Evidence in Management and Organizational Science: Assembling the Field's Full Weight of Scientific Knowledge Through Syntheses

Denise M. Rousseau
Heinz School of Public Policy and Management
And Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213 USA
Email: rousseau@andrew.cmu.edu
Tel: 1-412-268-8470
Fax: 1-412-268-5338

Joshua Manning
Carnegie Mellon University
Pittsburgh, PA 13213
joshuam@andrew.cmu.edu

David Denyer
Cranfield School of Management
Cranfield University
Cranfield
Bedford MK43 0AL UK
44-1234-750111
david.denyer@cranfield.ac.uk

Prepared for the *Annals of the Academy of Management*, volume 2.

An H.J. Heinz II Professorship supported this writing. Thanks are due to Mark Fichman, John Gibbons, Paul Goodman, Gary Latham, Joe Mahoney, Pietro Micheli, Rochelle Mucha, Jeff Pfeffer, and Sara Rynes for their insights in the development of this chapter. Thanks also to Kat Ranck for her research assistance on this project.

This chapter advocates the good scientific practice of systematic research syntheses in Management and Organizational Science (MOS). A research synthesis is the systematic accumulation, analysis and reflective interpretation of the full body of relevant empirical evidence related to a question. It is *not* a conventional literature review. Identifying where MOS research findings are clear (and where they aren't) is critical to establishing which conclusions science supports. It is also important for identifying contested findings and productive lines for future research. Uses of MOS evidence, that is, the motives for undertaking a research synthesis include scientific discovery and explanation, improved management practice guidelines, and formulating public policy. We identify six criteria for establishing the evidentiary value of a body of primary studies in MOS. We then pinpoint the stumbling blocks currently keeping the field from making effective use of its ever-expanding base of empirical studies. Finally, this chapter outlines a) an approach to research synthesis suitable to the domain of MOS and b) supporting practices to make synthesis a collective MOS project.

Evidence in Management and Organizational Science

It is the nature of the object that determines the form
of its possible science (Bhaskar, 1998, p. 3).

Each discipline has an intact critique of its competitors
(Fischhoff, 1991, p. 844)

Uncertain knowledge is better than ignorance
(Mitchell, 2000, p. 9)

This chapter is motivated by the failure of Management and Organizational Science (MOS) to date to make effective use of its available research evidence. Failure to make full *use* of scientific evidence is a problem both management scholars and practitioners face. This lapse has many causes, addressed in this chapter. Two are central in the failure to use MOS evidence well: 1) overvaluing novelty to the detriment of accumulating convergent findings and 2) the general absence of systematic research syntheses. These two causes are intertwined in that as we shall show the use of research syntheses ties closely with how a field gauges the value of its research. This chapter's subject, the systematic research synthesis, is not to be confused with a conventional literature review, its less systematic, non-representative counterpart. Systematic research syntheses assemble, analyze and interpret a comprehensive body of evidence in a highly reflective fashion according to six evidentiary criteria we describe. The why, what, and how of research synthesis in MOS is this chapter's focus.

The explosion of management research since World War II has created knowledge products at a rate far outpacing our current capacity for recall, sense making, and use. In all likelihood, MOS's knowledge base will continue to expand. We estimate

over 200 peer-reviewed journals currently publish MOS research. These diverse outlets reflect that fact that MOS is not a discipline; it is an area of inter-related research activities cutting across numerous disciplines and subfields. The area's expansion translates into a body of knowledge that is increasingly fragmented (Rousseau, 1997), transdisciplinary (Whitley, 2000), and interdependent with advancements in other social sciences (Tranfield, Denyer & Smart, 2003). The complicated state of MOS research makes it tough to know what we know, especially as specialization spawns research communities that often don't and sometimes can't talk with each other.

The Danger We Face

Organizational researchers who adopt public positions regarding management practices and organizational decisions risk making embarrassing gaffes. The reason is simple: it is difficult to be well-informed on an issue when no systematic summary of the relevant evidence exists. A case in point is the claim that incentive pay is not an effective motivator of individual job performance (cf. Pfeffer, 1998). This assertion contrasts with a descriptive literature review by Rynes, Gerhart and Parks (2005). They conclude that incentive pay does in fact tend to increase individual performance (cf. Lawler, 1971). Rynes and colleagues further contend that the evidence supports two mechanisms through which pay impacts individual job performance. The first is an immediate incentive effect where increased performance is motivated by a targeted reward. The second is a more sustained sorting effect where people with higher abilities and motivation self-select into workplaces with performance-based rewards. In defense of all who make good faith efforts to deploy MOS research on the public's behalf, we note that

no one has yet provided the one critical contribution that could resolve the pay-performance dispute--a systematic research synthesis (this chapter's focus). A systematic synthesis musters the full and comprehensive body of available evidence to provide the best-available answer to a question of interest. Such synthesis is uncommon in MOS, placing its scholars and educators in the undesirable position of making purportedly science-based claims to the public, to students, and in textbooks, without the benefit of complete and trustworthy information regarding research findings.

Other fields also have faced this problem. Recall the false consensus regarding fat and heart disease in the 1980s:

> "In 1988, the surgeon general, C. Everett Koop, proclaimed ice cream to be a public-health menace right up there with cigarettes. Alluding to his office's famous 1964 report on the perils of smoking, Dr. Koop announced that the American diet was a problem of 'comparable' magnitude, chiefly because of the high-fat foods that were causing coronary heart disease and other deadly ailments. He introduced his report with these words: 'The depth of the science base underlying its findings is even more impressive than that for tobacco and health in 1964.' That was a ludicrous statement, as Gary Taubes demonstrates in his new book meticulously debunking diet myths, 'Good Calories, Bad Calories' (Knopf, 2007). The notion that fatty foods shorten your life began as a hypothesis based on dubious assumptions and data; when scientists tried to confirm it they failed repeatedly. The evidence against Häagen-Dazs was nothing like the evidence against Marlboros." (Tierney, 2007)

If the evidence were weak, why would the American Surgeon General take such a position? Why did hundreds of popular heart health books follow it? Why did it prompt search for fake fats avoiding the purported bad effects of the real thing? It appears that the Surgeon General was expressing the consensus on the part of medical doctors, nutritionists and public health specialists, who themselves were relying on secondary sources such as popular books and trade articles rather than primary studies. No systematic review then existed to confirm or refute the Surgeon General's claim. When the U.S. National Academy of Science (NAS) ultimately completed such a review, both the government and the public rejected its findings. It's important that NAS checked out Koops' claims, an action contributing to the shift toward evidence-based medicine (Sackett, et al., 1996; 2000). However, this case also demonstrates that beliefs based on an ill-informed *use* of research can be tough to counter.

Uses of evidence in MOS as well as other applied fields run the gamut from scientific discovery and explanation to management education and practice guidelines, and on to the formulation of public policy. The quality and efficacy of any use of evidence depends on availability of carefully conducted research syntheses.

Research Syntheses are the Way Forward

This chapter calls for adoption of the good scientific practice of systematic research syntheses. Systematic means comprehensive accumulation, transparent analysis, and reflective interpretation of all empirical studies pertinent to a specific question. Reliance upon any sampling or subset of the literature risks misrepresenting its diversity in findings, outcomes methods, and frames of reference. The pertinent complement of

evidence typically is available via the internet, readily to scholars with electronic library access and ties to the broad scientific community. Systematic research syntheses evaluate a field's knowledge claims while recognizing omissions, limits, and untested assumptions. Syntheses separate replicable findings from noise. If a synthesis were available to identify where stable effects do in fact exist, researchers would be less likely to interpret apparent inconsistency as intractable complexity (Schmidt, 1992). Importantly, syntheses can also yield insights unrecognized by the original researchers or other less systematic literature reviews.

Systematic research syntheses are important too as quality control. Peer-review serves more as a check on a primary study's published report. The original data themselves seldom are subject to scrutiny. We note that in a multi-discipline survey of American scientists (Martinson, 2005) 0.3% admitted to falsifying data (in comparison 8.0% report ignoring human subject protocols). Although ethical violations may be few, any breach is cause for concern, another reason for cumulating studies and comparing their patterns of results before drawing conclusions.

Most MOS research is never subjected to careful accumulation, analysis or reflection. Traditional literature reviews are skewed by familiarity and availability bias and the implicit preferences of the reviewer (Goodstein & Brazis, 1970; Tranfield et al., 2004). The descriptive literature review typical of MOS can give short shift to alternative points of view, related research from other disciplines or methodologies. It commonly fails to fully represent the research relevant to the question at hand (Salipante, Notz & Bigelow, 1982; Tranfield, et al., 2003). A case in point, approaches such as meta-analysis combining quantitative indicators (Hedges & Okin, 1985) or meta-ethnographies

integrating qualitative findings (Estabrook, Field & Morse, 1994; Noblit & Hare, 1988) are limited in that each ignores the other (potentially leading to different conclusions, cf. Guzzo, Jackson & Katzell, 1987).

Without systematic research syntheses, MOS is in real danger of knowledge loss. As time passes, earlier studies become less likely to be appropriately interpreted and integrated into current thinking (Tranfield et al, 2003). Limited and unsystematic use of the available evidence base plays a role in the dilemmas MOS faces regarding its training of doctoral students and their development of successful academic careers (Zammuto & Connolly, 1984; Pfeffer, 1993; Glick, Miller & Cardinal, 2007; Rousseau, 2007). The field's newcomers have difficulties choosing productive research topics and identifying ways to make useful contributions. When it is tough to determine what is known and what isn't, it even tougher to know what's important to know next.

This chapter's goal is to promote the good scientific practice of research synthesis as an aid to knowledge accumulation and, potentially, more rapid advancement of MOS as a field. In doing so, we address key stumbling blocks hampering uptake of research syntheses in MOS and ways to overcome them. We next examine alternative forms of syntheses various fields currently use. Lastly, we formulate a framework for research syntheses to better deploy MOS research in answering scientific and practical questions.

The Meaning of Evidence in MOS

MOS is a practically-oriented broad-ranging social science (Whitley 2000). It encompasses theory and research on organizations and associated human behaviors, as workers, managers, and customers. It yields an array of products from *facts* about

workers and organizations (e.g., their characteristic properties and processes) to *tools* based on scientific knowledge (e.g., psychometric tests, management science algorithms).

Evidence is the essence of human knowledge. It deals with the regularities our senses and measuring tools can detect. Scientific evidence is knowledge derived through controlled test and observation. It differs from the firsthand evidence people derive from experience or the testimony others provide. Firsthand evidence and testimony play important roles in management practice; so does analyzing business data in making managerial decisions (Davenport, 2006; Pfeffer & Sutton, 2006). Evidence-based management is the complimentary use of scientific evidence and local business evidence. The former is difficult for practitioners to access because as yet MOS evidence is seldom organized in a way that lets would-be users know what the scientific evidence supports.

Science seeks general knowledge, explanations that make useful predictions about a common reality in a replicable fashion. It creates theories that explain regularities in our natural and human-made world (Bogen & Woodward, 1988). Regularities like sunrise and sunset are directly observable. Indeed some regularities may be deemed so obvious that no empirical study is needed. (N.B. This is a reason why no randomized controlled trials have been conducted to see if jumping out of an airplane wearing a parachute prevents death, Smith & Pell, 2003). Most aren't quite so obvious. Establishing that a phenomenon is potentially real and meaningful is science's critical first step (Merton, 1987). This often is accomplished via observations that lead to formulating a theory that can be tested. In MOS for example, descriptive research has identified the existence of work/nonwork boundary issues firms and employees face (Hochschild, 1997; Perlow, 1997). Theory can then be formulated and tested to account

for certain observed regularities, such as why successful workers who negotiate special arrangements to manage work/family conflict might subsequently experience far poorer career outcomes (Hornung, Rousseau & Gleser, in press). In the process of developing and testing theory, the systematic accumulation of empirical observations constitutes the evidence for judging the theory's merit.

Evaluating evidence involves six basic criteria. The first requirement of evidence is C*onstruct Validity*. To establish that a phenomenon is potentially real and meaningful basically means that the regularities scientists use to identify it can be consistently demonstrated. For example, organizational commitment on the part of employees has been conceptualized lots of different ways, but primarily as a psychological attachment to the organization. To establish it as a real and meaningful phenomenon, purported observations of commitment must be consistent with its conceptualization as a psychological attachment (displaying attachment features like goal acceptance and desire to remain). The potential reality of commitment is supported when employee reports of their commitment coincide with their support for and continued employment in the organization. Concomitantly, the meaningfulness of commitment as a phenomenon in its own right requires that its characteristics be distinct from other potentially related phenomena, such as work ethic or values (e.g., Morrow, 1983). Construct validity is an evidentiary requirement for all scientific constructs. Any test of cause-effect relationship must establish the construct validity of both the presumptive cause and effect.

Second, *Internal Validity* is the extent to which a study properly demonstrates a causal relationship between a presumed cause and effect. Many causal factors play a role in a given bit of data (Bogen & Woodward, 1988; p. 317), including the phenomenon

itself, the instrumentation used and other factors impossible to fully account for (e.g., respondent mood, order effects of items or measures). Internal validity is actually an amalgam of features required of an informative body of evidence. *Covariation* means that indicators of cause and effect are inter-related. *Temporal Precedence* means that studies are designed such that the "cause" precedes the "effect" in time. *Non-Spuriousness* means that no plausible alternative explanations exist for their observed covariation.

Measurement quality is particularly important to reducing spuriousness. Poor measurement quality creates alternative explanations for observed covariation where measures are unreliable (i.e. contain substantial error, as when respondents have difficulty answering complicated questions) or invalid (i.e. lacking construct validity, as in the case of a general intelligence test that taps cultural knowledge but not mental abilities per se). Careful design of primary studies promotes these three conditions of internal validity, but seldom eliminates them. Threats to internal validity are overcome when accumulated studies with different designs yield comparable findings.

The third criterion, *Effect Size* is a measure of the strength of the relationship observed between two variables (Hedges & Okin, 1985). In research on causal relationships, the key indicator is an effect judged significant according to the decision rule established in advance (e.g., statistical significance). It is less apparent whether its *size* is important given the host of factors that can constrain it including the observed variance in variables (Fichman, 1999). Moreover, some effects can be so easily induced that their effect size is less important than the fact that they are likely to be relatively pervasive (e.g., in-group/out-group effects; Prentice & Miller, 1992). Other difficult-to-influence dependent variables such as successful organizational change can have

cumulative effects over time (e.g., where initial small effects escalate as the change develops; Goodman & Rousseau, 2004). However, effect sizes are the common currency of meta-analyses summarizing statistical findings across multiple studies (Hedges & Okin, 1985). The *significance* of the effect size is essential information in a meta-analysis. Relative size of the effect may be less critical depending upon the research's purpose. However, studies combining effect sizes with cost/benefit information, indicating for example whether great benefit can be attained at low cost or with little effort, can have important evidentiary value.

Fourth, *Generalizability* or external validity refers to the extent to which a result holds across populations, settings, procedures, and times. Some results are idiosyncratic to particular research contexts and don't occur outside them (e.g., answers to a particular survey item; Fischhoff, 1991; Sudman, Bradburn & Schwarz, 1998). Most experimental settings never occur naturally, providing instead special "pure" conditions (Festinger, 1953). A study has evidentiary value when it provides information (qualitative or quantitative) regarding the conditions to which a treatment effect or phenomenon is generalizable. Robust causal relationships are those that are stable across situations, such as trust's impact on cooperation (Fichman, 2003) and in-group/out-group effects on liking (Brewer, 1979). Less robust effects may depend on context, such as how leadership styles impact follower behaviors (e.g., Porter & McLaughlin, 2006). Generalizability is a matter of judgment based upon information a set of studies provide about participants, treatments, and settings (Steckler & McLeroy, 2008). Unfortunately, research reports often fail to include in sufficient detail the facts needed to assess generalizability (e.g.,

participant background, the study's time frame or circumstances; Rousseau & Fried, 2000)

Generalizability is particularly threatened when only findings from published studies are considered. A purpose of traditional literature reviews is to identify whether findings are stable across researchers, methods, measures and times (Salipante, et al., 1982) to provide a firm foundation for advancing knowledge (Webster & Watson, 2002). However, identifying the stability of findings requires that relevant unpublished as well as published studies be reviewed, to overcome the bias many journals have against publishing non-significant results. Statistical meta-analyses that make special effort to overcome the "file drawer problem" provide more generalizable results than reviews limited to published materials (Schmidt & Hunter, 1990). A review that effectively identifies the generalizability of findings "facilitates theory development, closes areas where a plethora of research exists, and uncovers areas where research is needed" (Webster & Watson, 2002, p. xiii).

Fifth, *Intervention Compliance* refers to the occurrence of all conditions required to induce a particular cause or apply a specific treatment. "Treatment" non-compliance by care givers and patients is a source of inconsistency in the findings of medical research. Attention to compliance raises questions regarding whether all practices specified in the protocol were followed. It questions how well-trained, skilled or competent were those responsible for implementation. Compliance is a particular concern in MOS because of variations in organizational and management practices. Pritchard, Harrell, DiazGranadeos & Guzman (in press) investigate between-study differences in how thoroughly an organizational analysis and assessment system, PROMES, was applied. Several

implementation-related factors were identified in primary studies, including the extent users adhered to the PROMES process and the quality of the feedback provided. These features of intervention compliance were found to impact the overall productivity gains associated with PROMES. Generally speaking there is widespread variation in how organizations implement routines (e.g., performance appraisal) or interventions (e.g., quality programs). A study providing information regarding differences in actual implementation, and the sensitivity of outcomes to it, has considerable evidentiary value.

Sixth, *Contextualization* is empirical evidence regarding how context influences the phenomenon under study. It goes beyond generalizability in that contextualization identifies the limits of a phenomenon or cause-effect relationship by providing information regarding *why* it is limited. One important form of evidence to identify contextual supports, that is, co-occurring conditions not part of the phenomenon itself, which influence its occurrence or consequences. Such is the case where the effects of high involvement work systems depend upon organizational supports such as workforce training and appropriate rewards (cf. MacDuffie, 1995). In contrast, absence of contextual support is indicated when an organizational setting is inhospitable to a new management practice or other intervention. Prior events or historical factors, such as a previously failed change, can lead an otherwise constructive management practice to fail. Investigations into conditions of failure as well as success can contextualize the occurrence and operation of a given practice or intervention.

In addition, context can impact generalizability by altering the meanings associated with the phenomena studied. Such factors are commonly noted with respect to location (e.g., industry or country) and time frame (e.g., pre-internet, cf. Rousseau &

Fried, 2000). In the case of location, national culture differences for example are known to influence how MOS phenomena are socially constructed. Such in the case where directive leadership produces more positive responses in countries valuing authority than in more egalitarian nations (e.g. House, et al., 2004). In the case of time frame, historical forces can influence MOS phenomena as witnessed in the organizational responses to market forces and the shifting dynamics between capital and labor (Barley & Kunda, 1988; Guillen, 1994). In interpreting studies of employee-employer relationships, time frame must be factored in as in the case where unionism (greater in the 1970s than the 90s) and globalization (the reverse) may influence the perspectives of actors and the roles institutions play (e.g., government, stock market participation).

Primary studies have some evidentiary value along the above criteria. However, only in synthesizing an accumulated body of studies can the full meaning and quality of evidence be gauged.

Challenges to Research Syntheses in MOS

Having identified the criteria for establishing what evidence is in a systematic review, we now turn to investigate the dynamics within MOS that make it difficult to actually synthesize research. Three factors must be addressed to move the area toward greater use of research syntheses:

1. Alternative models of science disputing what is knowable and how to determine it;

2. Divergent political and cultural assumptions regarding the appropriate focus and methods in studying workers and managers, organizations and markets, and the institutions in which they are embedded; and
3. Professional rivalry in gaining legitimacy, institutional support, and scarce resources.

Alternative Models of Science

Alternatives views of science exist within MOS and underlie disputes regarding the nature of evidence. Epistemology, the theory regarding the nature of knowledge, acknowledges several approaches to scientific knowledge (Morgan & Smircich, 1980). The form evidence takes in each approach depends on its assumptions regarding the nature of reality. MOS grew out of the social sciences and its epistemological approaches reflect these roots. Let's start with two poles of an epistemological continuum most relevant to management and organizational science, positivism and relativism. (Note we exclude a fourth epistemology sometimes associated with MOS, pragmatism, because it ignores the essential role theory plays in the value of evidence. Pragmatism's focus is "can something be made to work" not why it works; James, 1907). Rooted in logic, Positivism argues that evidence is what can be observed. Empirical study and independent verification are the proper basis for developing and evaluating natural explanations of observable phenomena. In positivism, evidence constitutes cause and effect relationships that can be repeatedly observed (e.g., rewards increase the frequency of targeted behavior, aspirin reduces inflammation). Positivism seeks explanations founded on the notion of a unified reality governed by observable laws. In social science, the positivist perspective is found in behaviorism (e.g. Skinnerian psychology).

Positivism's emphasis on universality leads its advocates to make assumptions that limit the research questions they ask. The question "under what conditions will individuals resent the use of reinforcement systems" is unlikely to come up since "resent" is an emotional state not a behavior and an individual's viewpoint falls outside positivism's chosen domain. Positivism downplays the role of context and history making it less compatible to the study of organizational practices and workplace experiences. It has limited applicability in MOS where research often focuses on subjectivity, including individual and collective interpretations of events (e.g., Martin, 1990), not observation per se.

At the opposite pole is relativism, an epistemology skeptical about the existence of a reality beyond what is socially constructed (see Bechara & Van de Ven, 2007). No universal reality is presumed. Observation constitutes a faulty basis for knowing. In rejecting the notion of reality, relativism manifests in a family of approaches including feminist criticism and postmodernism. Each focuses to a large extent on the many explanations or narratives that account for the human experience using text as a basic datum. Instead, relativist approaches analyze verbal behavior, speech and print, with a goal of broadening understanding of the actors' perspectives. In relativism, evidence constitutes any themes or modes of expression actors repeatedly convey. Relativist scholars interpret these repeated expressions in terms of the conscious and unconscious processes of social actors and the larger social settings. It excludes phenomena associated with organizations as open systems embedded in the natural world including regularities in environmental adaptation and other systematic functions (e.g., Miller, 1971). Limited

in the same fashion as positivism, relativism cannot address phenomena incompatible with its view of reality.

The middle ground between positivism and relativism is occupied by critical realism. More pluralistic in its approach to evidence, critical realism includes any epistemological position that maintains the existence of an objectively knowable reality, while acknowledging that perception and cognition mediate human comprehension of that reality (e.g., Bhaskar, 1997; 1998). In this context "critical" means critique of the mind, especially judgments of fact and value. Assumptions and alternative interpretations are probed, compared, and tested. The evidence critical realism focuses upon are the patterns observed in data and their support or refutation of the mechanisms theory specifies.

This chapter argues that syntheses in MOS research are best accomplished using a critical realist epistemology. Critical realism acknowledges that all methods have limits. Each conflates to some degree what is experienced with what exists. As such, rather than advocate one method over another (e.g., quantitative/qualitative, objective/subjective indicators, laboratory experiment/field study), critical realism makes such a choice unnecessary. Instead it emphasizes triangulation across methods and forms of data. This triangulation is especially valuable in creating systematic research syntheses that address the array of evidentiary criteria specified above. Adopting critical realism means that scientific knowledge is conditional, although based upon general and invariant processes. In the case of MOS, a meso or cross-level approach to theory building and research (e.g., top/down control mechanisms and bottom/up emergent processes) is compatible with critical realism (House et al., 1995).

Our approach to evidence synthesis is based upon an understanding that reality exists independent of human cognition (i.e., objective), and that all facts, observations and data are theory-laden (i.e., subjective), there being no value-free inquiry. Theoretical pluralism and triangulation are necessary because reality requires multiple perspectives to understand its regularities (Bhaskar, 1998). All evidentiary criteria apply, from internal validity to contextualization, consistent with the acknowledged value of multiple research methods. Research methods differ in their appropriateness depending upon the entities studied. Thus, surveys suitable for assessing individual predispositions may be less appropriate for characterizing cultural institutions.

Political and Cultural Implications of Evidence

A second challenge to evidence synthesis lies in the political and cultural assumptions conflated with research evidence. Douglas McGregor advised, "Tune your ear to listen for assumptions about human behavior, whether they relate to an individual, a particular group, or people in general" (McGregor, 2006; p. 9). Political and cultural perspectives enter into the dispute regarding what is evidence. First, any single study is likely to be "ethnocentric" to some degree, making implicit and unexamined assumptions regard the interests it represents. These assumptions could be societal in nature where Americans, French, and Chinese researchers might well be expected to view the world, and organizational phenomena, differently. For example, the ways TQM are implemented and the factors that give it local meaning are highly dependent on cultural processes (D'Irbarne, 2000), which in turn influences the nature of both commitment and compliance to this global management system. Even within the same society, scholars hold divergent viewpoints (e.g., labor/ management, free market/governmental support),

with consequences for their approach to research, questions asked and methods they apply (Ghoshal, 2005).

Implicit assumptions can lead to choices of research methods that reinforce certain beliefs, regardless of their validity. The commitment research conducted in the individualistic United States at one time largely focused upon individual-level studies of workers in a single firm (e.g., Mathieu & Zajac, 1990). Consistent with historically negative views of workers (e.g., McGregor's, 2006, Theory X; Green, 2006), it is little wonder that scholars once attributed commitment to the vagaries of worker motivation and their personal predispositions. Subsequent inquiry into commitment expanded to consider employer effects. Multi-unit/multi-firm studies began finding that organizations themselves influence workforce commitment (e.g., Shore & Tetrick, 1991). Narrative reviews and statistical meta-analyses conducted on the early commitment research would have been misleading by failing to question whether the organization itself played a role.

A similar example from public health displays the risks of the traditional medical definition of evidence and its biases. McGuire (2005) observes that epidemiological studies typically focus on the individual correlates of disease, measuring the incidence of illness across with particular demographics. Such studies tend to ignore upstream factors, such as racism, poverty, and unemployment in the case of hypertension. Because individual-level studies are more likely to involve the randomized control trials (RCT) traditionally employed as medical research's gold standard, they better conform to conventional notions of evidence. The result is the potentially harmful situation where contributing societal factors and variations in implementation quality often associated with these are ignored. Reliance on RCTs as evidence downplays the role that non-

individual factors might play in health and disease, as example of implicit assumptions having unintended effects on the way research findings are interpreted and used.

Frames of reference and untested assumptions create invisible biases in research. In a classic demonstration of invisible cultural effects in research, Latham, Erez and Locke (1988) describe their efforts to reconcile divergent results reported by goal setting researchers. The result they investigated was the finding that worker participation was important to promote goal acceptance in Israeli studies, but not in North American ones. By working side by side and comparing their research protocols, the researchers found that this effect was due to the ways in which experimenters in the two countries had given their subjects instructions. In the non-participation condition, the Israeli "tell" strategy, assigning goals by direct order, contrasted with the North American "sell" approach, assigning goals in a fashion that motivated subjects to accept them. As such, cultural practices and assumptions can seep into research methods in unexpected ways.

We know from the history of science that the uses to which research is put can skew how that research is interpreted in ways that are fundamental about status and power (Baratz, 1974). A case in point, the assessment center is a method for leader selection that uses experts and trained observers to assess individuals for their leadership potential. This methodology is widely understood to be egalitarian in nature, its assessments both unbiased and valid with respect to the assessment of men and women, minorities and majority group members (Huck, 1973; Zal, 1998). However, their original use reflected a political agenda. During World War II, the British military had a dearth of office candidates given the large loss of life and its traditional recruiting of officers from the upper classes. Buddy ratings, where enlisted men rated the leadership

capabilities of peers, were found to be an effective way to identify officer candidates (Hollander, 1954; Weitz, 1958). A military review board during the war found this method "too democratic" (Cutcher-Gershenfeld, 1982), and initiated assessment centers as a substitute for the otherwise valid information enlisted personnel provided. Thus the use or non-use of particular practices may have more to do with their cultural conformity than their validity. A systematic review evaluating assessment centers, now widely used, risks passing along this non-egalitarian bias, unless the historical context of its application is represented in the analysis.

Lastly, history plays a significant role in shaping the underlying cultural and political perspective operating on both phenomena and scientist. Joseph Scanlon, an innovator in workplace motivation, was purportedly fond of saying "face the past and back into the future." Consider the case of industrial accidents. Viewed as acts of God from the outset of the industrial revolution until the early 20th century, industrial accidents came to be attributed at least in part to the accident proneness of workers, "inherent psycho-physiological predisposition toward accidents" (Viteles, 1932, p. 334. Subsequent changes in technology and work practices led to creation organizational practices that reduced accident rates to a point where their occurrence, if rare, came to be seen as a consequence of systemic factors (cf. Perrow, 1984) or worker inexperience in a specific job (e.g., Gauchard et al., 2007). The conceptual categories used to identify and understand events such as accidents are socially and historically determined. A reflective review thus needs to consider the prevailing political, cultural, or historical trends relevant to the context of its primary studies.

The Professional and Scientific Infrastructure

A third challenge to conducting evidence syntheses is the fragmentation characteristic of management research (Whitley, 2000). In his treatise on the sciences, Whitley classifies MOS as a field with low task and functional dependence. Its researchers, he asserts, experience less need to systematically incorporate their colleagues' results and ideas into their research than is the case in other fields. In contrast, in psychology or economics competition for reputation and centrality in these more unified fields leads to more integration (Whitley, 2000; Pfeffer, 1993). Low integration is complicated by widespread disinterest in replication on the part of MOS journals, preferring novelty to accumulation (Kilduff, 2007).

Fragmentation and specialization are dysfunctional when they undermine a field's integration and knowledge accumulation. Diversity of approaches is not troublesome in itself. Integrative pluralism is a constructive consequence of critical realism, when different theories can be used to account for the same phenomenon in compatible and non-conflicting ways. We suspect that MOS researchers may feel the need to downplay the array of research relevant to their own research programs out of the mistaken belief that pluralism reflects the immaturity of the area as portrayed by Thomas Kuhn (Mitchell, 2002). Instead, we argue that MOS's pluralism is a result of the complexity of its subject matter. Pluralism in theory and method can arise from different levels of analysis and in the assumptions underlying phenomena at different levels (Mitchell, 2002). Individuals, workgroups and organizations are each distinct and irreducible. Groups and organizations cannot be reduced to people because they can pre-exist their current membership and have causal power in themselves (Bhaskar, 1998). Each has causal mechanisms at its own

level (immediate or proximal) while mechanisms at other levels operate more distally (e.g., see Hackman, 2003; House, Rousseau & Thomas- Hunt, 1995).

Not all divergent explanations are competing. Instead what we have in MOS's different approaches and methods is a scientific division of labor. Nonetheless, on-going efforts to integrate findings obtained in diverse ways increase the likelihood that convergence will occur (cf. Staw, Sandelands & Dutton, 1980). Systematic reviews help identify whether differences across research domains are substantive or semantic, signs of different starting points or disciplinary assumptions, or authentic differences in the phenomena studied.

Conclusion

Management and organizational research is essentially a human science (Foucault, 1965) or a science of the artificial (Simon, 1996). In contrast to the physical sciences, MOS's domain is knowledge about human-made objects and phenomena. That is, it focuses on the human-made world of organizations and the behavior of people in relation to them. Human-made organizations, work groups, markets, and social institutions are by their very nature complicated. These multi-level human constructions necessitate critical realism's multiple methods to generate the requisite systematic knowledge. This conclusion is in line with an observation by a proponent of critical realism, "It is the nature of the object that determines the form of its possible science" (Bhaskar,1998, p. 3).

For integration to occur, MOS research from different theoretical and methodological perspectives needs to be accessed, organized, and interpreted into a

synthesis of the evidence. Political and cultural assumptions operating on both researchers and the phenomena studied must be reflected upon in the review process. Adopting a critical realist perspective enables a balanced assessment of the full array of research relevant to MOS's scientific and practical questions.

Approaches to Systematic Research Syntheses

We next describe how systematic syntheses have been conducted across various fields. Syntheses constitute a family of methods for accumulating, organizing, and interpreting research (Dixon-Woods, Agarwall, Young, Jones & Sutton, 2004). As a quintessential human creation, science requires not only methods to represent phenomena that cannot be observed directly but also on-going *critiques* of methods and interpretation to achieve authenticity and value. Systematic synthesis is such a critique: it evaluates both data and the interpretations made of them. Existing methods fall into four categories: aggregation, integration, interpretation and explanation.

Synthesis by aggregation

The characteristic method of aggregative synthesis is statistical meta-analysis. Aggregative syntheses extract and combine findings from separate studies to increase the effective sample size (Schmidt & Hunter, 1990). They undertake to provide evidence of "what is true" or 'what works' by summarizing an overall net effect (e.g. Kluger & DeNisi, 1996). This effect may be a correlation or a treatment effect, including measures of the performance or effectiveness for a given class of interventions. The synthesis's starting point is a narrow, focused question, such as 'does X lead to outcome Y', with 'X' and 'Y' and the 'outcome' being defined as tightly as possible. Since the validity of a

meta-analysis is depends on the quality and homogeneity of the primary studies on which it is based, its synthesis process is structured with the aim of reducing bias. Prior to actual synthesis, the researcher is encouraged to set a priori criteria specifying the types of studies will be included. Next extensive searches of published and unpublished studies are conducted and their methodological quality assessed to distinguish between reliable and unreliable research. In several fields regularly employing aggregated syntheses, the process has been formalized with the goal of producing rigorous and replicable reviews (Greenhalgh, 1997). See for example, the Cochrane Collaboration (www.Cochranecollaboration.org) in medicine and the Campbell Collaboration (www.campbellcolloboration.org) in education and criminal justice.

Promoters of aggregative synthesis stress the importance of controlling for bias, which includes use of a hierarchy of evidence to rank research designs according to their internal and external validity. Within this synthesis tradition, randomized controlled trials are often regarded as the 'gold standard' for judging 'what works' (Evans, 2003). Acknowledging the problems of observational (non-manipulated) studies, critics stress the limited scope of RCT methods for studying important social and organizational issues (Pawson, 2002a,b), noting the progress made in statistics and social science in working with observational data (e.g., Nagin, 2005; Rubin, 2007). .

A key problem in aggregative synthesis is that primary studies are rarely homogeneous. Studies can be pooled despite widely ranging effects reflecting unmeasured complexity (e.g., differences in quality of implementation, measured outcomes, or contextual factors that influence observed effects). Unless primary studies carefully report implementation processes and context, aggregated results can mask the

mechanisms underlying effects (Pawson, 2002a). Meta-analysis advocates (e.g. Schmidt & Hunter, 1990; Shadish & Sweeney, 1991) suggest that such differences can be accounted for by identifying mediators and moderators. However, the aggregated synthesis's design must anticipate the presence of mediators and moderators, and the relevant information must be available from the primary studies (cf. Gough et al. 2003). Pooling effect sizes, the additive model of evidence aggregated syntheses rely upon, can fail to provide insights into the mechanisms and contextual factors relevant to organizations and related human behavior. A meta-analyses on the other hand can inform both theory and practice when undertaken in a manner that allows tests of causality (e.g., Orlitzky, Schmidt & Rynes, 2003) or makes use of supplemental information regarding context or nature of the intervention (e.g., Kluger & DeNisi, 1996, described below). The basic approach to meta-analysis may itself be adaptable to the addition of more diverse sorts of information, including the inclusion of supplementary information from relevant qualitative work and expert practice (Gregson, Rodger, Neal & Avis, 2002).

Synthesis by Integration

Integrative synthesis involves the collection and comparison of evidence containing two or more data collection methods to investigate patterns across primary research studies. It can compensate for weaknesses in any single research design to improve the internal and external validity of findings. Data from one study design can be triangulated to corroborate findings from others or to investigate different aspects of a problem. One important use of integrative syntheses is to capture procedural knowledge related to how interventions or treatments might best be carried out in a particular context. Recognition is growing recognition that including qualitative data in syntheses

can provide ‘situation-specific wisdom’ (Campbell, 1984). This particularly useful where treatments or practices involve some improvisation (e.g., care givers, teachers or managers adapting a practice to local conditions). Qualitative data can help capture local ways of doing things as well as the subjective human experience, often excluded from more arms’ length studies (Dingwall et al, 1998).

Integrative synthesis typically employs predetermined questions and selection criteria. This approach often asks multiple questions at the same time so that a review of diverse research studies addresses issues no single study can. First, typical questions framing an integrative synthesis relate to effectiveness and cause-effect relationships (e.g., Does the intervention work? What cause-effect relationships are supported across studies? What are the benefits and harm? Who will benefit from application of these findings?). Second, the review queries the appropriateness of the intervention or causal relationship with regard to affected parties (e.g., What is the experience of the consumer, manager, or worker? What issues are important to them? Do the parties view the outcomes as beneficial?). Third, questions address the feasibility of applying the findings and in particular any procedural knowledge that might aid its implementation (e.g., What resources are required for the effects to occur or intervention success? Will the intervention be accepted and used? How should it be implemented? What are its economic implications?).

An exemplar of an integrative synthesis is the systematic review of information systems outsourcing conducted by Dibbern, Goles, Hirschheim and Jayatilaka (2004). Putting the phenomenon in the context of industry (e.g., financial, e-Business) and history, Dibbern and colleagues pull together the literature’s diverse definitions of

outsourcing, and then offer a synthesis of them. They organize and review the literature from positivist, descriptive and interpretive epistemologies. Their final product is a synthesis of constructs and result relationships with respect to an array of research questions (e.g., why outsource?).

Integrative syntheses can address a broad range of questions and accommodate heterogeneous sources of data (Dixon-Woods et al., 2004). Judgment and interpretation are crucial to this synthesis and cannot be eliminated by proceduralization, making it "…a sense-making exercise and not a *mechanical* one" (Pawson, 2002a). This synthesis may be especially useful to MOS where its findings have practical implications. An integrative synthesis in MOS might identify both facts (declarative knowledge) and ways to use them (procedural knowledge) as in the case of goal setting research's identification of the effect of specific versus do-your-best goals and the procedures organizations might use to implement it (cf., Locke & Latham, 1984; Locke & Latham, 2002).

Synthesis by interpretation

Interpretive approaches to synthesis are associated with relativist epistemologies (e.g., phenomenology or social construction). They are typically limited to published findings, rather than primary data in part because qualitative data are not as readily shared unless in a published form. The central focus of qualitative studies is often different from that of quantitative ones (e.g., experiences of participants versus assessments of effects). As such the motives for integrative syntheses often differ from those of aggregated reviews, driven more by interests in human experience and social phenomena rather than interventions or mechanisms. Their purpose is often to translate key interpretations from one study to another.

One approach in the interpretive synthesis family is meta-ethnography (Noblit & Hare, 1988: 5-6). It involves a flexible or open coding system developed by coders reviewing the data to identify categories that emerge. Coders compare the imagery and themes that surface across studies (Beck, 2001). Advocates argue that this approach is a unique form of synthesis because it preserves the interpretive qualities of the original data by, "carefully peeling away the surface layers of studies to find their hearts and souls in a way that does least damage to them" (Sandelowski, et al., 1997: p. 370). In meta-ethnography (see for example, Campbell, et al., 2003), reviewers also identify higher-order concepts not evident in the primary studies. The product is the construction of larger narratives and more generalizable theory (Estabrook, Field & Morse, 1994; Sandelowski et al., 1997). Interpretive synthesis compiles descriptive data and exemplars from individual studies, building them into a mosaic or map (Hammersley, 2001). The process is one of conceptual innovation and reinterpretation (Campbell et al. (2003), while attempting to preserve the original study's integrity or wholeness (Pawson, 2006).

Another method of interpretive synthesis is described by Popay, Roberts, Sowden, Petticrew, Arai, Rodgers, and Britten (2005). It focuses on understanding treatments or practices and involves four steps: 1) developing a theoretical model of how the interventions work, why and for whom; 2) deriving a preliminary synthesis; 3) exploring relationships in the data; and 4) assessing the robustness of the synthesized output. Primarily, this narrative approach relies on words and text to capture and explain synthesis findings. Its goal is to 'tell the story" of studies included in the review (Popay, et al. 2005: 5). In healthcare, qualitative reviews have proven useful in providing more complete understanding of how patients initiate strategic changes in their own regimens,

such as diabetics who vary the timing of insulin administration in response to job demands, not as a form of deviance, but a form of more effective coping (Campbell et al, 2003).

Interpretive syntheses are useful where there is a relatively comparable body of qualitative data. In MOS, such might be the case for example if Hochschild's (1997) study of work/family time pressures where interpreted in relation to Perlow's (1997) study of flexibility in work hours. There is no reason why quantitative data cannot be incorporated in an interpretative synthesis as a means of identifying a common explanation across bodies of data, but few examples of this exist (Dixon-Woods, et al. 2004). Unlike aggregative synthesis which seeks to avoid or mitigate bias, the reviewer is central to the process of interpretative synthesis. As such the synthesis provides a feasible explanation of the studies findings rather than a replicable explanation (Noblit & Hare, 1988). Because qualitative research tends to emphasize the richness and depth of contextual analysis (Fielding & Fielding, 2000), critics of this approach (e.g., Estabrooks et al., 1994) argue that interpretive synthesis is 'context-stripping,' not an interpretive act. Another controversy within the community of qualitative researchers is use of structured methods such as purposive sampling, grounded theory, multiple coders, and respondent validation and triangulation as a means for demonstrating rigor (e.g., Wolcott, 1990; Lincoln & Guba, 2005; Dellinger & Leech, 2007). Such practices do not appear to be widely accepted.

Synthesis by explanation

The explanatory approach to synthesis focuses on identifying causal mechanisms and how they operate. It seeks to discover if they have been activated in a body of research

and under what conditions (Sayer, 1992: 14). This synthesis involves a critical realist approach:

> "Social systems are the product of literally endless components and forces. When social science tries to focus on what seems a uniform pattern of behavior it soon discovers that it is shaped by historical forces, with the result that it may occur in one culture but not the next. Secondly institutional forces play an inevitable part. They render behavioral patterns susceptible to change under different organizational arrangements and political structures. Thirdly, behavioral regularities are, of course, also influenced by the volition and choices of people who act them out" (Pawson, 2006: 18)

Explanatory synthesis starts by articulating likely (alternative) underlying mechanisms and then interrogates available evidence to find out whether and where these mechanisms are applicable. Primary studies are then used "to test, revise and refine the preliminary theory" (Pawson & Tilley, 1997: 74). Each relevant published article is described and discussed in terms of its contribution to the emerging theory,

> "…. the reviewer's basic task is to sift through the mixed fortunes of a programme attempting to discover those contexts (C+) that have produced solid and successful outcomes (O) from those contexts (C) that induce failure (O-). The review process is then repeated across a number of additional studies featuring the same underlying mechanism, with the aim of gathering together the various permutations of success and failure… the aim is to differentiate and accumulate evidence of positive and negative CMO configurations" (Pawson 2002b: 345).

In one synthesis Pawson (2006) addresses the question "does Megan's law work?" (a state law publicizing sex offenders living in the community). Available data did not permit determining whether naming (the Intervention) led to re-offending (Outcome). Instead, Pawson capitalized on the broad array of related data and focused on the broader use of public shaming. Shaming in this synthesis constituted the mechanism explaining what system characteristics make naming effective in reducing re-offending. Pawson then built a theory of naming and shaming by developing an understanding based on the accumulating reports from a wide range of programs (e.g., bad checking writers, deadbeat parents, etc.). Note some of these programs had positive results and others not. Through synthesis, Pawson was able to build a theory from the body of evidence regarding why naming and shaming worked in which situations and with which types of people.

Explanatory synthesis recognizes no hierarchy of evidence. The worth of the study can only be determined in synthesis, by each study's contribution to pattern-building (Pawson, 2006). However, unlike the traditional narrative review an explanatory approach is comprehensive in its accumulation of relevant evidence. Representing the array of approaches to tackling a specific research question is critical to this form of synthesis. It treats the vital evidence from primary studies to include the original researchers' interpretations and explanations, not just the results. The scope of the review includes a wide range of studies, research types and data forms to promote a full understanding of the phenomenon of interest. Its product is a revised model intended to explain for whom, in what circumstances, in what respect and why, certain causes or interventions produce preferred outcomes (Pawson, 2002b).

The explanatory approach is useful for synthesis in fragmented and methodologically diverse fields and where little consensus exists regarding what constitutes quality research. MOS is such a field. The role of theory is crucial to the explanatory approach and this theory is made explicit. Advocates of interpretive synthesis view the explanatory approach as merely a form of interpretive synthesis and the quality of the review is heavily reliant on the appraisal and interpretation skills of the analyst (Boaz, et al. 2006). In psychology, Zajonc and Markus's (1975) review of birth order effects on intelligence is an example of this approach. Statistical findings, qualitative information and primary researcher interpretations were synthesized to identify underlying mechanisms accounting for observed patterns.

Implications

No single consensus exists regarding what is 'best evidence,' as the above four approaches to systematic synthesis demonstrate. Nonetheless, the four approaches have certain similarities. They all have a goal of optimizing use of primary studies. Putting together the findings of individual studies can demonstrate insights not otherwise evident. All are question-driven. Each approach yields what can be considered complete studies in themselves by virtue of their formal methods. They differ from primary studies by the fact that they make use, at least in part, of published material. Lastly, though they display no consensus regarding what constitutes *quality* evidence, all use "tough but necessary tests for evidence"' (Solesbury 2004): How relevant is this to what we are seeking to understand or decide? How representative is this of the population that concerns us? How reliable, how well-founded theoretically, empirically is it? Features from these four approaches are informative in developing appropriate approaches to syntheses in MOS.

Features of MOS Research Syntheses

MOS can learn from the approaches to synthesis other fields use. One commonplace complaint MOS doctoral students voice is how difficult it can be to identify and make sense of a body of research. Typically, students receive little training in conducting literature reviews. Developing an approach to synthesis especially adapted to MOS can serve several purposes from better doctoral student education to accelerated scientific progress and more effective practice (Rousseau, 2007). Another motive for developing an MOS-oriented approach to synthesis is the essential role that systematic reviews play in other disciplines adopting evidence-based approaches to practice (Tranfield et al., 2003). However, critics have framed the use of synthesis as a competitive matter (e.g. Learmonth & Harding, 2006), giving rise to "'paradigm wars' and corresponding mudslinging on behalf of 'quantitative' versus 'qualitative', 'positivism' versus ' phenomenology', outcomes versus 'process' and so on" (Pawson, 2001: 3). MOS's inherent pluralism suggests the need to find an alternative to such self-defeating pursuits. We concur with Fischhoff's (1991, p. 844) observation that each discipline has an intact critique of its competitors, blinding proponents to other perspectives. Rather than privilege a narrow set of research methods, we propose a framework for systematic research syntheses congruent with MOS's pluralism in methods, phenomena and potential end users (scholars, textbook writers, practitioners, and policy makers). We recommend a four-stage synthesis process as a starting point for learning and further development.

Step 1: Question Formulation: Reflection, Debate, and Reformulation

To be useful a review must have a clear purpose and an audience (e.g., scholars, practitioners, policy makers). Beginning with the end in mind, the review question(s) must reflect the review's intended use. Let's say the intended use is scholarly explanation. Targeting explanation might lead to a question such as "what mechanisms operate in the influence of pay on motivation?" In contrast, if the intended use is a practical application, the review process might begin with a question such as "how can the effect of pay on motivation be optimized?"

Synthesizing MOS research is complicated by the pluralism of values held by their stakeholders and the complexity of organizations (cascading effects across levels). The process of question formulation requires attention to both. Attending to stakeholders at the question stage can introduce a variety of meanings for key constructs. Returning to matter of pay and individual motivation, adopting an employer's perspective might concentrate attention on motivation in the form of willingness to engage in in-role performance. In contrast, an employee's point of view might lead to considering whether motivation reflects commitment (internalized motivation with broad behavioral effect) or compliance (externalized controls narrowing the range of likely behavior. In that case both in-role and extra-role performance may be used as proxies for motivation. Any construct a review question contains must be defined in the context of that question to provide an organizing framework for the review. Hence relevant perspectives and underlying value judgments are important to surface early on.

Scientists too make value judgments. These value judgments reflect prevailing scientific beliefs as well as the scientist's professional education, personal motives, and response to institutional pressures (cf. Baratz, 1974; Campbell, Daft & Hulin, 1982).

The complexity of the phenomena associated with the review question(s) derives from MOS's multi-level nature. Studies from several levels of analysis may be pertinent to a review question. Take for instance the question of whether incentive pay impacts individual performance. As described above, Rynes and colleagues (2005) identify several ways in which incentives can impact worker performance. An individual-level motivation effect can occur along with position- or firm-wide sorting effects where highly motivated workers are attracted and retained and the rest screened out. The studies required to address the impact of incentive pay on individual performance involve phenomena and data from several levels. The wording of the question itself is insufficient to determine what levels of measurement the studies reviewed should reflect. Reviewers need to develop knowledge with respect to the theories relevant to the issues at hand before determining the range of studies to examine.

Given the many purposes of evidence, from explanation to forecasting, any question framing needs to factor in the time frame in which evidence has been gathered (i.e. contextualization, as described above). Moreover if the purpose of the review is to advise management practice, it needs to account for the context of future use. As noted above, the historical nature of evidence risks drawing conclusions biased by existing or past practices and prevailing conditions (Clegg, 2005). So how then can we overcome the limits of the historical past when the future is invisible not only to past researchers but ourselves?

On-going changes can alter the nature and meaning of a causal relationship or the effectiveness of an intervention. As the nature of work, the competencies of workers, and the sophistication of management and information systems change, the factors driving

outcomes can change too. In those syntheses where effects are the focus, a basic question needs to be asked: Compared to what? In a review focused on effective management practice, questions might be included targeting sectors and subgroups of particular relevance. Let's say that a relevant fact to management practice is that in the course of a decade, workers in China and India are transitioning from low-wage jobs to global knowledge work. The reviewer knows that better educated workers are likely to be managed differently than their less educated counterparts. It is also the case that sophisticated global firms have different incentive systems and supports than many domestic ones. The questions asked might seek a more fine-grained look at aspects of the research most relevant to the anticipated nature of future work and firms. Nonetheless, as the mix of workers and firms change, fluctuating ranges in variables (broadening or narrowing) can alter observed effects and their stability. It is important to remember that science describes our world as it has been and is, not a logically necessary one (Mitchell, 2000, p. 251), nor the world that will be.

The framing of the question also may acknowledge other limits. Past research has focused largely on European and North American samples, with research from Asia or other parts of the world appearing only recently. Because the contexts of past observations may be limited, the initial set up for the review can acknowledge such limitations in order to make a concerted effort to offset them (e.g., seeking relevant non-English language primary studies), to provide a basis in subsequent steps for comparing North American findings with those obtained elsewhere, or to establish boundary conditions.

Although we are never free of the effects of personal values, history and culture, we can consciously critique our own question framing to surface as possible assumptions and implicit biases in order to consider these factors in the review. In formulating review questions, seeking out the opinions and feedback from others with diverse points of view helps yield more informative questions. We note that having questions reviewed by teams of scholars and practitioners is a common practice in fields with longer histories of conducting systematic reviews (e.g. Cochrane and Campbell Collaborations). At this early stage in the synthesis process, having an array of scholars and practitioners assess and interpret the question from various points of view can enhance the ultimate value of the completed review. Doing so can increase the mindfulness of reviewers regarding the disputes and alternative frames of reference associated with the review's subject. In all, in formulating a review question (or questions) we have two goals: a well-specified and informative question, asked in a manner that avoids obscuring contested issues.

Stage Two: Comprehensive Identification of Relevant Research

The next step is comprehensive identification of research relevant to the central review questions. Salipante and colleagues (1982) have pointed out that the reviewer is inevitably limited by the studies available. In assembling the material to be reviewed, in most cases, the broadest possible array of relevant research should be gathered to compensate for researcher value judgments and uncontrolled validity threats. For most uses, multiple forms of data from descriptive to archival, quantitative to qualitative, are likely to be relevant. Heterogeneity also makes possible the investigation of contextual factors influencing the study's design and its findings. Potential sources include

published and unpublished material, data bases, and direct contact with the authors of the research. As MOS scholars work in a myriad of countries, it can be necessary to access work in several languages, advantaging those review teams whose members have relevant language fluency.

The more heterogeneous the distribution of uncontrolled validity threats in a set of similar findings, the greater the validity of the findings from the set (Salipante and colleagues (1982). In the pursuit of appropriate inclusion of relevant research, precise inclusion/exclusion criteria should be specified, applied, recorded, and monitored (Tranfield, et al. 2003).

In the case of systematic reviews involving organizational practices, our current understanding of how managers and employees actually engage in these practices (deciding when to offer or seek out a reward, promotion, etc.) isn't well developed. Moreover, both organizational practice and policy execution entail some degree of improvisation (cf. Feldman & Pentland, 2003). Systematic syntheses addressing practice-related questions require primary studies addressing implementation and its consequences as well as descriptions of how the manager and the worker "responds in a unique way to the unique case" (Dingwall, Murphy, Watson, Greatbatch & Parker, 1998, p. 168, referencing the clinician).

Step 3: Organizing and Interpreting

After gathering the appropriate array of relevant studies, the planning for data extraction begins. The first key issue is what information to derive and code from primary sources. The purposes reflected in the framing of the review questions guide the information extracted and interpreted. Is explanation provided via a single theory,

competing theories, or complimentary ones? Studies might be coded to identify *mechanisms* associated with a phenomenon if competing theories are investigated. Or, to identify *phases* of a phenomenon if complimentary theories are examined.

Use of multiple extractors, as readers and coders, is important at this step in order to reduce mistakes in data recording as well as to avoid omission of relevant material. We anticipate that this recommendation will be criticized as "rigor displayed for its own sake." Note, however, in adopting a critical realist perspective we acknowledged the fallibility of all observations. This limitation applies to the information compiled in systematic reviews. A rigorous, transparent process review process is therefore wholly in keeping with MOS's epistemological basis.

The fallible nature of judgment and rating processes is compounded in systematic reviews by the non-standard reporting styles even published sources use. Data need to be available for checking. Agreement between extractors should be assessed in terms of percentage agreement and/or inter-rater reliability as appropriate. Divergent judgments should be reconciled. In the case of qualitative data, disagreement between extractors should be recorded in order to capture relevant aspects of their interpretations.

Next, the review entails the systematic organization of the data into formats that allow summary and display in the synthesis report. This body of evidence is then probed, sifted, coded, aggregated, and cross-tabulated in numerous ways. All the while, reviewers should engage in a mindful questioning of a priori beliefs regarding the scope and implications of relevant research.

It is important to examine whether the type of method or perspectives researchers have adopted play a role in their findings. Scientific theories explain facts about

phenomena rather than facts about data. If an alleged phenomenon is only detectable via highly specialized body of data, this raises suspicion that the phenomenon is spurious (Bogen & Woodward, 1988, p. 318). Spuriousness turned out to be the case in empirical studies of Herzberg's two factor theory. The hygiene/satisfier distinction showed up in interviews formatted so that interviewees credit themselves for things they like (using their skills at work) and blame the company for what they don't (pay)—but not in surveys or interviews structured differently (House & Wigdor, 1967; King, 1970). Facts and phenomena are not found in the data themselves, but in their interpretation. Interpretation is informed by more than data per se. It is informed by the sensemaking and reflection scholars engage in.

Research background, discipline, or social ties can predispose scientists to analyze and interpret data in a way that confirms firmly held or taken-for-granted beliefs. Such appears to have been the case with Least Preferred Coworker research, popular in the 1960s and 70s (Fiedler, 1967; 1983) where only the findings of the theory's creator and associates, consistently supported the theory. This pattern can be attributed to the intensity of analytic methods used, including the addition of control variables until observed results were fell in line with theory (Graen, Alvares, Orris & Martella, 1970). To identify such patterns, it is necessary to look for clusters and outliers in the results.

The product at this stage is a summary (descriptive and/or statistical), identifying regularities, inconsistencies, and co-occuring conditions that potential influence findings. Information summaries provide include effect size (r or d), patterns of co-occuring features or conditions, and the phenomenon's antecedents and consequences. A case in point is Kluger and DeNisi's (1996) statistical meta-analysis (607 effect sizes

representing 23,663 observations) looking back at several decades of research on performance feedback's effect on task or job performance. The effect proved to be conditioned upon the nature of the feedback given. If the form of feedback given called more attention to the self than to task, performance declined. If feedback were more task-focused, subsequent performance improved. Comparison of effect sizes was used to identify the underlying phenomena.

Descriptive information also included in summaries may capture the functions or meanings of a given practice (e.g., promotions that typically are seniority based in one set of studies or merit-based relative to one's cohort in another). They may also locate the phenomenon studied in its social, industry or cultural context (as in the case where a meta-analysis of assessment center outcomes for men and women is interpreted with respect to research on the way actual promotion decisions are made). As always, however, the reviewer, and consumers of reviews, must recognize that another level of threats exist to entire sets of studies—the conditions limiting the actual settings included and the types of studies conducted (Salipante, et al.,1982, p. 345).

Step Four: Synthesis

The resulting synthesis is an informed explication of what the scientific evidence says with respect to the question and related issues surfaced in the process. Triangulation identifies convergent findings (consistent patterns of effect sizes, comparable meanings and/or goals attached to particular practices or interventions). It also involves making make sense of inconsistent results, instead of ignoring them. Differences can be attributable to error, to method, or in some cases to meaningful factors such as boundary conditions and situational constraints on the distributions of observed variables (e.g.,

effects of education on occupational performance may be difficult to detect from studies within a given type of work since education is often a pre-condition of occupational entry). Integrative explanations focus on explaining interdependencies between aspects of a problem, its boundaries and context (van de Ven, 2007). Exploiting multiple perspectives makes salient robust features of reality by distinguishing the patterns across them. Some argue that judgment itself can be systematized using decision making technologies or Bayesian methods to integrate weighted judgments (e.g., Sutton, 2001). Informed conversation across stakeholders and interest groups regarding the meaning of syntheses' findings and future implications seems critical for MOS evidence to be well-interpreted and understood. Opening the synthesis to public discussion not only provides a check on quality but can deepen its value and authenticity.

Interpretation need not stop when synthesis is achieved. Some syntheses will review that the evidence is inconsistent or weak, prompting critical research to resolve contested issues. Interpretation of the synthesis can lead to a simulation of results to replicate synthesized conditions. It can be a catalyst for redesign of practices and routines. It can lead to the development of policy implications, informed by how stakeholder interests can be differentially affected by its findings. Perhaps most important, reviews can form the basis for consensus conferences of scholars, practitioners, and policy makers to engage in well-informed conversations about the evidence now in hand.

Implications

> "A design representation suitable to a world in which the scarce factor is information may be exactly the wrong one for a world in which the scarce factor is attention."
> (Simon, 1996; p. 144)

This chapter is intended to encourage the good scientific practice of evidence synthesis. All studies have limits. Only in their combination does evidence emerge. The goal of science is to produce useful models of reality. Nonetheless, a good deal of human history saw little impact of science on human life. Research has not always had a net benefit. In MOS for example, some scholars have critiqued its research for espousing dysfunctional views of labor and capital (Ghoshal, 2005). We believe that the opportunity to both uncover and develop constructive solutions to the problems workers and firms face can be realized through a careful, reflective consider of MOS evidence. But assuredly little impact will occur until we embrace the process of research synthesis and evidence convergence. Indeed doing otherwise only increases the likelihood that researchers continuing to confront seemingly inconsistent results, will choose one of two paths. In the absence of systematic reviews, they may throw up their hands and abandon promising lines of research due to an inability to distinguish complexity from randomness. Or, they may (unconsciously) escalate their efforts to produce the results they prefer. Collectively we lose either way.

Syntheses are needed to provide access to the evidence that would inform MOS's development as well as the teaching and practice of evidence-based management. In this regard the *Annals* series and the Academy of Management generally have an important role to play in promoting systematic syntheses and their public discussion. Note, however, we have no expectation that the availability of evidence directly translates into action. Even rational models of decision making focus on what knowledge decision makers already possess. Few people search for relevant evidence before making

decisions. Unfortunately failure to seek relevant evidence affects the decisions of scholars too. Herb Simon used to say "When an academic starts a sentence, 'As a (fill in the blank),' I always know I am not going to learn anything." You may fill in the blank with the words "economist," "psychologist," "sociologist," or any "ist" of your choosing. Long ago, Dearborn and Simon (1958) demonstrated that a subject presented with a complex stimulus sees what he or she is ready to see. As human as laboratory subjects and executives, scientists themselves tend to see what their frames of reference prepare them to see (Walsh, 1988; Waller, Huber & Glick, 1995). So, if the information provided by an evidence synthesis is not enough to promote it use, then what?

Interpreting evidence is essentially a community project. Intensive communication and networks are needed to impact the understanding, uptake and application of evidence. Evidence syntheses and their products belong in the day-to-day conversations of MOS's constituencies, its scholars, practitioners, and policy makers. We echo Bhaskar's (1998) aspiration that research synthesis will become an emancipatory social practice: How much more effectively might we use the resources at our disposal to foster change if we interpret the world adequately?

References

Arthur, W. Day, E.A., McNelly, T.L., & Edens, P.S. 2003. A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology* 56 (1), 125–153.

Baratz, L. 1974. *The servants of power; a history of the use of social science in American industry*. Westport, Ct: Greenwood.

Barley, S.R. & Kunda. G. 1992. Design and devotion: Surges of rational and normative ideologies of control in managerial discourse. *Administrative Science Quarterly*, 37: 1-30.

Bechara, J.P. & VandeVen, A.H. 2007. Philosophy of science underlying engaged scholarship. In A.H. VandeVen, *Engaged Scholarship: A guide for organizational and social research.* (pp. 36-70) New York: Oxford.

Beck, C.T. 2001. Caring with nursing education: A metasynthesis, *Journal of Nursing Education*, 40 (3), 101-110.

Bhaskar, R. 1997. *A realist theory of science*. 2nd edition. London: Verso.

Bhaskar, R. 1998. *The possibility of naturalism: A philosophical critique of the contemporary human sciences* (3rd edition), New York and London: Routelege.

Bogen, J. & Woodward, J. (1988) Saving the phenomena. *The Philosophical Review*, 97, 303-352.

Boaz A., Ashby D., Denyer D., Egan M., Harden A., Jones D.R., Pawson R. & Tranfield D. 2006. A multitude of syntheses: a comparison of five approaches from diverse policy fields. *Evidence and Policy* 2: 479-502.

Brewer, M. 1979. In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. Psychological Bulletin, 84: 307-324.

Briner, R.B. 1998. *What is an evidence-based approach to practice and why do we need one in occupational psychology?* BPS Occupational Psychology Conference.

Campbell Collaboration 2005. http://www.campbellcollaboration.org/. Accessed December 5.

Campbell, J.P., Daft, R.L., Hulin, C.L. 1982. *What to study: Generating and developing research questions*. Newbury Park, Calif.: Sage.

Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. & Donovan, J. 2003. Evaluating meta-ethnography: A synthesis of qualitative research on lay experiences on diabetes and diabetes care. *Social Science and Medicine*, 56, 671-684.

Cartwright, N. 1983. *How the laws of physics lie*. Oxford: Oxford University Press.

Clegg, S. 2005. Evidence-based practice in educational research: a critical realist critique of systematic review. *British Journal of Sociology of Education*, 26, 415-428.

Cochrane Collaboration 2005. http://www.cochrane.org/index0.htm. Accessed December 5.

Cutcher-Gershenfeld, J. 1982. Oral history interview with Eric Trist, Toronto, Ontario. Personal communication with Denise Rousseau, Chicago, September 19, 2007.

Davenport, T.H. 2006. Competing on analytics. *Harvard Business Review*, January,

Dean, C. 2006. When question of science come to a court room, truth has many faces. *New York Times*, December 5.

Dearborn, D.C. & Simon, H.A. 1958. Selective perception: A note on the departmental

identifications of executives. *Sociometry*, 21, 140-144.

Dellinger, A.B., Leech, N.L. 2007. Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, 1, 309-332.

Dibbern, J., Goles, T., Hirschheim, R. & Jayatilaka, B. 2004. Information systems outsourcing: A survey and analysis of the literature. *The DATA BASE for Advances in Information Systems*, 35 (4): 6-102.

Dingwall, R., Murphy, E., Watson, P., Greatbatch, D. & Parker, S. 1998. Catching goldfish: Quality in qualitative research. *Journal of Health Services Research Policy*, 3, 167-172.

D'Iribarne, P. 2002. Motivating workers in emerging countries: Universal tools and local adaptations. *Journal of Organizational Behavior*, 23, 1-14.

Dixon-Woods, M., Agarwal1, S., Young, B., Jones, D. & Sutton, A. 2004. *Integrative approaches to qualitative and quantitative evidence.* Department of Health Sciences, University of Leicester2. Department of Psychology, University of Hull. Available at www.hda.nhs.uk.

Estabrooks, C.A., Field, P.A. & Morse, J.M. 1994. Aggregating qualitative findings: An approach to theory development. *Qualitative Health Research*, 4, 503-511.

Evans D. 2003. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing* 12(1): 77-84.

Feldman, M.S. & Pentland, B.T. 2003. Reconceptualizing organizational routines as a source of flexibility and change *Administrative Science Quarterly*, 48, 94-118.

Festinger, L. 1953. Laboratory experiments. In L. Festinger and D. Katz (eds.),

Research methods in the behavioral sciences. New York: The Dryden Press.

Fichman, M. 1999. Variance explained: Why size does not (always) matter. *Research on Organizational Behavior*,21, 295-331.

Fichman, M. 2003. Straining towards trust: some constraints on studying trust in organizations. *Journal of Organizational Behavior*, 24, 133-157.

Fiedler, F. E. 1967. *A theory of leadership effectiveness.* New York: McGraw-Hill.

Fiedler, F.E. 1983 Assessing the validity of Fiedler's Contingency Model of Leadership Effectiveness: A closer look at Strube and Garcia, *Psychological Bulletin* 93: 404–8.

Fielding, N.G. & Fielding, J.L. 2000. Resistance and adaptation to criminal identity: using secondary analysis to evaluate classic studies of crime and deviance. *Sociology*, 34: 671-89.

Fischhoff, B. 1991. Value elicitation: Is there anything in there? *American Psychologist*, 46, 835-847.

Foucault, M. 1965. *Madness and civilization: A history of insanity in the age of reason* (R. Howard, trans.). New York: Pantheon.

Gauchard, G.C., Mur, J.M., Touron, C., Benamghar, L., Dehaene, D., Perrin, P. & Chau, N. 2007 Determinants of accident proneness: a case–control study in railway workers. *Occupational Medicine*, 56, 187-190.

Ghoshal, S. 2005. Bad management theories are destroying good management practices. *Academy of Management Learning and Education*, 2005,4,75-91.

Glick, W.H., Miller, C.C. & Cardinal, L.B. 2007. Making a life in the field of

organization science. *Journal of Organizational Behavior*, 28, 849-857.

Goodman, P.S. & Rousseau, D.M. 2004 Organizational change that produces results: The linkage approach. *Academy of Management Executive*, 18 (October), 7-21.

Goodstein, L. D. & K. L. Brazis. 1970. Credibility of psychologists: An empirical study. *Psychological Reports*, *27*, 835-838.

Gough, D.A., Kiwan, D., Sutcliffe, S., Simpson, D. & Houghton, N. 2003. A systematic map and synthesis review of the effectiveness of personal development planning for improved student learning. http://eppi.iow.ac.uk/EPPIWeb Content/reel/review_groups/EPPI/LTSN/LTSN_June03.pdf

Graen, G., Alvares, K.M., Orris, J.B. & Martella, J.A. 1970. Contingency model of leadership effectiveness: Antecedent and evidential results. *Psychological Bulletin*, 74, 285-296.

Green, J. 2006. *Death in the Haymarket: A story of Chicago, the first labor movement, and the bombing that divided gilded age America.* New York: Pantheon.

Greenhalgh, T. 1997. Papers that summarise other papers (systematic reviews and meta-analyses), *British Medical Journal*, 315, 672-675.

Gregson, P., Rodger, W., Meal, Andrew G., & Avis, M. 2002. Meta-analysis: the glass eye of evidence-based practice? *Nursing Inquiry*, 9, 24-30.

Guillen, M.F. 1994. *Models of management: Work, authority, and organization in a comparative perspective*. Chicago: University of Chicago Press.

Guzzo, R. A., Jackson, S. E. & Katzell, R.A. 1987. Meta-analysis analysis. *Research in*

Organizational Behavior Greenwich, CT. JAI Press. Vol 9, 407-442

Hackman, J.R. 2003. Learning more from crossing levels: Evidence from airplanes, orchestras, and hospitals. *Journal of Organizational Behavior*, 24, 1-18**.**

Hammersley, M. 1997. Qualitative data archiving: some reflections on its prospects and problems. *Sociology* 31: 131-42.

Hammersley, M. 2001. On 'Systematic' Reviews of Research Literatures: a 'Narrative' Response to Evans and Benefield, *British Educational Research Journal*, 27, 543-554.

Hedges, L. V., & Olkin, I. 1985. *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hochschild, A. R. 1997. *Time bind: When work becomes home and home becomes work.* New York: Henry Holt.

Hollander, E.P. 1954. Buddy ratings: Military use and industrial implications. *Personnel Psychology*, 7, 385-393.

Hornung, S, Rousseau, D.M. & Glaser, J. 2008. Creating flexible work arrangements using idiosyncratic deals. *Journal of Applied Psychology*, in press.

House, R.J., Rousseau, D. M. & Thomas-Hunt, M. 1995 MESO: An integration of macro and micro OB. In L. L. Cummings and B. M. Staw (Eds.), *Research in Organizational Behavior*, *17*, 71-114. Greenwich, CT: JAI Press.

House, R.J. & Wigdor, L.A. (1967) Herzberg's dual factor theory of job satisfaction and motivation: A review of the evidence and a criticism. *Personnel Psychology*, 20, 368-389.

House, R.J., Gupta,V., Dorfmann, P., Javidan, M. & Hanges, P. 2004. *Culture, leadership, and organizations: The GLOBE study of 62 societies.* Thousand Oaks, CA: Sage.

Huck, J.R. 1973. Assessment centers: A review of the internal and external validities. *Personnel Psychology*, 26, 191-212.

James, W. 1907. *Pragmatism*. New York: The American Library.

Kilduff, M. 2007. Top ten reasons why your paper might not be sent out for review. *Academy of Management Journal*, 32, 700-702.

King, N. (1970) Clarification and evaluation of the two-factor theory of job satisfaction. *Psychological Bulletin*, 74, 18-31.

Kluger, A. & DeNisi, A. 1996. The effect of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119: 254-284.

Kovner, A.R., Elton, J.J. & Billings, J.D. 2005. Evidence-based management. *Frontiers of health services management*, 16(4): 3-24.

Latham, G. 2007. *Work motivation: History,, theory, research, and practice*. Thousand Oaks, CA: Sage.

Latham, G.P., Erez, M., & Locke, E.A. (1988). Resolving scientific disputes by the joint

design of crucial experiments by the antagonists: Application to the Latham-Erez dispute regarding participation in goal setting. *Journal of Applied Psychology*, 73, 753-772.

Lawler, E.E. 1971. *Pay and organizational effectiveness. A psychological view*. New York: McGraw-Hill.

Learmonth, M. & Harding, N. 2006. Evidence-based Management: the very idea. *Public Administration*, 84, 245-266.

Lincoln, Y.S. & Guba, E.G. 2005. Paradigmatic controversies, contradictions, and emerging confluences. In N.K. Denzin & Y.S. Lincoln (eds), *The Sage handbook of qualitative research.* (3rd edition, pp. 191-215).

Locke, E.A. & Latham, G.P. 1984. *Goal setting: A motivational technique that works.* Englewood Cliffs, NJ: Prentice-Hall.

Locke, E.A. & Latham, G.P. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *Annual Review of Psychology*,57: 705-717.

Lodahl, J. B. & G. Gordon. 1972. The structure of scientific fields and the functioning of university graduate departments. *American Sociological Review, 37*, 57-72.

Lomas, J., Culyer, T., McCutcheon, C., McAuley, L. & Law, S. 2005. *Conceptualizing evidence for health system guidance*. Final Report to Canadian Health Services Research Foundation, Ottawa, Ontario, Canada.

Martin, J. 1990. Deconstructing organizational taboos: The suppression of gender conflict in organizations. *Organization Science*, 1, 339-359.

MacDuffie, J.P. 1995. Human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *Industrial and Labor Relations Review,* 48, 197-221.

Martinson, B.C., Anderson, M.S. & deVries, R. 2005. Scientists behaving badly. *Nature*, 435: 737-738.

Mathieu, J.E. & Zajac, D.M. 1990. A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin*, 108, 171-194.

McGregor, D. 2006. *The human side of enterprise*. Annotated Edition. New York: McGrawHill.

McGuire, W.L. 2005. Beyond EBM: New directions for evidence-based public health. *Perspectives in biology and medicine*, 48, 557-569.

Merton, R.K. (1987) *Social theory and social structure*. New York: Free Press.

Miller, J.G. 1978. *Living systems*. New York: McGraw-Hill.

Mitchell, S.D. 2000. Dimensions of scientific law. *Philosophy of Science*, 67: 242-265.

Mitchell, S.D. 2002. Integrative pluralism. *Biology and Philosophy*, 17, 55-70.

Morgan, G. & Smircich, L. 1980. The case for qualitative research. *Academy of Management Review,* 5, 491-500.

Morrow, P.C. 1983. Concept redundancy in organizational research: The case of work commitment. *Academy of Management Review*, 8, 486-500.

Nagin, D. 2005. *Group-based modeling of development.* Boston: Harvard University Press.

Noblit, G.W. & Hare, R.D. 1988. *Meta-ethnography: Synthesizing qualitative studies.*

Newbury Park, CA: Sage.

Nutley, S. Water, I. & Davies, H.T.O. 2003. From knowing to doing. *Evaluation*, 9, 125-148.

Oakley, A. 2001. *Experiments in knowing: gender and methods in the social sciences.* Cambridge: Polity Press.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. 2003. Corporate social and financial performance: A meta-analysis. *Organization Studies*, 24(3), 403-441.

Pawson, R. 2002a. Evidence and policy and naming and shaming. *Policy Studies*, 23/3-4: 211-230.

Pawson, R. 2002b. Evidence-based policy: In search of a method. *Evaluation*, 8, 157-181.

Pawson, R. 2006 *Evidence-based Policy: a realist perspective*. Sage Publications.

Pawson, R., & N. Tilley 1997 *Realistic evaluation*. London: Sage.

Perlow, L. A. 1997. *Finding time: How corporations, individuals, and families can benefit from new work practices.*

Perrow, C. 1984 *Normal accidents: Living with high risk technology*.

Pfeffer, J. 1993. Barriers to the advance of organization science: Paradigm development as a dependent variable. *Academy of Management Review*, *18*, 599-620.

Pfeffer, J. 1998 Six dangerous myths about pay. *Harvard Business Review*, May-June, 108-119.

Pfeffer, J., & R.I. Sutton 2006. Evidence-based management. *Harvard Business Review* January: 62

Popay, J., H. Roberts, A. Sowden, M. Petticrew, L Arai, M. Rodgers, & N. Britten 2006. *Guidance on the conduct of narrative synthesis in systematic reviews, A product from the ESRC Methods Programme,* http://www.conted.ox.ac.uk/cpd/healthsciences/courses/short_courses/qsr/NSguidanceV1-JNoyes.pdf.

Porter, L.W. & McLaughlin, G. B. 2006. Leadership and the organizational context: Like the weather? *Leadership Quarterly*, 17: 559-576.

Prentice, D.A. & Miller, D.T. 1992. When small effects are impressive. *Psychological Bulletin,* 112, 160-164.

Pritchard, R.D., Harrell, M.M., DiazGranados, D. & Guzman, M. (in press) The Productivity Measurement and Enhancement System: A meta-analysis. *Journal of Applied Psychology*.

Rousseau, D.M., 1997. Organizational Behavior in the new organizational era. *Annual Review of Psychology,* 48, 515-546.

Rousseau, D.M. 2007. Outstanding in the fields of organizational science. *Journal of Organizational Behavior,* 28, 849-857.

.

Rousseau, D.M. & Fried, Y. 2000. Location, location, location: Contextualizing Organizational Behavior. *Journal of Organizational Behavior*, *22*, 1-15.

Rubin, D.M. 2007. The design versus the analysis of observation studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, 26, 20-36.

Rynes, S.L., Gerhart, B. & Parks, L. 2005. Personnel Psychology: Performance

evaluation and pay for performance. *Annual Review of Psychology*, 56: 571-600.

Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., & Richardson, W.S. 1996. Evidence based medicine: What it is and what it isn't. *British Medical Journal* 312:71-2.

Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W., & Haynes, R.B. 2000. *Evidence-based medicine: How to practice and teach EBM*. New York: Churchill Livingstone.

Sandelowski, M., Docherty, S. & Emden, C. 1997. Qualitative metasynthesis: Issues and techniques , *Research in Nursing and Health*, Vol. 20, No. 4, pp. 365-371.

Sayer, A. 1992. *Method in social science: realist approach*. London: Rutledge.

Salipante, P, Notz, B. & Bigelow, J. 1982. A matrix approach to literature reviews. *Research on Organizational Behavior*, B.M. Staw and L.L.Cummings (eds.) pp. 321-348. JAI Press.

Schmidt, F.L. 1992. What do data really mean? Research findings, meta-analysis and cumulative knowledge," *American Psychologist*,47, 1173-1181.

Schmidt, F.L. & Hunter, J. 1990. *Methods of meta-analysis: Correcting error and bias in research finding*, Thousand Oaks, CA: Sage Publications.

Shadish, W.R. & Sweeney, R.B. 1991 Mediators and moderators in meta-analysis: There is a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Counseling and Clinical Psychology*, 59, 883-893.

Sherman, L.W. 2002. Evidence-based policing: Social organization of information for

Social control. In E. Waring and D. Weisburd (eds.), *Crime and social organization,* New Brunswick, NJ: Transaction Publishers.

Shore, L.M. & Tetrick, L.E. 1991. A construct validity study of the survey of perceived organizational support. *Journal of Applied Psychology*, 76, 637-642.

Seijts G.H., & Latham G.P. 2005. Learning versus performance goals: When should each be used? *Academy of Management Executive*, 19: 124-131.

Simon, H.A. 1996. *The sciences of the artificial*. 3rd edition. Cambridge, MA.: MIT Press.

Smith, G.C.S. & Pell, J. 2003. Parachute use to prevent death and major trauma related to gravitational challenges: Systematic review of randomised controlled tasks. *British Medical Journal*, 327: 1459-1461.

Smith, R. 1991. Where is the wisdom? The poverty of medical evidence. *British Medical Journal*, 303, 789-799.

Solesbury, 2004. Evidence based policy: whence it came and where it's going. ESRC UK Centre for Evidence Based Policy and Practice, Department of Politics, Queen Mary, University of London, Mile End Road, London E1 4NS, Oct 2001. 10pp (Working Paper 1). http://evidencenetwork.org/cgi-win/enet.exe/biblioview?404

Staw, B.M. Sandelands, L.E. & Dutton, J.E. 1981. Threat-rigidity effects on organizational behavior. *Administrative Science Quarterly*, 26, 501-524.

Steckler, A. & McLeroy,K.R. 2008. The importance of external validity. *American Journal of Public Health*, 98, 9-10.

Sudman, S., Bradburn, N.M. & Schartz, N. 1998. Answering a survey question:

Cognitive and communicative processes. *Thinking about Answers*, pp. 55-79.

Sutton, A. 2001. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research,*10, 277-303.

Taubes, G. 2007. *Good calories, bad calories*. New York, Knopf.

Tierney, J. 2007. Diet and fat: A severe case of mistake consensus. *New York Times*, October 9.

Tranfield, D., Denyer, D. & Smart, P. 2003 Toward a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Managment*, 14, 207-222.

Tsang, E.W.F. & Kwan, K.-M. 1999. Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24: 759-780.

Van de Ven, A.H. 2007. *Engaged scholarship: A guide for organizational and social research.* New York: Oxford University Press.

Van de Ven, A.H. & Schomaker, M.S. 2002. Commentary: The rhetoric of evidence-based medicine. *Health Care Management Review*, 27(3), 88-90.

Vecchio, R.P. 1977. An empirical examination of the validity of Fiedler's model of leadership effectiveness**.** *Organizational Behavior and Human Performance,* 19, 180-206.

Viteles, M.S. 1932. *Industrial psychology*. New York: W.W. Norton.

Waller, M.J., Huber, G.P. & Glick, W.H. 1995. Functional background as a determinant of executives' selective perception, *Academy of Management Journal*, 38: 943-974.

Walsh, J. P. 1988. Selectivity and selective perception: An investigation of managers' belief structures and information processing. *Academy of Management Journal*, 31: 873-896.

Walshe, K. & Rundall, T.G. 2001. Evidence-based management: From theory to practice in health care. *The Milbank Quarterly,* 79: 429-457.

Webster, J. & Watson, R.T. 2002. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26 (2), xiii-xxiii.

Weitz, J. 1958. Selecting supervisors with peer ratings. *Personnel Psychology*, 11: 25-35.

Westen, D. & Bradley, R. 2005. Empirically supported complexity: Rethinking evidence-based practice in psychotherapy. *Current Directions in Psychological Science*, 2005, 14, 266-271.

Whitley, R. 2000. *The intellectual and social organization of the sciences*. 2nd edition. Oxford: Oxford University Press.

Wolcott, H. 1990. On seeking—and rejecting—validity in qualitative research. In E. Eisner & A. Peshkin (eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121-152.) New York: Teachers College Press.

Zal, J.N. 1998. Assessment centre methods. In PJD Drenth, H. Thierry & C.J. deWolff (eds.) *Handbook of Work and Organizational Psychology. Volume 3: Personnel Psychology (2nd edition,* (pp. 89-122). Psychology Press.

Zajonc, R.B. & Markus, G.B. 1975. Birth order and intellectual development. *Psychological Review*, 82, 1, 74-88.

Zammuto, R. F., & Connolly, T. 1984. Coping with disciplinary fragmentation.

Organizational Behavior Teaching Review, 9, 30-37.

Table 1

Alternative Epistemologies in Management and Organizational Research

	Positivism	**Critical Realism**	**Relativism**
Reality	Objective	Objective yet human interpretations effect observed reality	Socially constructed
Application of Evidence	Confirmatory. Only what is observable exists	Falsification	Critical
Data	Concrete and quantitative	Observations, judgments, and interpretations; quantitative and qualitative	Text--Spoken or Written
Focus	Observation as reality	Causal mechanisms identified via fallible observations	The sense people make of the social world

Table 2

Forms of Research Syntheses

	Aggregation	Integration	Interpretation	Explanation
Goal	Combine effect sizes to increase sample size and reduce bias. Answer specific questions. Predict intervention results by generating a more exact estimate than achievable by a single study.	Synthesize research from different methodologies. Answer specific questions. To explore when interventions are more likely to be appropriate.	Synthesize and interpret research to build higher-order theoretical constructs. Create tentative theories of phenomena including patterns of social construction.	Synthesize for explanation-building. Generate theory to explain how the world operates and under what circumstances.
Method	Quantitative combination of results of primary studies Calculation of the mean effect for a group of intervention Statistical meta-analysis.	Triangulation across multiple studies and methods. Different methods can investigate various aspects of problem. Mixed method synthesis; Bayesian meta-analysis.	Compilation of descriptive data and exemplars. Concepts across studies are identified and translated into new categories Narrative review; meta-ethnography.	Theories refined in the light of confirmatory and contradictory findings. Discern patterns behind multiple explanatory claims. Realist synthesis; explanatory synthesis.
Data	Evidence hierarchy. Favors randomized controlled trials where possible. Published and unpublished studies and data sets	Evidence hierarchy, somewhat contested. Typically published studies.	No evidence hierarchy. Typically published articles with qualitative data on highly comparable subject matter. Includes primary researcher interpretations.	Multiple forms of evidence accepted. Typically from published studies. Qualitative and quantitative data, including primary research's theory and interpretations.
Strengths	Minimal method bias.	Allows multi-	Synthesizes	Identifies boundary

	Answers precise question. Systematic, replicable process. Reputable method within the natural sciences.	method analysis. Combines statistical and qualitative synthesis. Highlights promising interventions to develop and test.	multiple qualitative studies. Allows reviewer to be reflexive and critical. Takes into account the nature of the context (its circumstances and the subjects involved)	conditions and contextual factors. Focuses on why and where interventions lead to certain outcomes. Takes into account the nature of the context (circumstances and the subjects involved) Data can be gathered from fragmented or methodologically diverse fields. Highly pragmatic; focused on informing decisions
Weakness	Useful only for studies with homogeneous statistical methods. Social systems often too complex for controlled intervention. Only quantitative data can be combined using statistical meta-analysis. Only narrow range of studies suitable for inclusion.	No standardization of methodology for review. Contested definitions of quality. Replication difficult. Skills in both statistical meta-analysis and qualitative analysis required.	Combining studies using diverse methodologies is contested. Meta-ethnographic approach under development. Yields many possible explanations. Replication difficult. Information from quantitative data can be lost. Coding relies on reviewer skills; method requires reflexivity.	Methodology is under development. Difficulties ensuring transparency and reproducibility Requires detail about the context and implementation of interventions which might not be available in primary studies. Highly dependent on skills of reviewer

Table 3

Synthesis Method

	SYNTHESIS METHOD			
EVIDENCE	**Aggregation**	**Integration**	**Interpretation**	**Explanation**
Construct Validity	?	✓	✓	✓
Internal Validity				
Covariation	✓	✓	✓	✓
Temporal Precedence	✓	?	?	✓
Non Spuriousness	✓	O	O	✓
Effect Size	✓	O	O	?
Cost/Benefit	?	O	O	?
Generalizability	✓	✓	✓	✓
Intervention Compliance	?	?	?	?
Contextualization	O	✓	✓	✓

✓= Yes ? = Unclear or inconsistent O = No