-bearers explicitly used by humans…sentences, maps, pictures, bit-strings, video recordings, or other more abstract representations of actual or possible processes" (p. 13). But it is very difficult to understand how a picture can be reasonably construed as consisting in "a collection of primitives, along with ways of modifying them, combining them to form larger structures, transformations that can be applied to the more complex items,

[69] However, as Bickhard (2000) observes, "Representation, in fact, is commonly considered to be a special kind of information. It must be a *special* kind, because otherwise all of the myriad instances of informational relationships in the universe would be representational" (p. 1).

mechanisms for storing, matching, searching, and copying them…" To be clear, pixilated *digital images* and/or *video* might accord with this description; as demonstrated in Chapter 2, how to render images in just such a mechanically reproducible manner was among the explicit concerns of Hartley (1928) and Shannon (1948). It is worth considering how a non-pixilated, hand-drawn sketch of, say, a smiley face could be viewed as meeting these criteria. But even if we generously accept that images can be fairly characterized as a collection of primitives (e.g., points, lines, colors, etc.), the possibility of such a componential analysis is not what *makes* a smiley face drawing *represent* a smiley face. Rather, it is the fact of being viewed (or described, interpreted, regarded, etc.) *as* a smiley-face that imbues the smiley face drawing with its representational property.

Sloman (2011) believes that empirical research might someday discover new forms of representation:

> There probably are many more forms of representation (more types of information-bearer) than we have discovered so far… The continued investigation of the space of possible forms of representation, including the various options for forming more complex information contents from simpler ones, and the tradeoffs between the various options, is a major long term research project. (p. 15)

In claiming that it is possible to discover new forms of representation, Sloman might be lampooned as claiming that linguists might discover new words, previously unknown to any speakers.

Bearing in mind that Sloman (2011) "uses 'representation' to refer to any kind of information bearer" (p. 03), consider his apparently reasonable assumption that:

> Information cannot play a role in any process unless there is something that encodes or expresses the information: an 'information bearer' (B), and some user (U) that takes B to express information I (i.e. interprets B). The same bearer B may be interpreted differently by different users, and the same user, U may interpret B differently in different contexts (C). (p. 3)

Accordingly, to discover new forms of representation, or "information-bearers" would be to discover new forms of information users, that is, representation users. Given how little Sloman claims that we know about representation, it is surprising to encounter his claim that "*known* examples…[of] information bearers in biological systems…include chemical structures and patterns of activation of neurons" (p. 13, emphasis added). For the sense in which it is supposedly "*known*" that "chemical structures and patterns of activation of neurons" bear information is precisely what is at issue when Sloman (2011) asks the question (*What's information for an organism or intelligent machine?...)* that begins his paper.

It is not the case that sensory signals, chemical structures, and patterns of activation of neurons are *known* to bear information; rather, they are *assumed* to do so, although the "nature of psychological information and its structure" are of course, *unknown* (Lachman et al, 1979, pp. 74-75). Accordingly, Sloman (2011) seeks an implicit definition for this sense for 'information.' This leads him to confess:

> Some of the most important and least well understood parts of a theory about information are concerned with the variety of roles it plays in living things, including roles concerned with reproduction, roles concerned with growth, development, maintenance and repair, roles concerned with perception, reasoning, learning, social interaction, etc. (p. 25)

So, it is most definitely *not* known that patterns of neuronal activation bear information, although it *is* known that such patterns of neuronal activation are causally related to *informed* behaviors, such as perceiving a food source, as well as behavioral capacities that do not involve information in any obvious way, such as the capacity to write a poem. The issue at hand is, how suited is the "somewhat impressionistic" sense of 'information' to the role of *explaining* the relationship between such neurological and behavioral phenomena, respectively (Dennett, 2005, p. 14)?

The Homunculus Fallacy

Who or what "user…takes" sensory signals, chemical structures, and patterns of activation of neurons "to express information" (Sloman, 2011, p. 18)? Sloman alludes to, but does not identify by name or citation, an argument known as *the homunculus fallacy* (Kenny, 1972). Briefly, this is the error of supposing that the epistemic and/or agentive abilities of a system can be explained by reference to the epistemic and/or agentive abilities of a *part* of that system, typically imagined to be *inside* that system in the manner that we regularly characterize thoughts and feelings as *inside*. Dennett (1991, 2005) rightly caricatures this as the assumption of a "Cartesian Theater…where an inner show of remarkable constructions is put on parade for a (material) *res cogitans* sitting in the audience" (Dennett, 2005, p. 02).[70] The positing of a homunculus raises a problem of infinite regress. For example, if it is supposed that understanding utterances involves unconsciously translating spoken languages into a language of thought consisting of mental representations (Fodor, 1975), then this raises the question of how the mentally-represented language of thought is understood.[71]

Sloman (2009, 2011) counter-claims that in characterizing sensory signals, chemical structures, patterns of activations of neurons, and so on as bearers of 'information', one is not thereby assuming a full-blown information user (homunculi) somewhere inside the organism. He defends this practice by cleaving between 'information use' and 'rationality', "Living things use informed control. *This does not presuppose rationality.* We can study living functional control, systems using the designer stance (McCarthy) without adopting the 'intentional stance' (Dennett)..." (Sloman, 2009, Slide 12, emphasis added).

Not only are we free to do this, according to Sloman (2011), but in order "to understand biological organisms and design sophisticated artificial systems, we *need* what McCarthy…labels 'the designer stance'" (p. 23, emphasis added). What are

[70] Dennett (2005) supposes that this fallacy is avoided by a "quite banal and uncontroversial" practice of assuming that information-processing talk is "*not* of (personal level) *experience* but of [a "subpersonal process"], say, *data from the ventral stream*" (p. 14). But it is the clarity of just such subpersonal uses that is at issue. It is admirable that Dennett draws this distinction but it is not clear how simply observing the distinction resolves any ambiguity in the subpersonal cases.

[71] A classic defense is to claim that the internal language has causal and semantic properties such that mental representations do not require interpretation in the manner that non-mental ones do (e.g., Fodor, 1975, 2008).

McCarthy's (2008) *designer stance* and Dennett's (1978) *intentional stance*, respectively? McCarthy clarifies that the:

> designer stance is related to Daniel Dennett's design stance, but Aaron Sloman has persuaded [him] that [he] was not using it quite in the way that Dennett used design stance…In so far as we have an idea what innate knowledge of the world would be useful, AI can work on putting it into robots, and cognitive science and philosophy can look for evidence of how much of it evolved in humans. This is the *designer stance*.
> (p. 2004)

Whatever McCarthy's (2008) designer stance may have to recommend it as an AI research strategy, it does not function to support Sloman's (2009, 2011) arguments because it too simply presupposes the coherence of the information-processing vocabulary. For example, McCarthy's (2008) assumptions include that "human performance is limited by how slowly we process information. If we could process it faster we could do better, and people who think faster than others have advantages" (p. 2006). With the above-noted confusion between *capabilities* and *mechanisms* in mind, consider that what counts as a person's mind and/or brain 'processing information faster' is that *the* person *thinks* faster, for example, reacts to stimuli and/or solves problems faster. What is the justification for the assumption that *thinking faster* is identical with *processing information faster*? None is offered. Sloman (2009, 2011), McCarthy (2008), and very many other theorists appear to confuse such theory-laden descriptions of behavioral data with the phenomenon of interest itself, as if the proposition that *certain organisms process information faster than others* is as clear as the proposition that *certain organisms learn faster than others and/or react faster than others*.[72]

Sloman and Logan (1998) make it clear that the Sloman/McCarthy 'design stance' is motivated by the observations that "mentalistic" attributions are "irresistib[ly]…useful":

> no other vocabulary will be as useful for describing, explaining, predicting capabilities and behavior…So, instead of the self-defeating strategy of trying to avoid mentalistic language, we need a disciplined approach to its use… banning [mentalistic terms] altogether from explanatory theories, would be as crippling in the study of intelligent agents as it would be in the engineering design of complex control systems. (p. 1)

It is easy to agree that "mentalistic" language is necessary because such terms constitute the pre-theoretical vocabulary that defines the phenomena of interest, for example, understanding, interpreting, perceiving, reasoning, etcetera. But there is an important difference between objecting to psychological terms in general as would a methodological behaviorist, and criticizing the *misapplication* of those terms. To criticize the *mis*application of psychological terms *presupposes* that they can be coherently applied. So to argue that "banning [mentalistic terms] altogether from explanatory theories, would be as crippling in the study of intelligent agents as it would be in the engineering design of complex control systems" is a red herring (Sloman & Logan, 1998, p. 1). The issue at hand is not *whether* mentalistic language can be validly applied but rather, *how*? In particular, what does 'information' mean when ascribed to neural and/or mental processes?

Having established that no in-principle objection to mentalistic language will be offered here, particular claims regarding *information-processing* are further examined. Sloman and Logan (1998) contrast their position with Dennett's (1978, 2005):

> Dennett (1978) recommends the 'intentional stance' in describing sophisticated robots, as well as human beings. *That restricts mentalistic language to descriptions of whole agents, and presupposes that the agents are largely rational…*By contrast, we claim that mentality is

[72] McCarthy also glides over the differences between ascribing 'memory' to computers and to persons: "Compared to computers, humans have very little short term memory. In writing a computer program it is difficult to restrict oneself to a short term memory of 7±2 items," argues McCarthy, making a passing (uncited) reference to Miller's (1956) famous *Magical Number…*article (p. 2006).

> primarily concerned with an 'information level' architecture…"
> (Sloman & Logan, 1998, p. 1, emphasis added)[73]

Sloman and Logan suppose that they have developed a justification that frees psychologists from "restrict[ions of] mentalistic language to descriptions of whole agents [that]…are largely rational" and that information is a commanding concept in this logical liberation.

What Is a Stance?

In order to understand Sloman and Logan's (1998) argument, it is helpful to examine Dennett's "intentional stance" from which Sloman seeks to distinguish his own views.[74] Dennett argues (2005) that:

> it is an empirical fact, and a surprising one, that our brains–more particularly, *parts* of our brains–engage in processes that are *strikingly like* guessing, deciding, believing, jumping to conclusions, etc. And it is *enough* like these personal level behaviors to warrant stretching ordinary usage to cover it. (p. 11)

At first blush, this appears to be consistent with Sloman's views, so it is easy to appreciate how McCarthy (2008) initially neglected the subtle differences between Sloman and Logan's (1998, and also Sloman, 2011), and Dennett's (1978, 2005) positions. Although both sanction the ascription of 'information use' to parts of organisms, there is an important difference between how they defend this practice. Dennett (2005) claims that there are similarities between the information-use of whole agents and the informational process that information-processing psychologists ascribe to agent's *parts*. He has:

[73] Sloman also mentions Newell's (1982) work in this passage. It is beyond the scope of this project to complete an exhaustive review of the similarities and differences among Dennett's (1989), Newell's (1982), and Sloman's views, but because Sloman contrasts his arguments with Dennett's in multiple places, aspects of Dennett's 'intentional stance' are examined here.

[74] See Bennett and Hacker (2003) for a critical analysis of Dennett's notion of 'stances.'

> defended such uses of the intentional stance in characterizing complex systems ranging from chess-playing computers to thermostats, and in characterizing the brain's subsystems at many levels, The idea is that when we engineer a complex system (or reverse engineer a biological system like a person or a person's brain), we can make progress by breaking down the whole wonderful person into sub-persons of sorts, agentlike systems that have *part* of the prowess of a person, and then these *homunculi* can be broken down further into still simpler, less person-like agents, and so forth–a *finite*, not infinite, regress that bottoms out when we reach agents so stupid that they can be replaced by a machine. (p. 12)

In other words, Dennett recommends a reductionistic strategy that he supposes can bridge the gap between rational and mechanical phenomena. He assumes that being "stupid" is intrinsically dehumanizing and that therefore, the simple decisions of stupid agents are essentially identical with the mindless mechanical processing of a machine (Dennett, 2005, p. 12). There are aspects of this argument that make sense. For example, rational agenthood does come in degrees. So, for example, the limited problem-solving abilities of non-human organisms (and/or, for more liberal thinkers, machines) could be offered as evidence that they, like human beings, are rational agents, only differently or less so. But Dennett's (1994) claim is that it is valid to attribute "intentional" (i.e., roughly, psychological) predicates to *parts of* organisms. What does he offer the critic who cries *homunculus fallacy*?

> One may be tempted to ask: Are the subpersonal components *real* intentional systems? At what point in the diminution of prowess as we descend to simple neurons does *real* intentionality disappear? *Don't ask.* The reasons for regarding an individual neuron (or a thermostat) as an intentional system are unimpressive, but not zero, and the security of our intentional attributions at the highest levels does not depend on our identifying a lowest level of real intentionality. (p. 240)

This constitutes a dodge, not a defense (*Don't ask…),* followed by a red herring. *Of course* the "security of our intentional attributions" to whole organisms (or, whole systems) "does not depend on our identifying a lowest level of real intentionality." For the security of "intentional" attributions to whole organisms, is *presupposed* by wondering what the "lowest level of real intentionality" is. The security of whole-organism intentional attributions is a consequence of it being possible to explain what such attributions *mean*. And it *is* possible to explain what it means to attribute beliefs to a whole organism, such as a human being. To borrow an example (that is criticized below) from Fodor (1981): “anyone who believes John is tall is very likely also to believe someone is tall...And anyone who believes everyone in the room is tall and also believes John is in the room will very likely believe John is tall” (p. 95). The cliché *seeing is believing* also helps to explain what it is to believe *that p*. Fodor’s example describes paradigmatic relationships among particular beliefs whereas *seeing is believing* describes the relation between (one modality of) perceiving and the phenomena of believing. These examples do not constitute scientific explanations of the capacity to believe. Rather, they identify the class of phenomena that a scientific explanation of believing would have to explain. It would be absurd to suppose that the security of attributing a belief that John is tall to Jack depends upon establishing that certain of Jack’s brain cells also believe that John is tall.

Dennett’s (1994, 2005) argument is framed in terms of ‘intentionality’ whereas Sloman and Logan’s (1998, Sloman, 2011) leans on ‘rationality.’ To translate Dennett into Sloman, it appears to be fair to say that Dennett (1994, 2005) believes that information can be involved in both rational and non-rational, mechanical processes. The perceived behavioral similarity between “stupid” intentional agents and mechanisms is somehow sufficient to convince Dennett that “don’t ask” constitutes a satisfying response to the questions guiding this analysis. Much to his credit, Sloman (2009, 2011) does ask such difficult questions. However, there are problems with his responses, as well as with the assumptions that he seeks to defend. For example, Sloman (2009) claims that “most animals are neither rational nor irrational: yet they process information—they acquire, use, store, transform, modify, combine, and in some cases communicate it” (Slide 13). He goes on to argue that “for most organisms, evolution removed the need to be rational, by providing genetically determined mechanisms for selecting among possible alternatives in all circumstances. *Adopting the intentional stance towards their information-processing is a form of anthropomorphism*” (Slide 13, emphasis added).

There are myriad problems with this argument. First, if the issue is framed in terms of 'intentionality,' then the substantive question is not whether it is advisable to adopt "the intentional stance towards…information-processing" (Sloman, 2009, Slide 13), but whether or not ascribing information-related processes is *intrinsically* intentional. That is, does ascribing information-related processes entail that the subjects of such ascriptions have the capacities to be both informed and/or *un*informed, to *know* and/or not know? Sloman (2009) anticipates this objection, which he views as committing a converse crime of anthropomorphism. But this too is a dodge, not a defense. *Don't adopt an intentional stance towards information-processing* is not an answer to the question: *Is ascribing information-processing intrinsically intentional?*

Yet, Sloman (2011) is correct in intuiting that 'anthropomorphism' is relevant. For it is relatively clear *what it is* for a human being, and/or that which resembles a human being (Wittgenstein, 1953) to use representations and/to "acquire, use, store, transform, modify, combine, and…communicate [information]" (Sloman, 2011, p. 13). In contrast, it is quite mysterious what *counts as* "animal brains us[ing] *still unknown forms of representation* to encode information about the environment..." (pp. 13-14, emphasis added). And it is not the anthropocentrism of the critic that prevents the information-positing psychologist from identifying what *counts as* a neurological and/or mental structure processing information. For if "information cannot play a role in any process unless there is…an information bearer (B), and some user (U) that takes B to express information I (i.e., *interprets* B)" (p. 3, emphasis added) then the critic who cries anthropomorphism is right and ascribing information-related processes *presupposes* paradigmatically human interpretive abilities. Sloman's (2009) recommendation to avoid an "intentional stance towards…information-processing" (Slide 13) is thereby akin to recommending that an object be painted red but left uncolored.

On Coin Sorting and Control Information

There are probably very many automatic, mechanical processes that can be coherently described in information-related terms. For example, one automatic coin-sorting machine might be said to use information about the sizes of coins in order to sort them whereas another design might use information about the mass of coins in order to sort them. There does not appear to be any definitional difficulty associated with the use of ‘information’ here because actual mechanisms that use information about the size and/or mass of coins to sort them can be identified and described. It might be helpful to clarify that, with respect to *this sense* of ‘information use,’ the coin-sorting machine cannot *not* use information about the size of the coins which it receives from depositors (although it can stop working). The use of information by a coin-sorting machine is different from the use of information by an agent that might also have be *un*informed, for example, the inheritor of a coin collection might be ignorant as to the origins of very many of the coins she inherits.

It is better to seek particularity and distinction rather than unification and generality here. For if we are seeking a unifying definition of ‘information’ then we are faced with the impossible task of reconciling incompatible claims. If it is accepted that “information cannot play a role in any process unless there is …an information bearer (B), and some user (U) that takes B to express information I (i.e., interpret*s* B)” (Sloman, 2011, p. 3) then it is misleading to conceive of mechanisms as using information. What is needed is conceptual *analysis*, not conceptual *synthesis*.

Multiple Distinctions versus Metaphysical Dichotomies

Sloman (2009, 2011) and Dennett (1978, 1991) suppose that in establishing that certain instances of *non-rational* (Sloman) or *subpersonal* (Dennett) information-predication are sensible, they have solved a class of general, dichotomous problems, which solution licenses the ascription of information-use far and wide. But it is preferable to avoid metaphysical dichotomies and instead focus on the clarity of particular claims and/or examples. Because, for example, it is possible to explain what it means to say *a coin-sorting machine uses information about the masses and/or sizes of the coins it sorts*, describing its sorting mechanism in terms of ‘using information’ does not raise a logical problem. If we accept the premise that both human coin collectors and mechanical coin-sorters, respectively, use information in dealing with coins, must we also conclude that every information-related activity that can be predicated to rational agents can also be predicated to mechanical processes? Clearly not. But this is easily confused when it comes to neurological mechanisms that are regarded as *parts of information*-using organisms.

Information and Mereology

Sloman (2011) baldly dismisses "the philosophical claim that only a whole human-like agent can acquire, manipulate and use information" (p. 23). What exactly is this putatively "false" philosophical claim? Sloman alludes to (but does not cite) a line of argument crystallized by Bennett and Hacker's (2003) identification of what they call the *mereological fallacy* in neuroscience. 'Mereology' is the study of part/whole relations, and mereological fallacy is that of attributing to *parts* of an animal attributes that are properties of the *whole* being. This critique was famously and poetically expressed by Wittgenstein (1953): "Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious (para. 281). That is, it is clear what *counts as* a human being—and/or *what resembles a human being*—having sensations, seeing, hearing, and so on.

Now consider Sloman's (2011) contention that an "information-user can have parts that are information users" and his denial of "the philosophical claim that only a whole human-like agent can acquire, manipulate and use information" (p. 23). It is only natural to ask what *counts as* a neurological and/or mental structure using information. We know what counts as an *organism* using information—for example, a hunting dog following a trail of scent, a person referring to a map or diagram, looking up a word in a dictionary, and/or checking a thermometer all count as 'using information.'[75]

Who Can V? versus What Counts as Ving?

I argued above that Sloman (2009, 2011) was correct in identifying *anthropocentrism* as relevant to questions concerning information-predication. However, as discussed above, it is also easy to identify cases for which attributing information-use to a mechanism, such as an automated coin-sorter is no less intelligible. There is an important difference between clarifying what *behaviors count as V*ing and stipulating on non-behavioral, in-principle grounds who-or-what subjects can *V.*[76] Although Sloman (2011) baldly asserts that "the philosophical claim that only a whole human-like agent can acquire, manipulate and use information is false," he does not provide any

[75] Of course, it is not *always* possible to be *certain* of whether or not an observed behavior counts as 'information use' in any particular case. For example, consider a very prosaic situation of observing a person as she consults a thermometer. Typically, this will result in the observed party having acquired some information about temperature. However, it is also possible for a person to be preoccupied and fail to acquire the information that she might have otherwise acquired, had she not been so distracted. If a preoccupied individual walks away from the thermometer she just consulted without having learned the temperature, then she has not become informed, and/or used and/or acquired any information. However, to an observer, unaware of her distracted state, she would likely *appear* to have learned what the temperature is and, thereby, have acquired information about the temperature. But she has not, and so she appears more informed than she, in fact, is. This exemplifies what, in technical terms, is called the *defeasibility of behavioral criteria:* that behavioral observations license the application—but do not guarantee the accuracy—of such epistemic predicates as 'becoming informed' and/or 'using information' (Baker & Hacker, 1984b)

[76] As an example of a legitimate application of epistemic capacities to agents that were previously believed to lack such abilities, on the basis of empirical evidence, Bennett and Hacker (2003) observe that "Susan Savage-Rombough has produced striking evidence to show that bonobochimpanzees, appropriately trained and taught, can ask and answer questions, can reason in a rudimentary fashion, give and obey orders, and so on. The evidence lies in their behavior – in what they do (including how they employ symbols) in their interactions with us. This was indeed very surprising. For no one thought that such capacities could be acquired by apes. But it would be absurd to think that the ascription of cognitive and cogitative attributes to the brain rests on comparable evidence. It would be absurd because we do not even know what would show that the brain has such attributes." (p. 72) We know what it is for a person and/or a chimpanzee to use information, for a book to contain information, for a map to present visual information, etcetera. But we do not know what it is for neural structures or 'information-processing architectures' to use information.

argumentation against such mereological criticisms. He even identifies mereological concerns of his own:

> An information-user can have parts that are information users. This leads to complications such as that a part can have and use some information that the whole would not be said to have. E.g. your immune system and your digestive system and various metabolic processes use information and take [*sic*] decisions of many kinds though we would not say that you have, use or know about the information. (p. 23)

This "complication" is a consequences of accepting the premise that "an information-user can have parts that are information users" in the absence of an explanation of what information-for-parts *is*.

Ironically, although Sloman's (2011) paper asks *What's information for an organism or intelligent machine…*the substantive issue concerns what information is for *parts* of organisms, such as neurological structures and processes, as well as mental ones, although mental properties might be more clearly described as an *aspect* of an organism rather than as a *part* of it. The impossibility of explicating what *counts as* information in the information-processing-psychology-sense led Sloman (2011) to search for an implicit definition. By the end of his investigation, his aspirations are considerably more modest. He concludes that "the ideas sketched here help us to focus more clearly on aspects of information processing that are not yet understood" (p. 28). With this, the reader can easily agree. I, in turn, hope that the remarks above help to demystify this lack of understanding.

No Implicit Definition Found

Sloman (2011) argued that the 'information' in 'information-processing psychology' might be implicitly defined by the manner in which:

> the current scientific community well-educated in mathematics, logic, psychology, neuroscience, biology, computer science, linguistics, social

> science, artificial intelligence…and philosophy… currently understand and use the word 'information.' (p. 27)[77]

He optimistically supposes that "as regards information-processing our state of knowledge could be compared with Galileo's knowledge of physics" (p. 28). But the content of his investigation demonstrates that this is wishful thinking.

[77] I have deleted "physics" and "cosmology" from this quote because both my investigations and Sloman's (2011) are particularly concerned with the 'information' in 'information-processing psychology.'

Chapter 5.

Conclusion

What facts are there to face about 'information' in psychology? This investigation revealed that (1) There are widespread myths regarding the degree of substantive continuity between Shannon's (1948) information-theory and information-processing psychology and (2) that Miller's (1951, 1953, 1956) celebrated and pioneering works in information-processing psychology involve serious distortions of Shannon's (1948) information-theory.

In Chapter 2, sources of apparent encouragement to information-related confusion were identified in the works of authors whose names are most closely associated with the phrase information theory (Hartley, 1928, Weaver, 1949; Wiener, 1948, 1950). Certain of Shannon's misstatements were also clarified. Overall, however, the aim of this section was to underscore Shannon's (1948) clarity and establish that *his* information theory involves, precisely-defined concepts whose range of application is sharply bounded. A working definition of the everyday non-technical sense of 'information' was also developed. It emphasized the epistemic rather than semantic character of the everyday sense information, as a corrective to the idea that Shannon's (1948) *H* statistic constitutes a quantified and de-semanticized version of the everyday sense of 'information.[78]

In Chapter 3, a variety of specific errors ((a)-(k)) were identified in the works of George Miller. These investigations emphasized that certain, but definitely not all,

[78] For example, as exhibited in Miller's (1953) contention that in computing *H* values, "only the *amount* of information is measured—the amount does not specify the content, value, truthfulness, exclusiveness, history, or purpose of the information" (p. 3).

psychological experiments can be coherently described in Shannon's (1948) information-theoretic terms. In particular, it was demonstrated that one of the more celebrated articles of 20th Century psychology (Miller, 1956) involves a cluster of confusions regarding the distinction between Shannon's statistical concept of 'information' and the everyday epistemic sense of the term. In particular, the idea that statements, perceptions, and/or sensations can be coherently viewed *in general* as 'messages,' in the information-theoretic sense was sharply criticized. I emphasized the discontinuity between Shannon's information theory and information-processing psychology. Finally, it was demonstrated that in his subsequent commentary on his famous 1956 paper, Miller's (2003) self-assessment supported my own critique of that article.

In Chapter 4, the idea that it is acceptable (e.g., Dennett, 2005), inevitable, and/or even preferable (Sloman, 2011) that the 'information' in 'information-processing psychology' remains undefined was sharply criticized. Sloman's (2011) recent argument that information-processing psychology's sense of 'information' might *be implicitly defined* was thoroughly investigated. It was demonstrated that Sloman exhibits a degree of optimism that is unsupported by the content of his investigations. Furthermore, I strongly objected to his contention that empirical research might reveal what is meant by the 'information' in 'information-processing,' on which point I regard Sloman's views as representative of 'information'-processing psychologists in general. I emphasized the distinction between logical and empirical questions, and argued that determining *what counts as* information-processing in particular cases can only be resolved through *decision*, and not by *discovery*.

In summary, the investigations above demonstrate that there are widespread misperceptions about the history of information-processing psychology, lamentable apathy towards its current, definition-less state of affairs, and unjustified, wishful thinking regarding the future of 'information' as a term of art in the psychological sciences.

Why face facts about 'information' in psychology?

Finally, I will describe the context in which I chose to perform the investigations summarized above. *Why face facts about 'information' in psychology?* I'll speak to that question by describing some memorable, stimulating discussions. In the spring of 2008, through a series of conversations, Bill Turnbull of Simon Fraser University challenged me to explain something to him as well as to myself. It only took me 6 years to come up with this response. Bill asked me how I could simultaneously hold what he perceived to be potentially incompatible beliefs. I will personify these beliefs by using a figure of speech and drawing a contrast between The Critic and The Participant.

On the one hand, The Critic believes that it is astonishing that the concepts of ‘information’ and ‘processing’ in ‘information-processing psychology’ are undefined. Why is that astonishing? Because statements of the form *I study the Xing of Y but I don’t know what ‘X’ or ‘Y’ means* are internally contradictory. Now, ambiguity and absurdity can obviously be a virtue in certain contexts, such as in poetry or jokes. Sense is prerequisite for truth; there is no way to evaluate the truth of a claim if it does not make sense. For example, it is neither true nor false that colorless green ideas sleep furiously, (Baker & Hacker, 1984). In more general terms, you can’t test the truth of hypotheses about *Y-Xing* if you can’t identify Xing and Ys. And you can’t identify *Xing or Ys* unless you know what the concepts of ‘*X*’ and ‘Y’ mean. So, the Critic thinks that there is something deeply, and interestingly wrong about psychologists speaking in terms of ‘information-processing’ without being able to say what ‘information-processing’ means.

Yet, on the other hand, as Bill Turnbull observed, The Critic is also, to some extent, The Participant. And what ‘participating in information-processing psychology’ means here is just believing that information-processing psychology, at least at its best, involves genuine, respectable scientific discoveries. That is, according to the Participant, it is possible and perhaps even preferable, at least in some cases, to set aside issues of linguistic precision and appreciate that information-processing psychologists really study real things. What kind of real things? To take just one non-obscure example of seemingly obvious social importance, Dodge and Crick (1990) offer “a social information-processing model of children's aggressive behavior,” and write that

> a child's behavioral response…is a function of five steps of processing: encoding of social cues, interpretation of social cues, response search, response evaluation, and enactment…Empirical studies are described in

> which children's patterns of processing have been found to predict individual differences in their aggressive behavior (p. 8).

In other words, these authors find that aggressive behavior among children is associated with a variety of particular attentional, interpretive, and reasoning biases. Now, if the predictions of the social-information-processing model of aggressive behavior are often correct, doesn't this mean that there must be some truth to the claim that "a child's behavioural response…is a function of five steps of processing" information (p. 8)? And if there is some truth to the claim that "a child's behavioral response…is a function of five steps of processing" (p. 8) information, doesn't this mean that 'information' and 'processing' must mean something? From the point of view of The Participant, it seems like the absence of definitions for 'information' and 'processing' are, at least in some cases, not a big deal.

So, there is an apparent paradox. When The Critic thinks in terms of X's and Y's, claims such as *I study the Xing of Y but I do not know what 'X' or 'Y' mean* appear nonsensical and absurd. And yet, when The Participant thinks in terms of the particular concepts of 'information' and 'processing', it seems to be trivially easy to identify cases for which the presence of apparently nonsensical language does not appear to interfere with the generation of true claims. Bill Turnbull challenged me to identify *how* such apparently true and important claims could be scaffolded to a seemingly nonsensical metatheory involving core terms that are undefined. He asked me certain basic questions that seemed fundamentally important. For example, if employing undefined terms such as 'information' and 'information-processing' is irrational and therefore unacceptable, how can researchers who think in such terms ever manage to discover anything? And, conversely, if one accepts that information-processing psychology involves at least some respectable discoveries, doesn't such acceptance diminish the force of the terminological objections? In other, oversimplified words, if the concept 'information-processing' is so bad for psychology, how can any good come out of it?

Broadly speaking, these were the concerns that motivated my 'information'-investigations. If, I supposed, I knew more about where the 'information' in 'information-processing' came from, then I would be able to figure out how The Critic and The

Participant co-exist. I set out to investigate how the discipline of psychology had developed such that 'information' and 'information-processing' could be regarded as *both* core concepts *and* undefined terms. The Participant believes that, in very many cases, it is quite easy to identify what a particular claim involving 'information-processing' *means*, despite the appearance of the undefined concept of 'information.' How can this be so?

The solution to this definitional problem is surprisingly simple. And, oddly enough, it appears right alongside the problem. For example, consider these words from the late Stanford cognitive scientist John McCarthy (1927-2011): "Human performance is limited by how slowly we process information. If we could process it faster we could do better, and people who think faster than others have advantages" (McCarthy, 2008, p. 2006). In this case, it is easy to see that to *process information faster* means to *think faster*, and the meaning of 'thinking faster' is clear enough. Examples include solving problems faster, or reacting to stimuli faster, or coming up with a snappy retort more quickly. So, the criterion for identifying the speed at which information is processed by a particular organism is identical with identifying how quickly that organism thinks, or, more precisely, how quickly that organism can do one of the very many things that count as thinking. This particular example illustrates a general point explaining how the irrational use of an undefined senses of 'information' can, at least in certain cases, be relatively benign. *If the phenomena of interest can be identified and described in terms that do not presuppose the concepts of 'information' and/or 'information-processing',* then psychologists can identify the phenomena that they subsequently describe in undefined, information-processing terms.

The question then becomes, why posit indefinable informational processes? Fred Dretske insightfully emphasized the aesthetic appeal of 'information' talk, writing that its "use in telecommunications and computer technology gives it a tough, brittle, technical sound" while it remains "spongy, plastic, and amorphous enough to be serviceable in cognitive and semantic studies" (Dretske, 1981, p. viii). Hopefully, my investigations will stimulate critical reflection on what services the concepts of 'information' and 'processing' provide. It would be overstated to claim that these concepts *could not* be of value. But it appears that there are very many reasons to believe that often, they aren't.

References

Adams, F. (1991). The informational turn in philosophy. *Minds and Machines, 13,* 471–501.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*(2), 353-356.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Baker, G. P., & Hacker, P. M. S. (1984a). *Language, sense and nonsense*. Oxford, England: Blackwell.

Baker, G. P., & Hacker, P. M. S. (1984b). *Scepticism, Rules and Language*. Oxford, England: Blackwell.

Barandiaran, X., & Moreno, A. (2008). On the nature of neural information: A critique of the received view 50 years later. *Neurocomputing, 71,* 681–692

Bar-Hillel, Y. (1955). An examination of ‘information’ theory. *Philosophy of Science, 22*(2)*,* 86-105.

Bar-Hillel, Y., & Carnap, R. (1953). Semantic Information, *The British Journal for the Philosophy of Science, 4*(14), 147-157.

Beebe-Center, J. G., Rogers, M. S., & O'Connell, D. N. (1955). Transmission of ‘information’ about sucrose and saline solutions through the sense of taste. *Journal of Psychology, Interdisciplinary and Applied, 39,* 157-160.

Bennet, M. R., & Hacker, P. M. S. (2003). *The philosophical foundations of neuroscience.* Oxford, England: Blackwell.

Bickhard, M. H. (2000). Information and representation in autonomous agents. *Journal of Cognitive Systems Research, 1*(2), 65-75.

Bickhard, M. H. (2003). Some notes on internal and external relations and representation. *Consciousness & Emotion*, 4(1), 101-110.

Broadbent, D. E. (1958). *Perception and communication.* Oxford, England: Oxford University Press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal, 47,* 139–159.

Cherry, C. E. (1957). On validity of applying communication theory to experimental psychology. *British Journal of Psychology, 48*(3), 176–188.

Chomsky, N. (1957). *Syntactic structures.* The Hague, The Netherlands: Mouton.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*(2), 67-90.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Davies, M., & Gardner, D. (2010). *Word frequency list of American English.* Retrieved from http://www.wordfrequency.info/files/entries.pdf

Dennett, D.C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, VT: Bradford Books.

Dennett, D.C. (1987). *The intentional stance* (6th ed.). Cambridge, MA: M.I.T. Press.

Dennett, D.C. (1991). *Consciousness Explained*. Boston: Little, Brown.

Dennett, D. C. (1994). Self portrait. In S. Guttenplan (Ed.), *A companion to the philosophy of mind* (pp. 236-244). Oxford, England: Blackwell.

Dennett, D. C. (2005, October 18). *Philosophy as naive anthropology: Comment on Bennett and Hacker.* Retrieved from http://ase.tufts.edu/cogstud/papers/bennetthacker.pdf

Devlin, K. (2001). Claude Shannon, 1916–2001. *Focus: The Newsletter of the Mathematical Association of America, 21,* 20–21.

Dodge, K. A., & Crick, N. R. (1990). Social information processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin, 16,* 8-22.

Dretske, F. (1981). *Knowledge and the flow of 'information'*. Cambridge, MA: MIT Press.

Dreyfus, H. (1972). *What computers can't do*. New York, NY: MIT Press.

Dupré, J. (1993). *The disorder of things*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese, 28,* 97-115.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1981). The mind-body problem. In M. Eckert (Ed.), *Theories of mind: An introductory reader (*pp. 81-96). Lanham, MD: Rowman & Littlefield.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford, England: Oxford University Press.

Frick , F. C., & Miller, G. A. (1951). *A statistical description of operant conditioning. American Journal of Psychology, 64*(1), 20-36.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46*(5), 373-380.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience: The biology of the mind*. New York, NY: Norton.

Hacker, P. M. S. (1990). *Wittgenstein: Meaning and mind, Volume 3 of an analytical commentary on the philosophical investigations, part I: Essays.* Oxford, MA: Blackwell.

Hacker, P. M. S. (2004). *Of the ontology of belief.* Retrieved from http://info.sjc.ox.ac.uk/scr/hacker/docs/Ontology%20of%20belief.pdf

Hacker, P. M. S. (2007). *The relevance of Wittgenstein's philosophy of psychology to the psychological sciences.* Retrieved from http://info.sjc.ox.ac.uk/scr/hacker/docs/Relevance%20of%20W's%20phil.%20of%20psychol.%20to%20science.pdf

Harris, R. (1981). *The language myth.* London, England: Duckworth.

Hartley, R. V. L. (1928). The transmission of 'information'. *Bell Systems Technical Journal*, *7*(3), 535-563.

Heil, J. (1981). Does cognitive psychology rest on a mistake? *Mind, 90,* 321-42.

Hobbes, T. (1994). Leviathan. In E. Curley (Ed.), *Leviathan, with selected variants from the Latin edition of 1668.* Indianapolis, IN: Hackett.

Information. (2010). Oxford English Dictionary (3rd ed.; Online). Retrieve from http://www.oxfordreference.com.proxy.lib.sfu.ca/view/10.1093/acref/9780199571123.001.0001/m_en_gb0410460?rskey=YMr4VJ&result=1

Jost, J.T. & Gustafson, D.F. (1998). Wittgenstein's problem and the methods of psychology: How 'grammar depends on facts.' *Theory & Psychology, 8*(4), 463-479.

Kenny, A. (1972). The homunculus fallacy. In M. Grene (Ed.), *Interpretations of life and mind* (pp. 65-74). London, England: Routledge & Keegan Paul.

Kenny. A. (1993). *Aquinas on mind.* New York, NY: Routledge.

Klemmer, E. T., & Frick, F. C. (1953). Assimilation of ‘information’ from dot and matrix patterns. *Journal of Experimental Psychology, 45,* 15-19.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Laberge, D., & Samuels, S. J. (1974). Toward a theory of automatic information-processing in reading. *Cognitive Psychology, 6*(2), 293–323.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information-processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Luce, D. R. (2003). Whatever happened to information theory in psychology? *Review of General Psychology, 7*(2), 183–188.

Mandler, G. (2002). Origins of the cognitive revolution. *Journal of History of the Behavioral Sciences, 38*(4), 339–353.

Maraun, M. (1998). Measurement as a normative practice: Implications of Wittgenstein's philosophy for measurement in psychology. *Theory & Psychology, 8,* 435-461.

Maraun, M., & Peters, J. (2005). What does it mean that an issue is conceptual in nature? *Journal of Personality Assessment, 85*(2), 128-133.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, 207,* 187-217.

McCarthy, J. (2008). The well-designed child. *Artificial Intelligence, 172,* 2003-2014.

Mechanism. (2012). *Canadian Oxford English Dictionary* [Online]. Retrieved from http://www.oxfordreference.com.proxy.lib.sfu.ca/view/10.1093/acref/9780195418163.001.0001/m-en_ca-msdict-00001-0042930?rskey=IPNgCv&result=3

Miller, G. (1989). George A. Miller. *In G. Lindzey (Ed.), A history of psychology in autobiography* (pp. 391-418). Stanford, CA: Stanford University Press.

Miller, G. A. (1951). *Language and communication.* New York, NY: McGraw-Hill.

Miller, G. A. (1953). What is information measurement? *American Psychologist, 8,* 3-11.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychological Review*, *63*(2), 1-97.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences, 7*(3), 141-144.

Miller, G. A., & Frick, F. C. (1949). Statistical behavioristics and sequences of responses. *Psychological Review, 56,* 311-324.

Newell, A.(1982). The knowledge level. *Artificial Intelligence, 18*(1), 87-127.

Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning: A study of the influence of language upon thought and of the science of symbolism.* London, England: Harvest/HBJ Book.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America, 24,* 745-749.

Pollack, I. (1953a). The assimilation of sequentially encoded information. *American Journal of Psychology, 66,* 421-435.

Pollack, I. (1953b). The information of elementary auditory displays. II. *Journal of the Acoustical Society of America, 25,* 765-769.

Putnam, H. (1975). *Mind, language, and reality.* Cambridge, MA: Cambridge University Press.

Putnam, H. (1989). *Representation and reality.* Cambridge, MA: MIT Press.

Putnam, H. (2002). *The collapse of the fact/value dichotomy and other essays.* Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60,* 20-43.

Rapoport, A. (1956). Three symposia: Problems in human communication and control; Information theory in psychology; Toward a unified theory of human behavior. *Behavioral Science*, *1*(4)*,* 303-309.

Ritchie, D. (1986). Shannon and Weaver: Unravelling the paradox of 'information'. *Communication Research, 13,* 278-198.

Sacks, H., & Jefferson, G. (Ed.). (1992). *Lectures on conversation, Volumes I and II.* Oxford, England: Blackwell.

Seow, S. C. (2005). Information theoretic models of HCI: A comparison of the Hick-Hyman Law and Fitts' Law. *Human-Computer Interaction, 20,* 315–352.

Shanker, S. S. (1998). *Wittgenstein's remarks on the foundations of AI.* London, England: Routledge.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 379–423 and 623–656.

Shannon, C. E. (1950). Programming a computer for playing chess. *Philosophical Magazine*, *41*(314), 256-275.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus of minus two: A commentary on capacity limitations. *Psychological Review, 101*(2), 357-361.

Simoncelli. E.P. & and Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24,* 1193-1216.

Skinner, B. F. (1953). *Science and human behavior.* New York, NY: The Free Press.

Slaney, K. L., & Maraun, M. D. (2005). Analogy and metaphor running amok: An examination of the use of explanatory devices in neuroscience. *Journal of Theoretical and Philosophical Psychology, 25*(2)*,* 701-720.

Sloman, A. (1999, August 26). *Architecture-based conceptions of mind.* Retrieved from http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-Impsfinal.pdf

Sloman, A. (2009, May 22). *What has life got to do with mind? Or vice versa?* Retrieved from http://www.cs.bham.ac.uk/research/projects/cogaff/talks/maggiefest-aaron.pdf

Sloman, A. (2011, January 23). *What's information for an organism or intelligent machine? How can a machine or organism mean?* Retrieved from http://www.cs.bham.ac.uk/research/projects/cogaff/misc/whats-information.pdf

Sloman, A., & Logan, B. (1998, April 1-4). *Architectures and tools for human-like agents.* Proceedings of the 2nd European Conference on Cognitive Modelling, Nottingham, England.

Sloman, A., & Scheutz, M. (2002, September). *A framework for comparing agent architectures.* Proceedings UKCI'02: UK Workshop on Computational Intelligence, Birmingham, England. Retrieved from http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-scheutz-ukci02.pdf

Tolman, R. C. (1938). *The principles of statistical mechanics.* New York, NY: Dover.

Turnbull, W. (2003). *Language in action: Psychological models of conversation.* New York, NY: Psychology Press.

Uncertainty. (2013). *Oxford English Dictionary* [Online]. Retrieved from http://oald8.oxfordlearnersdictionaries.com/dictionary/uncertainty

Vitello, P. (2012, August 1). George A. Miller, a pioneer in Cognitive Psychology, dead at 92. The New York Times. Retrieved from http://www.nytimes.com/2012/08/02/us/george-a-miller-cognitive-psychology-pioneer-dies-at-92.html?pagewanted=all&_r=0

Weaver, W. (1949). Recent contributions to the mathematical theory of communication. In C. E. Shannon & W. Weaver (Eds.), *The mathematical theory of communication* (pp. 1-28). Urbana, IL: The University of Illinois Press.

Weaver, W. (1955). Translation. In W. N. Locke & D. A. Booth (Eds.), *Machine translation of languages* (pp. 15-23). Cambridge, MA: MIT Press.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine.* Cambridge, MA: MIT Press.

Wiener, N. (1950). *The human use of human beings.* Boston, MA: Riverside Press.

Wittgenstein, L. W. (1953). *Philosophical investigations.* London, England: Blackwell.

Wittgenstein, L. W. (1972). *On certainty*. New York, NY: Harper Collins.