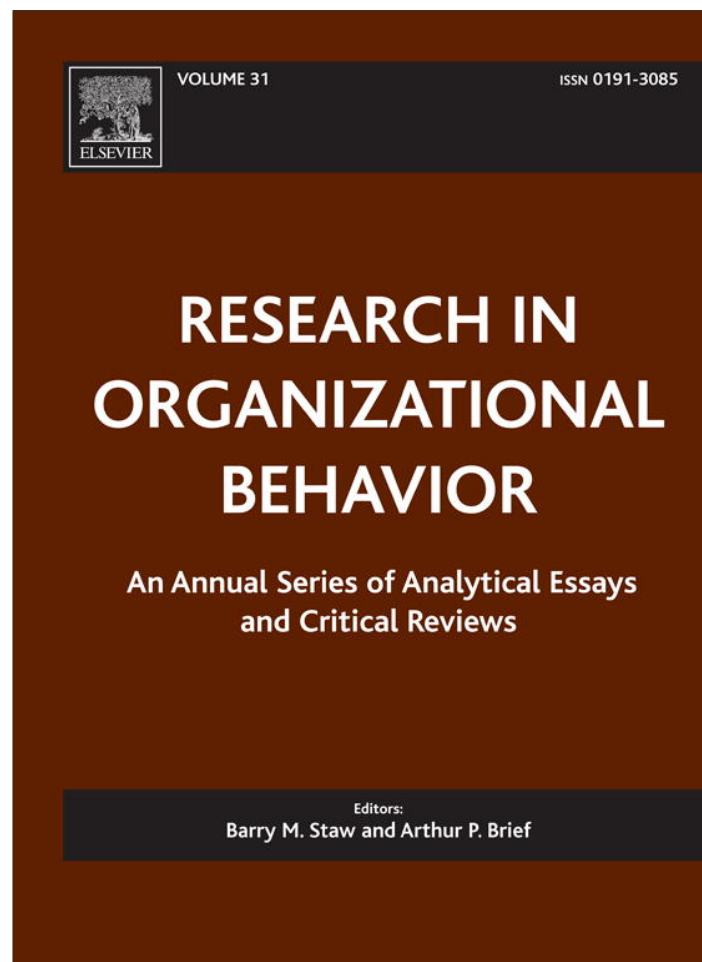




This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Best practices: How to evaluate psychological science for use by organizations

Susan T. Fiske^{a,*}, Eugene Borgida^b

^a *Department of Psychology, Princeton University, Princeton, NJ 08540, United States*

^b *Department of Psychology, University of Minnesota, N387 Elliott Hall, 75 East River Road, Minneapolis, MN 55455, United States*

Available online 8 November 2011

Abstract

We discuss how organizations can evaluate psychological science for its potential usefulness to their own purposes. Common sense is often the default but inadequate alternative, and bench-marking supplies only collective hunches instead of validated principles. External validity is an empirical process of identifying moderator variables, not a simple yes–no judgment about whether lab results replicate in the field. Hence, convincing criteria must specify what constitutes high-quality empirical evidence for organizational use. First, we illustrate some theories and science that have potential use. Then we describe generally accepted criteria for scientific quality and consensus, starting with peer review for quality, and scientific agreement in forms ranging from surveys of experts to meta-analyses to National Research Council consensus reports. Linkages of basic science to organizations entail communicating expert scientific consensus, motivating managerial interest, and translating broad principles to specific contexts. We close with parting advice to both sides of the researcher–practitioner divide.

© 2011 Elsevier Ltd. All rights reserved.

Contents

1. Common sense as the alternative	255
1.1. Generalizability: external validity	255
2. What theories and science get used?	257
2.1. What theories?	257
2.2. What psychological science?	257
2.3. What uses?	258
2.4. An in-depth example: the motherhood penalty	259
3. Criteria for quality and consensus	261
3.1. Quality science = peer-reviewed science	261
3.2. Establishing consensus	263
3.2.1. Surveys of experts	263
3.2.2. Meta-analysis	264
3.2.3. Narrative literature reviews	265
3.2.4. Adversarial collaboration	266
3.2.5. Professional society consensus reports	267

* Corresponding author. Tel.: +1 609 258 0655; fax: +1 609 258 1113.

E-mail addresses: sfiske@princeton.edu (S.T. Fiske), borgi001@umn.edu (E. Borgida).

4.	Linkage to real organizational life	268
4.1.	Communicating science to practitioners and persuading them to use it	268
4.2.	Organizational gatekeepers may not be interested.	268
4.3.	Generalizing from broad principles to specific situations.	269
5.	Parting advice	270
	References	271

In caricature, an earnest lab-coated scientist with an electronic tablet accessing hundreds of studies sits across the table from a designer-suited manager sneaking peaks at a smart phone accessing scores of corporate communications. The behavioral scientist fantasizes that mere scientific expertise suffices to make the case that the latest, exciting, reliable basic research should inspire the organization's practices. The manager prefers to anticipate the competition's latest plans. What are the proper origins for organizational innovation? Neither the one in the white coat nor the one in the power suit is making the best use of the other's goals and knowledge. This essay aims to help them talk to each other more effectively and to begin to build bridges that span these two different and complex perspectives.

Evidence-based management is gaining ground, but perhaps not fast enough; here we explore why. Using data to discern best practices, in legal or management decisions, is not easy: potentially relevant evidence is too abundant, hard to evaluate, possibly confounded, maybe not applicable, perhaps biased by conflict of interest, often self-selected, and harder to assimilate than a good story (Pfeffer & Sutton, 2006). Of all these pitfalls, this article suggests ways to manage the first two issues, overwhelming quantity and ambiguous standards, which together constitute one take on the criterion problem: how to understand the consensus of experts (to cut through the sheer quantity) and how to assess quality science (to focus on the best). What we mean by "criterion" is not an objective standard, because none exists, but rather how to judge credible expert consensus about high-quality science. To foreshadow, we advocate for reliance on peer-review to judge quality and on independently vetted summaries of the current science to assess consensus.

We write as basic behavioral scientists with some experience of organizations, besides working in them, and as consultants in the fraught domain of employment discrimination.¹ We have most deeply considered the use of psychological science in the courtroom (Borgida & Fiske, 2008; Fiske & Borgida, 2008, 2011), but here we seek to address a set of issues with broader applicability. We are hardly the first to grapple with these issues. Research-to-practice tensions are perennial in organizational psychology. For example, these debates have arisen in whether to design business schools as focused on *emics* (immersed relevance) or *etics* (scientific rigor) (Walsh, Tushman, Kimberly, Starbuck, & Ashford, 2007). Especially vociferous debates have surfaced recently in *Research in Organizational Behavior* (regarding the Implicit Association Test (IAT) specifically: Jost et al., 2009; Tetlock & Mitchell, 2009). In many of these debates, a pair of parallel concerns are the disrespect of practice (hence the inattention of the manager in our opening example) and its mirror-image, the neglect of scientific and especially laboratory-based evidence in management (hence the earnest demeanor of the scientist); the research and practitioner communities have long mistrusted each other (Vinchur & Koppes, 2011).

We do not pretend to resolve these tensions here. Of course, the research-practice synergy ultimately benefits both (Zedeck, 2011), and collaboration between them has long been recognized as ideal (Argyris & Schön, 1996); for example, the original practical theorist (Kurt Lewin) noted that people buy into new practices best when they help develop them (Lewin & Grabbe, 1945). Here we focus on one specific facet of basic research being used by organizations, namely, the criterion problem, how to decide *when* evidence should trump common sense. We presume that, if available, sound scientific evidence should trump common sense. But we also recognize that political and financial considerations and even professional biases (e.g., the skepticism about social and behavioral science that resides in some quarters) often play a role in situations where scientific evidence challenges everyday understandings.

¹ For Fiske: *Hopkins v. Price Waterhouse*, 618 F. Supp. 1109 (D. D.C. 1985); appeal: *Price Waterhouse v. Hopkins*, 825 F.2d 458 (D.C. Cir. 1987); Supreme Court review: *Price Waterhouse v. Hopkins*, 109 S. Ct. 1775 (1989); remand: *Hopkins v. Price Waterhouse*, No. 84-3040, slip op. (D. D.C. May 14, 1990). A subsequent case that also did not settle before trial: *Robinson v. Jacksonville Shipyards, Inc.* (M.D. Fla. 1989; Case No. 86-927). For Borgida, e.g., *Jenson v. Eveleth Taconite Co.* 824 F. Supp. 847 (D. Minn. 1993), later documented in the movie *North Country*. Other testimony involved consulting, and cases that settled.

1. Common sense as the alternative

Absent evidence, common sense is the default: it is convenient, believable, inexpensive, and most of all, comfortable. But organizations should not use mere familiarity as a criterion. Adopting organizational innovations on the basis of intuition and common sense can be costly in dollar terms, especially if and when subsequent evidence suggests that such innovations are ineffective and a waste of money (see Klenowski, Bell, & Dodson, 2010). Common sense is neither as commonly held nor as sensible as people think. The false consensus phenomenon alone shows that people's individual hunches do not overlap with other people's as much as they believe (Ross, Greene, & House, 1977). And common sense is famous for being contradictory, as any opening lecture in introductory social psychology shows by quoting conflicting aphorisms (opposites attract; but birds of a feather). Nevertheless, people persevere with familiar beliefs, resisting their overthrow (e.g., Bastardi, Uhlmann, & Ross, 2011, on biased evaluation of scientific evidence; Fiske & Taylor, 2008, on schemas).

Another caution against common sense is casuistry, namely the phony logic that anyone can muster to support desired outcomes (Norton, Vandello, & Darley, 2004). People choose and evaluate more favorably evidence that supports their preferred outcomes (Bastardi et al., 2011; Kunda, 1987), and people interpret as biased a third party who represents the objective middle ground between their view and an opposing view (Vallone, Ross, & Lepper, 1985). Worse still, people believe they are immune to these biases (Pronin, Gilovich, & Ross, 2004).

As an example of the weight of common sense: a content analysis of 2000–2008 federal race and sex discrimination cases (King et al., 2011) shows reports of both subtle and overt forms of hostile behavior. But only the most overtly intentional negative workplace behaviors that fit common sense ideas about discrimination actually increased the likelihood of pro-plaintiff legal outcomes. More subtle insults and denigrating comments about women and ethnic minorities had to occur more frequently and pervasively to meet common sense standards for race or sex discrimination. Yet science shows that under-represented group members routinely pick up and are undermined by the subtle micro-aggressions (Sue et al., 2007). Judicial behavior may not be behaviorally realistic in light of contemporary science: the law may not be “aligned with scientific consensus regarding the subtle nature of discrimination” (King et al., 2011, p. 56). So, too, with management, at least sometimes.

We raise common sense as the alternative to empirical evidence, not to disrespect practice, but to paint a worst-case scenario for the consequences of neglecting high-quality scientific consensus. Management-by-instinct, we suspect, is a frequent decision rule in many quarters, as attested by Jack Welch's *Straight from the Gut* (Welch and Byrne (2001)). However, hunches and intuitions are unlikely to use the most up-to-date scientific insights. Human decisionmakers, managers included, are ruled by individual heuristics and biases, as just noted.

Collective hunches are not necessarily more valid than individual hunches. Organizations may use benchmarks, adopting a practice from a successful organization just because that overall organization works well. Popular management techniques do increase the CEO's compensation, but they do not predict corporate performance (Staw & Epstein, 2000). Absent direct evidence, the role-model organization's success could be due to entirely other factors. Hence, even collective daily experience does not prove validity, which is an empirical question.

1.1. Generalizability: external validity

Common sense often prevails in arguments specifically about applications of basic science to organizations. External validity is the key issue here. Case in point: Social psychology experiments are in principle applicable to organizational settings (college sophomores are young adults, too), but the criticism often arises that laboratory studies must exaggerate effects because the rarefied environment can isolate the phenomenon of interest in a silent, hygienic, frictionless, and weird vacuum (Henrich, Heine, & Norenzayan, 2010a, 2010b; Sears, 1986). Surely, the noisy, dirty, windy, real world will demolish the delicate effects obtainable in rarified laboratory settings. But this is not necessarily so (Locke, 1986). For example, real-world ingroup/outgroup effects are often stronger than laboratory effects, when meta-analyses compare the two: people favor their ingroup even more in the real world than in the artificial minimal groups of the laboratory (Mullen, Brown, & Smith, 1992); people judge the outgroup as more homogeneous than the ingroup, even more in the real world than the lab (Ostrom & Sedikides, 1992); people preferentially remember stereotypic information about an outgroup, more so outside the lab than inside (Stangor & McMillan, 1992); people prefer familiar others more in the messy real world than in lab studies (Bornstein, 1989). Contrary to the physics laboratory, people are not particles: their behavior is often over-determined by several

correlated forces acting together in real-world settings. Externally valid settings may show stronger effects precisely because of being confounded.

Of course, field settings may sometimes show weaker effects than laboratory settings, as meta-analyses often examine (see upcoming section on establishing consensus with meta-analyses). Relative strength of field effects is an empirical question appropriate to what might be considered the third stage of understanding a phenomenon. First, basic research demonstrates an effect, and then the second stage documents processes (mediators). Then, the third stage explores boundary conditions (moderators) that produce stronger and weaker effects. The issue is not so much lab versus field, as a variety of identifiable conditions that interact with the phenomenon of interest (Locke, 1986). Older models of external validity assumed that effects simply would or would not generalize because of differences in participants, measures, and contexts, but more modern approaches treat generalizability as a conceptual and empirical stage of research. Indeed, the stages represent a cycle that can re-commence with field conditions that inform basic research. For example, Fiske's theory and research program on power-as-control contributing to stereotyping (1993) started with the realization that then-extant research spoke more to peers judging peers than to managers judging subordinates, one feature of the setting in her testimony for *Price Waterhouse v. Hopkins* (1989). But even without the power dynamics, basic processes of social cognition proved useful in understanding the organizational dynamics at work in the decision to deny partnership. The field setting informs the lab settings and vice versa. In fact, Cialdini (1980) long ago advocated for a similar approach to external validity concerns and called it “full-cycle social psychology.” In this approach (Cialdini, 1980, p. 44, emphasis added):

initial natural observation gives direction to subsequent controlled experimentation, the outcomes of which can then be given external validation through further natural observation that may stimulate still further experimentation. Systematic recourse to the evidence of the real world *both before and after performance of the experimental work* may thereby reduce the extent to which current social psychological research can be criticized as artificial and epiphenomenal.

Overall, organizations seeking evidence about implementing new policies and innovations or the efficacy of extant policies and innovations should give no priority to the idea that lab effects will *not* generalize. Indeed, what organizations should consider (assuming the quality and consensus criteria elaborated next) is that the principles, not the exact operationalizations will generalize (Banaji & Crowder, 1989; cf. Brief & Dukerich, 1991). That is, intuitions to the contrary, ecologically valid field methods do not guarantee generalizability of findings; instead ecologically valid methods are orthogonal to the generalizability of the results. What works in one narrowly defined but real setting might not work in another, but from such real-world research, one could not know the mechanisms to account for any differences in results. What matters is the systematic control over spurious relationships.

So, for example, both laboratory experiments and correlational studies (lab and field) have linked group cohesion and performance. Controlled experiments show that cohesion can improve performance, so now that principle is ready for export. What this particular field literature adds, besides testing generalizability, is longitudinal data showing that performance can also make groups more cohesive (Mullen & Copper, 1994). Both results get us beyond common sense, which might equally argue that cohesive groups become isolated and self-satisfied, undermining performance. (Common sense can argue any result and its opposite.) Although group cohesion has costs and benefits, the meta-analyses show benefits, and more important here, feedback between performance and cohesion. This illustrates the back-and-forth between lab and field as both generating and testing hypotheses. Lived experience or participant observation often generates theory, which is then fine-tuned in the lab and brought back into the field full-cycle (Cialdini, 1980). But as Staw (2010) suggests, this full-cycle approach to research is difficult to achieve, and caution must be exercised in applying experimental findings directly to workplace settings. Cases in point include the external validity concerns that constrain the integration of laboratory-based judgment and decision-making research (the JDM tradition) into organizational studies: “. . . we really don't know whether many of the most accepted biases and heuristics will stand up to the array of forces facing actors in natural settings such as work organizations” (Staw, 2010, p. 413). For a successful, full-cycle approach to these external validity issues, Staw suggests that JDM researchers need to be “more ambidextrous” in their methodological tool kit and choice of research settings, and they need to collaborate more often with organizational researchers whose own work is grounded in organizational studies.

2. What theories and science get used?

Before evaluating the criteria for using behavioral science evidence, it is important to specify what theories, what science, and what uses are involved in such an evaluation. Some illustrative issues provide the flavor of the problem, before the main event.

2.1. *What theories?*

Usefulness is a common criterion for judging applicable organizational theories (e.g., Miner, 1984). Usefulness refers to how much the theory guides effective action toward a goal. Not just any theory can apply in practice (Argyris, 1974). Useful theories must be coherent, and in two ways. First, internal consistency avoids the relevant variables being incompatible in practice, in that maximizing one would prevent the other. For example, increasing accountability may decrease privacy and individual sense of control, which may or may not be priorities. Or self-government may increase buy-in but create destabilizing tension, certainly not productive unless the cost is worthwhile. Second, coherent theories as used must also be congruent with the original theory as espoused. A leader may believe in democratic leadership but behave autocratically without self awareness sufficient to mind the gap.

Once coherent, the proof of a theory-in-practice depends on demonstrating its effectiveness. The relevant outcomes need to be facilitated, of course, but also other valued outcomes need to remain at acceptable levels. To judge such effectiveness, the theory-in-practice must be testable: specifying the antecedents, mechanisms, and consequences allows (dis)confirmation. Testing may create a cultural clash: As Argyris (1974) notes, in effective practice, one must be committed, confident, and decisive, but in testing a theory, one must take a tentative, skeptical, and experimental stance. Otherwise, the self-sealing theories fall prey to expectancy effects, impervious to testing. Argyris prescribes affirmation without dogmatism.

Even if theories-in-use fit all these criteria—being internally consistent, congruent, testable, and effective—the final criterion is whether the world they create is valuable, unintended consequences and all. This is an empirical question of a more grounded sort than basic science, but empirical nonetheless. The debate over the intended and unintended consequences of goal-setting, mentioned later, illustrates the comprehensive process of assessing organizational value on a variety of expected and potentially unanticipated dimensions.

All too often, theories are not seen as the pragmatic basis on which organizational decisions and policies should be made. That is, they may seem abstract, idealized, and irrelevant, while the managers must cope, on the ground, wrestling with the alligators in the mud. Theories may seem like the luxurious purview of academics with tenured job security and no practical accountability for their theories' effectiveness in the world.

On the other hand, perhaps theories need not be evaluated for their organizational usefulness. Theory can be useful in a broader sense than whether it immediately informs organizational practice (Brief & Dukerich, 1991). If usefulness prescribes directly what an organization should do, a more general practical theory sensitizes one broadly to how an organization might see itself in new ways. Does a useful theory need to prescribe action, or can it supply concepts for framing the issues? Requiring a theory to prescribe a specific action in a specific situation may be a mistake both because such a theory would not identify basic principles and because its prescriptions would be too specific to generalize. Further, as philosophy of science notes, an aggregation of specific prescriptions does not accumulate to scientific progress unless they constitute part of a larger testable theory that will ultimately make way for new theory.

Moreover, the requirement for immediately useful prescriptions may narrow theoretical vision, constrain creativity, neglect mechanisms, and distract from basic science (Brief & Dukerich, 1991). An overly specific set of recommendations is not a theory precisely because it is too narrow to explain anything beyond the particular application. In contrast, a generally practical (as opposed to immediately useful) theory can shed new light on general organizational issues, having utility in a broader sense. A practical theory can focus on principles—by elucidating general truths that reflect other truths—which in turn inform management, as well as teaching and research (Locke, 2009). Most important, practical but abstract principles can inform evidence-based management.

2.2. *What psychological science?*

Assuming that general principles from basic behavioral and social sciences can be practical for organizations, what kinds of science likely prove most relevant? Sampling recent handbooks in organizational science yields plenty of

topical overlap with psychological science: decision-making, leadership, power, personality, learning, creativity, challenge, stress, performance, goals, incentives, justice, values, culture, roles, stigma, stereotyping, teams, and negotiation (e.g., Anderson, Ones, Sinangil, & Viswesvaran, 2001; Locke, 2009; Zedeck, 2011).

As a specific example, the basic science of intergroup contact has settled on some principles that together predict successful conditions for diversity that works: authority support, equal status in the setting, common goals, and cooperation (Pettigrew & Tropp, 2006). What is more, as we now know, these principles operate not through increased knowledge but through the affective conditions that reduce anxiety, increase empathy, and allow friendships (Pettigrew & Tropp, 2008; Tropp & Mallett, 2011; Tropp & Pettigrew, 2005). These well-established principles offer practical insights for organizations building diverse workplace teams (see Plaut, 2010, on building a diversity science).

As another example, laboratory research famously shows abundant flaws in ordinary decision-making (e.g., Fiske & Taylor, 2008; Gilovich & Griffin, 2010; Kahneman, 2007). Ever since the first foibles in human inference were documented (Nisbett & Ross, 1980), optimists have opined that the lab tests are unnatural, unfair, unrealistic, or unfamiliar, and besides, errors are either self-correcting or socially corrected, the argument goes. So in this view, the research may under-estimate people's everyday decision-making (Staw, 2010). Nevertheless, as just one example, the track record in organizations prior to the Great Recession hardly inspires faith that somehow moving outside the lab makes decision-makers hyper rational or accurate. The economic model of the rational decision-maker is not faring well even in its home discipline (Gilovich & Griffin, 2010). The vulnerabilities in human judgment offer both cautions as noted and cures, in the form of using psychological science to make more mindful decisions (Weber & Johnson, 2009).

As a completely different example, potentially relating to liability issues, the media and entertainment industry minimize any link between video violence and actual aggression (Anderson & Gentile, 2008; Bushman & Anderson, 2001). Yet the basic science repeatedly shows that violent video games increase aggressive behavior, thoughts, and emotions, as well as decreasing empathy and prosocial behavior (Anderson et al., 2010). Further, the important conceptual variables in this literature generalize between lab and field (Anderson & Bushman, 1997). The links to organizations are several: Given product liability, producers of violent media might conceivably end up responsible for heightened aggression, especially in children. (Granted, the U.S. Supreme Court in June 2011 struck down a California law that prohibited the sale or rental of violent video games to minors, declaring that these video games are protected speech under the First Amendment; *Brown v. Entertainment Merchants Ass'n*, 2011.) Or perhaps organizations that permit on-site gaming (presumably during workday breaks) might inadvertently prime antisocial behavior. As a parallel, the presence of sexually objectifying media in the workplace primes sexual harassment of women, according to both basic science and judicial precedent (*Robinson v. Jacksonville Shipyards, Inc.*; Rudman & Borgida, 1995). Thus, principles of aggressive or sexual priming and spillover from media to behavior can inform the evidence-based use of media in the workplace.

As a final example, behavioral science can be practical in telling organizations what not to do. After a National Research Council report (Committee to Review the Scientific Evidence on the Polygraph, 2003) discredited lie detectors, the use of peripheral physiological responses to indicate deception should be suspect. Some agencies complied (e.g., Cumming, 2007), yet the practice (both by agencies of the federal government and by private corporations) continues precisely because public common sense believes in them, so polygraphs extract confessions (Iacono, 2008; Iacono & Lykken, 2005). The neuroscience of deception likely will have a similar pattern, acting as an inaccurate but intimidating polygraph, marketed by those who stand to profit thereby (Fiske & Borgida, 2008). Physiological measurement is always less than a royal road to unbiased assessment, so its role should be limited to well-validated uses (e.g., Kang, Inzlicht, & Derks, 2010).

2.3. What uses?

Assuming the practical theory and evidence, how do organizations use the science? The answer, apparent even from the outside: abundantly. Apart from the list of overlapping handbook chapter topics just noted, the domain itself is clear. If organizational psychology is the behavior of people at work (e.g., Zedeck, 2011), then the domain itself includes just about every human behavior except a full night's sleep (though maybe even this applies to airline pilots and long-distance truckers). In practice, business-related demands often focus organizational uses on management: selection, performance appraisal, teamwork, leadership, motivation, and job satisfaction.

Nevertheless, the practical use of theory and evidence are constrained only by the imagination. What comes to mind for us are areas where we ourselves have a scientific foothold but do not see organizations making sufficient use of our field. For example, the extensive use of psychological scientists as discrimination experts in litigation tends to focus on identifying and preventing bias due to gender, race, and age (e.g., examples to date in our edited volume: Banks, Eberhardt, & Ross, 2008; Crosby & Dovidio, 2008; Eagly & Koenig, 2008; Fitzgerald & Collingsworth, 2008; Heilman & Haynes, 2008; Nelson, 2008; Rudman, Glick, & Phelan, 2008; see also Nelson, 2009, and Dovidio, Hewstone, Glick, & Esses, 2010). Why could the cited science *not* be used to promote the best use of available human capital? If we know that deciders tend to use convenient categories and their attendant stereotypes, unless otherwise attentive and motivated, then *why not* focus the use of psychological science techniques on how to make the best (instead of the same-old) decisions? Critics, of course, ask whether these processes demonstrably exist in organizational contexts, and whether proposed remedies have been tested in organizational settings as well as in laboratory contexts. Mostly, the reviews show, the principles hold up well in both lab and field, but unquestionably more empirical work grounded in work organizations would go a long way to addressing the critics' concerns.

Taking this idea further, research useful outside most business organizations—but applicable in the justice system—may have lessons for issues in management. On the surface, eyewitness identification and testimony, false confessions, and invented memories would seem to have scant role outside the courtroom, but in fact, employees and managers have to report on their own and other's behavior all the time. This science, conducted for another practical purpose, may be useful in organizations more generally than in litigation contexts (for reviews, see our edited volume again: Kassin, 2008; Loftus, Garry, & Hayne, 2008; Wells & Hasel, 2008). In fact, organizational psychology shows considerable research interest in the prediction and modeling of counterproductive work behaviors. Detecting employee deception and dishonesty matters to many organizations, given costs associated with destructive employee behavior.

Another untapped domain familiar to us in psychological science describes the universal dimensions of social cognition. Because organizations consist of people, these principles are potentially useful. People making sense of other people rapidly decide if another individual or group has good or ill intentions (warmth, trustworthiness), and the ability to enact them or not (competence, capability) (for a review of the stereotype content model, SCM, see Fiske, Cuddy, & Glick, 2007). These dimensions appear across dozens of cultures (Cuddy, Fiske, et al., 2009; Durante et al., submitted for publication) and account for 80–90% of the variance in impressions (Wojciszke, 2005), predicting emotions and behavior (Cuddy, Fiske, & Glick, 2007). We know of two organizational applications.

Just recently, thanks to serendipity, these SCM dimensions have been applied to corporate images in branding surveys (Kervyn, Malone, & Fiske, in press). For example, J&J is seen as admirable: well-intentioned and competent, whereas BP is seen as disgusting: untrustworthy and incompetent. Google is enviable: respected but disliked. Amtrak is seen as pitiful: well-intentioned but incompetent. These four kinds of images predict purchase intentions and brand loyalty. Theory and evidence, from basic research in another domain, may prove useful in marketing. Images of other companies may also inform decisions about corporate allies and competitors.

Another SCM application focuses on managing diversity within an organization (Fiske, in press). Not all prejudices are alike, and not all are negative. For example, older and disabled people both are judged as warm but incompetent, whereas both Asians and female professionals are often judged as competent but cold. Poor people and immigrants are viewed as neither trustworthy nor competent. All this has implications for monitoring biases in employment decisions for under-represented groups. Another article in this volume focuses on warmth and competence judgments in organizations more generally (Cuddy & Glick, 2011).

2.4. *An in-depth example: the motherhood penalty*

A number of relevant, applicable studies address what happens to women in the labor market context (U.S. and elsewhere) when they become mothers. The most established disadvantage is the motherhood wage penalty: mothers earn about 5% less per child than other workers, above and beyond any gender wage gap (Benard, Paik, & Correll, 2008; Budig & England, 2001). This motherhood wage penalty is robust, despite statistical controls for past work experience, race, education, work status (full-time vs. part-time), and other occupational and background variables. Several basic-science theories help explain this problem and potential solutions.

Drawing on the just-noted SCM (Cuddy, Fiske, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002), for example, suggests that the motherhood penalty is in part driven by stereotypic associations about mothers as warmer but less

competent than employees without children. An experiment examined perceptions of working men and working women, with and without children, in the context of the SCM (Cuddy, Fiske, & Glick, 2004); student participants rated profiles of three fictitious management consultants on traits relating to warmth and competence, along with other measures of evaluation and hiring intentions. The data supported predictions: female consultants with children were viewed much as housewives are often seen – warm but incompetent (Eckes, 2002). Working mothers were viewed as warmer than female professionals without children, but as less competent; working mothers also were rated less hireable, promotable, or trainable compared to working fathers and female professionals without children. Working fathers were not comparably disadvantaged – their perceived warmth increased compared to men without children, yet they were able to maintain perceived competence. Finally, competence ratings were strongly correlated with behavioral intentions to hire, promote, and educate. The latter findings suggest that the boost in perceived warmth of working mothers is at the cost of perceived competence and their professional standing. The SCM framework suggests that organizations should place extra emphasis on the competence of the working mother as a professional and perhaps also as someone successfully managing complex competing demands. Commitment to the organization should also be evidenced by the decision to continue working after having a child.

Role congruity theory (Eagly & Karau, 2002) provides another theoretical framework for understanding the motherhood penalty and gender prejudice toward women who are mothers. Using a sample of managers and the subordinates who reported to them in a Midwestern division of a global *Fortune* 100 transportation company, a survey (Hoobler, Wayne, & Lemmon, 2009) examined the social-role-theory and gender-stereotype hypothesis that women, more so than men, are viewed from a nonwork, caregiving perspective, and that their nonwork demands (e.g., associated with an employee who is a mother or pregnant) will be salient to their workplace colleagues. Managers' perceptions of family-work conflict revolve around: "...whether women's promotability is hampered by their bosses' perceptions that they experience greater family-work conflict than men and thus have poorer fit with their organizations and jobs" (p. 951). A carefully controlled statistical analysis (structural equation modeling) supported the perception-promotability relationship among both male and female managers, who viewed women as carrying more family-work conflict, even controlling for actual family responsibilities and the women's own rated family-work conflict. Managers' perceptions of women's family-work conflict undermined their judged fit and perceived performance for the female employees. In turn, perceived fit predicted promotions and rated promotability. Another field study and a lab experiment show the same effects (Heilman & Okimoto, 2008; King, 2008).

Arguably, women are often stereotyped as warm and nurturing, so others then assume that women experience conflict between the competing demands of their work and family roles. What is more, people's default image for managers is closer to their image of men than women (Heilman, 1983), so women already face a lack of fit as managers that the expected caregiver role only exacerbates. The cure here is debunking the myths that women are necessarily caregivers, that women necessarily experience more work-life conflict than men, and that they necessarily make worse employees if they do. One could easily argue the contrary, that women are accustomed to juggling these role demands and have learned to handle them (Williams, 2010). Women succeed despite these challenges, so they should be compensated accordingly.

The motherhood penalty also fits the status characteristics theory, which suggests that stereotypes tend to associate greater competence and worth for members of some social categories, such as nonmothers, than for members of other social categories, such as women who are mothers (Ridgeway, 2001; Ridgeway & Correll, 2004). The performance of lower-status individuals, compared to higher-status individuals, is subjected to greater scrutiny and monitoring. Motherhood qualifies as a status characteristic based on the evidence that mothers are perceived to be less competent in general than fathers or women without children (Correll, Benard, & Paik, 2007). Mothers may be perceived as less committed to their work roles in part because they are victims of the clash between cultural beliefs about the ideal worker (with single-minded allegiance to work – a cultural schema of work devotion) and the ideal mother (with single-minded allegiance to her children – the family devotion schema) (Blair-Loy, 2003).

In line with mothers being judged by harsher performance standards than nonmothers, two studies examined parental discrimination in workplace settings (Correll, Benard, & Paik, 2007). In the first, college students rated two applicant packets, pairs of the same sex, race, and qualifications for the job. Mothers were rated as being less competent and committed than nonmothers, given lower starting salaries, and permitted fewer days of absence/tardiness than nonmothers. Mothers were also much less likely to be hired than nonmothers (87% vs. 47%). Conversely, fathers were favored over nonfathers when it came to commitment and salary recommendations, so fatherhood advantaged the men. Further, individuating information does not typically combat bias: participants saw a

fact sheet, brief resume, and interview notes including information career goals, college GPA, and past work experience. Discrimination still occurred. A linked audit study mailed the same pairs of resumes to actual employers in response to real job postings. Mothers were called back half as often as nonmothers, whereas the father–nonfather difference was not significant. And the effect sizes for the audit study were virtually identical to those obtained in the student sample.

As these studies suggest, organizations managing worklife policies must deal with broad-scale adoption of new ways of working in order to recruit and especially to retain valued employees who struggle to balance work-family issues (Williams, 2010). Changing workplace practices that disadvantage parents, both women and men, therefore represents a significant opportunity for organizations to utilize scientific evidence to develop remedies that increase the likelihood of continued productivity and well-being for employees who also are parents.

3. Criteria for quality and consensus

We have merely illustrated some theories and some science that might be useful to organizations, for much social psychology is relevant to organizational issues. Having provided exemplars, this article next addresses how to establish scientific consensus about what we know or do not know, emphasizing two criteria, quality and consensus. We begin with quality.

3.1. *Quality science = peer-reviewed science*

Elsewhere, we have argued that quality science merits a place in the legal system: What legal scholars describe as improving “behavioral realism” in the assumptions made by judges and juries, whose decisions of course set the law’s precedents (Fiske & Borgida, 2008, 2011; Kovera & Borgida, 2010; Krieger & Fiske, 2006). The definition of quality science requires that we delve deeper into the editors and reviewers who determine what constitutes quality science. In the law, at least, this gatekeeper role is crucial because judges and jury-eligible publics are non-experts who cannot be expected to differentiate quality from junk science (Cutler & Kovera, 2011). For example, neither judges nor prospective jurors appreciate avoiding confounds and blinding experimenters to condition, two fundamentals of psychological science methods (Kovera & McAuliff, 2000; McAuliff, Kovera, & Nunez, 2009).

The gatekeepers in organizations are a special subset of the jury-eligible public, in several respects. Although managers and HR specialists are likely to have more scientific training than judges and juries, they cannot be topical specialists in all the sciences that might inform management. Assuming they are interested in the useful sciences (an assumption we address later), they must allow proxies to vet the science for them. Management gurus cannot be presumed to rely on a responsible review of the literature. Some popular business magazines and some consulting firms dabble in the relevant science, but of course they have a vested interest in promoting their own sales. Let us assume the best case, in which an honest broker summarizes the relevant available science. How is that organizational messenger to judge quality science? Peer review is the scientific monitor of choice. Peer review trusts experts to vet each other. Nevertheless, skeptics of peer review ask: How good is peer review? What is to prevent mutual back-scratching, simply publishing each other’s work?

Peer review dates back to 1665, when the Royal Society licensed *Philosophical Transactions* with the stipulation that its content be reviewed prior to publication by members of its Council (Chubin & Hackett, 1990, p. 19). In the 19th and early 20th century, *Nature* and some medical journals followed suit (Weller, 2001). By 1962, three quarters of scientific journals sampled across 10 countries reported peer review. All journals vest some decisions in the editorial office, for example, publishing conference abstracts, archival documents, news, and obituaries, as well as the increased triaging of desk-rejected submissions without external review.

Defining peer review helps to decide what it does and does not do for organizational decision-makers interested in quality science. The simplest definition is that a peer-reviewed manuscript has been evaluated by someone other than the editor (Weller, 2001, p. 16). Reviewers’ professional expertise and relevance are vetted by the journal staff, the editor decides after weighing the reviews, and the author receives feedback, usually blinded to the identity of the reviewers. The peer aspect stipulates that one’s equals assess the quality of one’s scholarly work (Eisenhart, 2002). Peer review serves as a systematic method to evaluate the soundness of scientific methods, the plausibility of results, and the allocation of scarce resources, such as journal space, and (indirectly) tenure, promotion, reputation, and honors (Chubin & Hackett, 1990). Peer review also confirms the status of established scholars who have run its gauntlet (Miller, 2006).

To be sure, skeptics call peer review a chimera (Chubin & Hackett, 1990; Eisenhart, 2002), claiming that it is overextended, impedes creativity (due to its scientifically conservative bias), undermines author autonomy, demoralizes authors, wastes time, diffuses responsibility, and relies on peers who are competitors with intrinsic conflict of interest. Notably, none of this fits the fear that peer review simply creates a mutual back-scratching cabal.

As even cynics (Bedeian, 2004) agree, the underlying dilemma is how to preserve its representation of scientific integrity—impartiality, validity, and reliability—as well as preserving authors' voice. This process always entails compromises. Sometimes standards seem too rigid (affecting mainly one's own rejected manuscripts of course), but sometimes standards are too low because of unqualified reviewers (affecting mainly other people's unaccountably accepted manuscripts). Or standards may be too conventional because of reviewers' predispositions due to being more senior, established experts, averse to controversy (Eisenhart, 2002). Balancing innovation and quality can be a challenge (Rynes, 2006).

From some authors' perspective, reviewers seem ignorant, unmotivated, and unaccountable, while editors seem just plain ignorant (Tsang & Frey, 2007). Reviewers may appear biased by institutional prestige, presentational complexity, confirmation bias (Miller, 2006). The evidentiary utility of these opinions remains up in the air, but demonstrates that not all scientists appreciate the wisdom of peer review.

Disgruntled commentary aside, empirical studies of peer review are surprisingly scant (Bailar & Patterson, 1985), for several probable reasons. First, one has to agree on standards for assessing the process and the outcome. The process could be fair without the product being trustworthy. That is, the process could seem impartial and expedient without the product being dependable, innovative science (Chubin & Hackett, 1990). For example, the recent publication of Bem's work on ESP (2011) followed all the rules of peer review with repeated revisions, and eventual acceptance by a methodologically sophisticated editor, but few psychological scientists believe the results. This could be an error of commission (publishing something unworthy), but peer review also risks errors of omission (overlooking good work), partisan or prejudiced judgment, inconsistency of judgment, and lack of consensus among reviewers, according to the experiences of disgruntled authors (Chubin & Hackett, 1990).

Another reason that scientific evaluations of peer review are scant is that the investigator has to gather data, as Weller (2001) did from 139 journal-editorial-policy letters in which editors comment on peer review; the results are miscellaneous, but only about a quarter of journals across disciplines mention blind review, surely the gold standard.

The most specific research addresses reviewer agreement, most studies coming from psychology, sociology, and medicine (Weller, 2001). Editors agree on the theoretical and actual reasons for rejection or acceptance, as do investigators. Across studies, reviewers tend to agree about half the time, with the rejection of a particular manuscript averaging about 45% agreement, but acceptance agreement only about 22% (Weller, 2001). Another study of reviewer dissensus (Miller, 2006) reported a wider range (due to differing expertise, but also inconsistency). Editors may recruit reviewers to fill complementary areas of expertise, so they do not necessarily identify the same issues (Fiske & Fogg, 1990; Fogg & Fiske, 1993).

Successful peer review aims to be effective (sensitive to quality), rational, fair, reliable, valid, efficient, accountable for regulations (COI, IRB) and reviewer performance, responsive to shifting policy needs and new science, but robust to fads (Chubin & Hackett, 1990). Peer review can bridge from self-regulated, esoteric, sequestered science to policy about resource allocation (e.g., public control, agency mission, accountability, due process) (Chubin & Hackett, 1990).

Doubtless, peer review could improve, and the suggestions begin with increasing reviewer accountability by vetting them, preparing them, assessing them, and recognizing them (Bedeian, 2004; Chubin & Hackett, 1990; Miller, 2006; Raelin, 2008). At the next level, the process could allow authors to respond to the reviews before the editor commits to a publication acceptance or rejection (Chubin & Hackett, 1990). In a variant, authors could write discursive footnotes responding to reviews, or the reviews could be published (Bedeian, 2004). The most radical suggestion is to limit editors and reviewers to an accept/reject decision, with revisions made to accepted manuscripts solely at the author's discretion (Tsang & Frey, 2007); this as-is method preserves the authors' voice against excessive tinkering. Formal appeals and letters to the editor might also open up the process (Bedeian, 2004).

In theory, science might anoint a professional cadre of professional reviewers, as in art, theater, and literature (Raelin, 2008), but science moves so fast and requires training in the execution of the work itself. Science is not a spectator sport, and kibitzers cannot keep up with the latest technical developments. A lottery is not a serious alternative, nor is the popular election of individual scientists to publish whatever they want.

As a foundation of quality science, then, peer review is hardly flawless, but it serves as well as we can do. In the end, "No one questions either the role of referees in advising editors regarding the 'publishability' of a submission or that

referees can, on occasion, protect authors from themselves, as in cases involving slipshod work” (Bedeian, 2004, p. 206). Peer review is indeed a tripartite balance among authors, reviewers, and editors (Raelin, 2008). Scientific standards are socially constructed: consensual, contextual, argumentative, and human (Chubin & Hackett, 1990). The marginal costs of imprecise outcomes may be self-correcting, in any case, with replication (Bedeian, 2004). Meritocracy ultimately is undemocratic (though it need not be anti-democratic; Fiske, 2011). A wise methodologist once noted that peer review provides scientists’ only major on-the-job training. And it signals quality science for external use.

3.2. *Establishing consensus*

If the first criterion for psychological science being practical for organizations is its peer-reviewed quality, the second criterion is scientific consensus. Basic scientists need literature summaries as much as end-users do, and several methods address both needs. All methods aim to represent consensus as the general, majority sense of those experts who produce the science and who have the credentials to support their expertise. For example, the consensus of astrologers would not count as scientific authority because they are not recognized as scientists, which requires doctorates in the relevant accredited field, research support by federal agencies, and publicly accountable scientific methods. And, contrary to some naysayers, scientific consensus does not involve some sort of biased “popularity contest” (Tetlock, 2010, p. 253).

In fact, consensus need not imply unanimity. In any ongoing scientific field, the overall pattern contains a distribution of outcomes and judgments; outliers do not discredit a reliable average result with a tightly clustered set of sufficient replications. As we note below, moderator variables may indicate that more than one result occurs, depending on specifiable circumstances. Neither variance nor modulation indicates disarray and inconclusive patterns. Some dissenters interpret exceptions, variation, or moderators as discrediting an established effect. The burden rests on them to show that their alternative interpretation fits the majority of the available evidence. This section suggests several ways to do so.

Partisans may of course exploit disagreement, as in the hijacking of the IAT controversy in a move to discredit all social cognition research and its use in organizations, courtrooms, and policymaking. For example, IAT-skeptics claim major unresolved issues in definition and measurement, generalizability problems, and failure to predict behavior (Tetlock & Mitchell, 2009). IAT-endorsers counter-claim “The existence of implicit bias is beyond reasonable doubt, [providing] a refutation of ideological and methodological objections and executive summary of . . . studies that no manager should ignore” (Jost et al., 2009, p. 39). As part of their evidence, they note that the methods identifying implicit bias build on a century’s worth of cognitive science; to reject the evidence base, “one would need to set aside so much of social and cognitive psychology that both disciplines would be rendered unrecognizable to contemporary students and scholars” (pp. 39, 45). This kind of disagreement concerns incompatible readings of the accumulated literature, to which we now turn, after one last preambulatory comment.

Characterizing the state of the field differs in kind, not merely magnitude, from characterizing a particular datum. In court, for example, each side hires competing experts, and courts rightly evaluate both partisans’ experts for nonobjectivity. For example, in one test of adversarial bias in the evaluation of the psychopathy of violent offenders, empirically, nearly two dozen instances showed each side’s own certified experts differentially interpreted the same clinical data, and in the direction expected by the party retaining them, although normally inter-rater agreement is high (Murrie, Boccaccini, Johnson, & Janke, 2008). In management consulting also, advice readily lends itself to bias, perhaps what the client wants to hear. In some cases, companies may hire consultants de facto to justify what they already intend to do.

Of course, competing experts can characterize an entire literature differently to suit the party that retains their services, but more publicly reproducible methods provide superior support for reading the research consensus than does a particular, for-hire clinical or consulting judgment that relies on less objective methods, such as one person’s experience or hunches, however astute they may seem. Even collective hunches—in the form of benchmarking to successful firms—represent only collective hunches. Here, we suggest surveys of experts, meta-analyses, narrative literature reviews, adversarial collaborations, amicus briefs, and professional society consensus (“white paper”) reports.

3.2.1. *Surveys of experts*

When an organization seeks expert opinion, sometimes that consensus could result from a literal opinion poll. For example, to decide whether psychological science is relevant to the real world, one could ask psychological scientists.

One study did just that (Lipsey, 1974): 2340 graduate students (50% of those contacted in 109 departments) and 368 faculty (74% response rate) reported on their concerns about psychology as a science. About half of the faculty and students spontaneously raised the relevance of their science to social problems, the real world, and to application. From this, one could conclude something approaching a consensus 40 years ago that relevance needed more attention. The results are thought-provoking but do not imply specifically actionable directions for the field.

A more specific survey of experts focused on the reliability of findings in a particular domain by surveying expert opinion on the general acceptance of conclusions about eyewitness testimony (Kassin, Ellsworth, & Smith, 1989). As noted earlier, this literature might be relevant beyond the courtroom, for example in reporting on peers' behavior in 360° evaluations or in whistle-blowing. Of relevant published experts who responded, 80% endorsed the conclusion that eyewitness testimony can be biased by “the wording of questions, lineup instructions, misleading post-event information, the accuracy-confidence correlation, attitudes and expectations, exposure time, unconscious transference, show-ups, and the forgetting curve” (p. 405). More than 70% endorsed principles of lineup fairness, cross-race identification bias, and overestimating event duration.

Such surveys prove useful in other instances seeking expert agreement. A survey of editorial board members of major human-resources journals elicited consensus about the field's major findings (Rynes, Giluk, & Brown, 2007): intelligence predicts performance, goal-setting and feedback motivate, human resources matter to organizational outcomes, structured interviews outperform unstructured ones, selection criterion validity matters, and personality predicts performance. Such a survey identifies areas of consensus ripe for export into organizational settings, either as a basis for introducing new policies or as potential remedies for flawed practices.

Because surveys have well-established methods, and sampling frames can specify the criteria for selecting experts, as well as their response rate, surveys of experts are a viable way to establish the degree of consensus on a specific point. Such surveys do have limits, to be sure. Experts may agree more about the intended consequences of interventions but not about the unintended consequences (Brief, personal communication). For example, despite the hundreds of studies showing the effectiveness of goal-setting (Locke & Latham, 2002), and the consequent expert consensus on this point, few experts have considered some undesirable side-effects of goal-setting (Ordóñez, Schweitzer, Galinsky, & Bazerman, 2009).

3.2.2. *Meta-analysis*

A more common method, quantitative literature review, entails a meta-analytic search for the extent and consistency of answers to a focused empirical question. Having identified and isolated an effect from dozens of carefully sampled studies, often both published and unpublished, the investigator calculates the average effect size. More important, the meta-analyst identifies moderator variables that describe contextual variations in the effect size. Because meta-analyses have well-established methods and vetting by peer review, they reach a high standard of objectivity and utilize quality science for summarizing scientific consensus.

In addition, meta-analyses can identify a priori indicators of scientific quality, testing effect sizes with and without selecting studies on quality. For example, in one recent video violence meta-analysis (Anderson et al., 2010), quality criteria include whether the independent variable had appropriate levels (really violent and really nonviolent games), an otherwise properly operationalized independent variable (isolating hours of violent games from all games played), internal validity (subject self-selection not confounded with experimental conditions), appropriately operationalized dependent variable (aggression toward a person, not an object), a predictably sensitive dependent variable (state aggressiveness, not trait aggressiveness), and proper computation of the dependent measure (not averaging pre and post scores). As long as such quality criteria are applied without bias, they can enhance the information provided.

A variety of quality controls reinforce the value of the meta-analyses themselves. Published meta-analyses follow established scientific criteria, which typically require systematic sampling of the literature, including unpublished studies to counteract the “file-drawer” problem, and then the study's publication status itself becomes a moderator variable. Because meta-analyses specify their samples, methods, and calculations, in order to pass peer review, and because as in all science, they are subject to replication, rebuttal, and rejoinder, the level of confidence in their results is high.

As for organizational relevance, one of the most impressive programs of meta-analysis is Eagly's series on women leaders (for one collection and discussion, see Eagly & Koenig, 2008). From this series, one can conclude with some certainty that there is a strongly masculine cultural stereotype of leadership (Koenig, Eagly, Mitchell, & Ristikari, 2011) and that, compared with men, women are almost as motivated to lead (Eagly, Karau, Miner, & Johnson, 1994),

are equally effective when they do lead (Eagly, Karau, & Makhijani, 1995), do not much differ in task or interpersonal leadership style, though women do often lead in a less autocratic and more democratic style (Eagly & Johnson, 1990), a style that is favorably evaluated regardless of leader gender (Eagly, Karau, & Makhijani, 1995), and women engage in more transformational leadership that is effective regardless of gender (Eagly, Johannesen-Schmidt, & van Engen, 2003). This does not mean that women face no disparate challenges as leaders because clearly they do (Eagly & Carli, 2007). When women do lead in more autocratic styles, when the setting is male-dominated, or when evaluators are male, women leaders are evaluated less favorably (Eagly, Makhijani, & Klonsky, 1992). Women also emerge as task leaders less often and social leaders more often (Eagly & Karau, 1991), perhaps because women generally help in more long-term ways (Eagly & Crowley, 1986) and aggress in less overt ways (Eagly & Steffen, 1986). Most of these effects are similar whether inside and outside of organizational contexts.

Further, as Eagly and Wood (1991) note, all these meta-analyses not only summarize the research literature but also examine mediators and moderators, thereby testing the plausibility of competing explanations. Much of the theoretical and practical interest of meta-analysis lies in the mediators and moderators. The effect sizes averaged over moderators may appear small, but comparing the magnitude of even small effects within and across fields, psychology's effect sizes compare favorably to those in physics. Hedges (1987) compared 13 reviews from the Particle Data Group in physics and 13 reviews from psychology (e.g., the effects of teacher expectancy on IQ, the effects of desegregation on educational achievement, sex differences in spatial ability). The psychological reviews were slightly more consistent across studies than the physics reviews, despite evidence of some inconsistencies in both fields.

For organizational uses, the moderators identify settings that may qualify the export of the effects into a particular setting, making the effects likely larger or smaller than average. The moderator analyses therefore indicate potential cures for problems of the kinds identified in Eagly and colleagues' meta-analyses. For example, as discussed, even though differences in evaluations of male and female leaders are small, women leaders fare better in gender-integrated workplaces and when they are evaluated by both men and women, lending some plausibility to the role of gendered biases. That is, the reliable moderator pattern indicates that women leaders suffer most when evaluators are disproportionately male; this suggests the importance of representing women among the evaluators. Other moderator variables suggest other potential remedies for gender biases.

All in all, multiple replications have confirmed many findings in social psychology in particular and the social sciences in general; quantitative, meta-analytic reviewing has established their robustness. The identification of moderators may reveal more substantial effects under specifiable conditions, making the accumulated findings even more applicable to organizational settings. Moderators also play an important role in newer, more nuanced indicators of generalizability, specifying the extent to which effects occur under specified conditions.

3.2.3. Narrative literature reviews

For some broader purposes, qualitative literature reviews can summarize several themes in a general topic area. Peer-reviewed narratives can characterize trends in a subfield. The top-ranked journal in the field, the *Annual Review of Psychology*, commissions just such reviews, based on expert author selection and subsequent expert manuscript review (possible bias alert: Fiske currently edits the *ARPsy*, and Borgida has served on its Editorial Committee). The impact of this model is further indicated by new *Annual Review* volumes in clinical psychology, in industrial/organizational psychology, and in law and social science. Other psychology review journals also rank high in impact factor, the most notable being *Psychological Bulletin*; sometimes *American Psychologist* also publishes a narrative review of general interest. And of course, within more specialized subfields, each has its review journals (e.g., *Personality and Social Psychology Review*). Because such journals are peer-reviewed, typically with a 90% rejection rate, they meet the quality criterion.

The purpose of an *Annual Review* article differs from the more focused meta-analysis, in aiming to overview a moving window of a general topic's current standing, rather than (usually) answer one specific question in detail. For example, recent organization-relevant articles have described the perpetrators (Barling, Dupré, & Kelloway, 2009) and victims (Aquino & Thau, 2009) of workplace aggression. A more conceptual set of organizational questions appear in a review of affect in the workplace (Brief & Weiss, 2002). Organizationally relevant issues of human behavior arise in reviews of human-computer interaction (Olson & Olson, 2003) and the Internet's role in social life (Bargh & McKenna, 2004). Another organizationally related qualitative literature review identified conceptual and process issues in the extant diversity literature (van Knippenberg & Schippers, 2007). Clearly, practical issues appear in a review of what works in prejudice reduction (Paluck & Green, 2009). This last review provides a clear comparison

with meta-analysis. The Paluck-Green narrative review identifies a variety of methodological issues that plague the literature on contact, whereas the Pettigrew and Tropp (2006) meta-analysis focuses on a point-estimate of the effect sizes in published studies under different moderator conditions. Narrative and quantitative reviews are suited to answer distinct questions.

Competing qualitative literature reviews are harder to resolve than competing meta-analyses, though both involve selection and interpretation to differing degrees and explicitness. A narrative review aims to provide coherence and to identify themes, as an intrinsic goal. However, one author's theme is another author's slanted bias. For example, in the domain of gender and work, several factors might affect the differential hiring, evaluation, and promotion of women: stereotypes, gender roles, and subjectivity of criteria (respectively: Eagly & Koenig, 2008; Heilman & Haynes, 2008; Rudman, Glick, & Phelan, 2008; granted, these contributions are not peer-reviewed except by the authors of this article serving as editors). Other authors dismiss the applicability of the included laboratory research (Copus, 2005) or research conducted outside a work setting (Landy, 2008). Differing narratives must respond to the representativeness and comprehensiveness of the reviewed literature, just as in meta-analysis.

The best narrative literature reviews are strong in identifying themes, making an argument for a given interpretation, and in short, making sense of the sprawling research domain. The criteria here are subjective, but peer review and editorial discretion determine whether the perspective is convincing and well-documented. In contrast, the best meta-analyses test specific, theory-driven (and hence principled) hypotheses that are falsifiable. Just as in a narrative review, the conceptual foundations for meta-analysis also can make sense of a sprawling literature. From our perspective, a meta-analysis that does both is both rare and preferable, but narrative reviews continue to serve an indispensable function when results are more scattershot, replicated less often, involve incommensurate methods, or are not reducible to a few key tests. Narrative reviews also convey a particular expert voice, which usually increases the credibility of the review; the importance of expertise in judging overall themes and trajectories has merit.

3.2.4. *Adversarial collaboration*

Another approach to establishing scientific consensus is adversarial collaboration (Kahneman, 2007; Mellers, Hertwig, & Kahneman, 2001). Adversarial collaboration focuses more narrowly on reaching scientific consensus on a specific controversy. The approach brings together two sides on an evidentiary matter of scientific disagreement. Each party agrees about the methods necessary to operationalize competing perspectives on a phenomenon, and both parties supervise the proposed research, its analysis, and resultant publications. The first author participated in an adversarial collaboration, specifying the conditions for activating a portion of the medial prefrontal cortex, whether due more purely to any rewards, or more specifically social cognition (people are intrinsically rewarding). The results suggested that the answer appeared to involve both, especially when rewards were social in nature (Harris, McClure, van den Bos, Cohen, & Fiske, 2007; van den Bos, McClure, Harris, Fiske, & Cohen, 2007).

Calls for adversarial collaboration recently have come from critics of the IAT and the external validity of implicit bias research (Tetlock & Mitchell, 2009), for example, claiming “. . .no empirical basis for believing that personnel decisions studied in artificial laboratory settings approximate personnel decisions in real-world settings” (Tetlock, Mitchell, & Murray, 2008, p. 442) and that only an adversarial collaboration involving researchers from opposing camps will resolve the matter. These claims, however, tend to underscore the difficulty of successfully launching adversarial collaborations. Some critiques of the IAT and implicit bias research have raised pointed ideological and normative claims, as well as scientific challenges, that de facto have made it difficult to bring the opposing parties to any sort of adversarial collaboration table (Tetlock & Mitchell, 2009). Accusations of reductionist blinders and misrepresented scientific studies do not provide much of an incentive to engage in adversarial or any other type of collaboration.

For example, a science-based rebuttal to these criticisms defends implicit bias research and the IAT's predictive validity through scientific findings from cognitive and social psychology (Jost et al., 2009). However, Jost et al. reveal no inclination to resolve this controversy via adversarial collaboration in part because of the *ad hominem* quality of some of the critiques, but also in part because they argue that the dispute is more apparent than real, thus calling into question the necessity of an adversarial collaboration to resolve the dispute. Both sides do however endorse a research agenda that tests implicit bias in organizational settings using multiple methods and measurements.

3.2.5. Professional society consensus reports

Finally, professional organizations and societies occasionally commission expert consensus reports that come in different forms. One such form is the scientific white paper that brings together experts with different viewpoints who then craft a consensus position on the research questions posed. These white papers are usually scientific review papers that cite quantitative meta-analyses in their data base, but are not themselves typically quantitative meta-analyses. Once consensus among the assembled expert team has been achieved, these products are then reviewed, revised, and disseminated in venues that range from the journals published by the professional society that commissioned the report (e.g., the Society for the Psychological Study of Social Issues or the Association for Psychological Science) to the consensus reports issued by the National Research Council, an arm of the National Academy of Science. Some professional societies, such as the American Psychology-Law Society (AP-LS), have developed an extensive vetting procedure for white papers that are ultimately published in the Society's lead journal, *Law and Human Behavior*. AP-LS commissioned such a paper on police-induced confessions, assembling an intellectually diverse team of leading scholars in the field to review the empirical literature on interrogations and confessions in order to establish consensus findings and best practices recommendations (Kassin et al., 2010). The idea for the white paper was first approved by the Society's executive committee, then drafts were vetted by AP-LS's Scientific Review Paper Committee, four peer reviewers at the journal, and its Associate Editor and Editor.

Even extensively vetted consensus reports cannot guarantee that consumers will heed the advocated recommendations. Often, gold-standard scientific procedures clash with preferences for the administrative expediency of current practice. Such was the case, for example, with the field study of eyewitness identification conducted in three metropolitan police jurisdictions in Chicago (Diamond, 2008). Proposals for changing investigative practices in eyewitness identification, based on decades of experimental psychological science, focused on using a line-up administered in a double-blind, sequential (vs. simultaneous) fashion, to reduce the influence of administrator bias and to maximize the rate of accurate identifications. But in two of the three jurisdictions, the field study compared a non-blind simultaneous procedure to only a blind sequential procedure, thereby confounding the test. The obtained results, contrary to numerous lab studies, favored the former (doubly contaminated) technique to the latter. Despite the incomplete and therefore inconclusive research design (unable to resolve whether results were due to the simultaneous procedure or to non-blind line-up administrators), the study ended up on the front page of the *New York Times* as allegedly having showed that current police department practice was in fact not as biased as experts had thought. In response, a team of leading scholars, subsequently assembled by the AP-LS, examined this dispute and the inferential constraints associated with the methodological confound in the original study; even this high-powered team has had mixed results in persuading police departments that extant practice may still need reform (Schacter et al., 2008).

A second common form of expert consensus reports commissioned by professional societies involves *amicus curiae* briefs typically submitted to appellate courts and the U.S. Supreme Court, where the litigation touches on scientific issues of interest to a particular professional society. The American Psychological Association, for example, routinely submits *amici* in those cases where peer-reviewed research by APA members bears directly on the empirical claims at the heart of the legal issues at stake. In the Title VII sex discrimination case, *Price Waterhouse v. Hopkins* (1989), APA filed an *amicus* brief in order to highlight the body of social scientific knowledge that was consistent with the scientific evidence introduced at trial, and that provided a substantive science-based context for considering the case (Fiske, Bersoff, Borgida, Deaux, & Heilman, 1991). Other professional societies such as the American Sociological Association and the Law and Society Association also have filed *amicus* briefs to inform legal analyses at the appellate level. Most recently, the ASA and the LSA filed an *amicus* with the U.S. Supreme Court in the *Wal-Mart Stores, Inc. v. Betty Dukes, et al.* class action case that was decided during the 2010–2011 session. As with other expert consensus reports, the goal of this *amicus* was to address the empirical questions that pertain to whether the female Wal-Mart employees should be certified as a class:

Amici offer their expertise to aid the Court in evaluating the use of social science evidence such as the evidence in the record in this case as a source of valid information about the effects of corporate culture and policy on potential discrimination, including the effects of corporate culture on individual-level decision-making (p. 2).

This is an organizationally relevant, qualitative research approach to the legal record that the ASA refers to as an organizational case study.

In summary, established standards for peer-reviewed science vet its quality, and various summary methods (surveys of experts, meta-analyses, narrative reviews, adversarial collaborations, consensus reports) identify expert agreement.

The legal settings that we cite provide a particularly rigorous test of using psychological science because the adversarial setting is explicit, argument is open, and judgments are public. Most organizational choices between competing options are far less systematic, but face similar challenges to identifying quality scientific consensus.

4. Linkage to real organizational life

Having identified high-quality scientific consensus, academics agree that the evidence is ripe for export. Unfortunately, the most plausible users of that evidence do not receive it. MBAs and managers have no ready access (Rousseau, 2006), and intermediary publications fail to bridge the gap. Still, evidence-based management is possible (Guest, 2007). How? This section discusses (a) persuasively communicating quality scientific consensus to practitioners, (b) acknowledging why organizational gatekeepers may not be interested or convinced, and (c) how to generalize from broad principles to understand specific situations.

4.1. *Communicating science to practitioners and persuading them to use it*

The requirements for bridging from science to practice (Rynes et al., 2007) start with a clear scientific consensus, the focus of the preceding sections. Also, the findings must be important, nontrivial, and not too obvious. In human resources, for example, the role of intelligence and integrity assessment are under-estimated but should be central. However, a content analysis of intermediary publications (e.g., *Harvard Business Review*) showed that these topics were either neglected or covered inconsistently.

Even in leading professional journals, only about half the articles provide implications for practice (Bartunek & Rynes, 2010). All too often, harried practitioners seek management magazines and consultants who provide a more grounded, shared, easily-absorbed frame of reference that may or may not reflect the best evidence. After all, magazines and consultants need revenue, too, so they cut corners and may substitute appealing anecdote for rigorous data. So, what is a busy manager reasonably expected to do, in order to keep up?

Organization-relevant science needs ambassadors who speak the appropriate language and can convey useful science in accessible venues. Scientific societies that educate and advocate (e.g., SIOP, FABBS, APS, APA, SPSSI) provide a variety of venues such as evidence-based websites, speakers, interviews, blogs, and columns.

4.2. *Organizational gatekeepers may not be interested*

Even if quality information about scientific consensus is available, people use their cultural, professional, and personal worldviews to interpret it. Going back to the video-game violence effects, for example, even a highly-vetted, mega-sampled, unambiguous meta-analysis (Anderson et al., 2010) will not convince people who want to believe the evidence is still inconclusive because their identity, self-interest, or values lead to arguments for keeping such games available (Huesmann, 2010). People, including managers, tend to believe that everyone sees reality the way they do (see Kahan, Hoffman, & Braman, 2009, for a high-level example). And as noted, we all misperceive ourselves as being less biased than most people (Pronin, Gilovich, & Ross, 2004). This leads organizational gatekeepers to believe their own interpretation is the only reasonable one, especially if their disdain for scientific consensus is based on a strong preference for certain forms of knowledge (e.g., intuitive anecdote) over others.

Managers also want to know what other managers are doing. Organizations benchmark to what other organizations do. Just as in individual life, so too in corporate life, conformity to opinion leaders and evolving norms determine behavior, and what the scientists suggest seems less relevant than what peers are doing (Staw & Epstein, 2000). Although they do not predict the bottom line, a market leader's rated admirability, perceived innovation, and apparent quality make other companies adopt popular management techniques, following these leaders but not necessarily the science.

Practitioners may be less than uninterested; they may actively resist evidence from laboratory studies, worrying about external validity (specific-to-general inference). Of course, they may not even get so far as weighing lab versus field research, but when they do, this skepticism arises so regularly that some preemptively rebut it: "A blanket call for more field and fewer laboratory studies is, therefore, unwarranted. In the final analysis, researchers must select the mix of methods most appropriate to the research question(s) under investigation, taking into account the overall maturity of the theories and concepts underpinning their work" (Hodgkinson & Healey, 2008, p. 405).

The lab-field mix depends on one's purpose. Given an emerging principle, a mix of laboratory and field methods might be appropriate and amenable to Cialdini's (1980) full-cycle approach. Given well-established general principles and processes, the moderator question arises, so one might test in specific contexts because external validity is an empirical question. Pending such a test, the presumption that laboratory effects will be unrealistically large is not at all justified, as discussed earlier. Nor should one presume as a default that principles of human behavior do not generalize, when common sense is the only counter-evidence. For example, although laboratory and field studies of justice may not always agree, three types of justice—distributive, procedural, and interactional—do not differ much in their relationship to performance across lab and field settings (Cohen-Charash & Spector, 2001, 2002).

Nevertheless, social science lives by context, so a challenge to laboratory findings should specify what features of context (i.e., what moderators) likely matter (Locke, 1986). For example, Landy (2008) stipulates several features of performance evaluation that in his professional experience should matter in the workplace. In this view, contemplating decades of relevant laboratory research, various factors might prevent generalization: evaluator–evaluatee acquaintance, evaluator experience, evaluator information-seeking, evaluator accountability, evaluator outcome-dependency, evaluator frequency of evaluating this domain, evaluator's knowledge of evaluatee's history, and evaluator's complexity of judgment. Although each of these factors has in fact been manipulated in laboratory settings, the claim is that their combination creates a unique setting that must be mimicked precisely, in order to generalize (Tetlock & Mitchell, 2009). The implication is that research only in an organization similar in all its specifics will inform an understanding of scientific principles' utility for practice. Carried to an extreme, external validity is impossible because each context is unique. This begs the question, externally valid to what? How similar does the organization have to be? Basic science provides some principles to inform which dimensions of context matter (Locke, 1986).

4.3. Generalizing from broad principles to specific situations

Given all these challenges, how and when can social scientists effectively translate general, aggregate-level data to understand specific cases? In their role as expert consultants, can organizational and social scientists act as honest brokers to a particular company on a particular matter? This question underlies thinking about organizationally-specific research questions from the standpoint of peer reviewed, aggregate-level, social-science findings: in the absence of specific organizationally relevant research findings, to what extent can researchers interested in, say, the causes of robust economic disparities between male and female managers in a company, use general-causation principles from the pertinent body of social scientific studies to offer reliable organization-specific recommendations? In fact, some suggest that organizationally grounded research is absolutely crucial to addressing such questions about the extent to which individual-level biases drive institutional-level disparities (Smith, Brief, & Colella, 2010) and the extent to which individual-level approaches represent a viable way to remedy institutional biases (Henry, 2010).

This question applies to general-causation principles that an expert witness might use to offer an opinion about the specifics of a legal case. As just discussed, expert consensus reports in the form of *amicus* briefs filed in appellate cases have relied on quality, peer-reviewed science to answer the legal version of this question in the affirmative. For example, in the 2010 *amicus* filed in *Wal-Mart Stores, Inc. v. Betty Dukes, et al.* by the American Sociological Association and the Law and Society Association, the use of aggregate-level social science evidence by the trial expert as a source of scientifically valid information about the effects of corporate culture and policy on potential discrimination was affirmed, and the *amici* urged the Court to consider such evidence in this particular case about class certification.

More generally, questions about linking general social science research and specific case questions have generated a lively debate among social scientists and legal scholars with expertise in employment discrimination class action cases. Some scholars have argued that using aggregate-level science to provide an overall context for the fact finder (i.e., judge or jury) is appropriate, but making specific applications to the facts of the case, involving specific claims about causality, is inappropriate and should be made only by the lawyers in the case (Monahan, Walker, & Mitchell, 2008; Mitchell, Walker, & Monahan, 2011b; Mitchell, Monahan, & Walker, 2011a). For example, some argue that reasoning from general scientific data to individualized legal decision-making represents the challenge of “evidentiary incommensurability” (Faigman, 2010) – that general science may be probative but not determinative and often unacceptably speculative in particular cases. Advocates for appropriate fit, between the scientific evidence presented and the specific legal issues(s) to which it is directed, argue a gap between extant aggregate-level research findings and specific employer decisions in employment discrimination disputes (Faigman, Dasgupta, & Ridgeway, 2008).

By contrast, other scholars do not categorically exclude linkage between aggregate-level data that address general causation, and specific case questions. The legal argument holds that the Federal Rule of Evidence 702 (on expert testimony) is flexible and that expert testimony may properly include opinions on facts at issue in a case (Hart & Secunda, 2009). While the expert should not opine about discriminatory intent in a specific decision, “What the expert can do is offer his or her knowledge of the social science research and identify the characteristics of policies challenged in the particular workplace that research has linked with higher likelihood of bias and stereotype or lower likelihood of correction for bias” (p. 45).

While virtually all of these scholars (including Monahan, Walker, & Mitchell, 2008) agree that general-causation principles based on quality peer-reviewed scientific studies provide a useful framework for understanding specific case facts at trial, that is, for understanding how subjective discretion might facilitate biased performance appraisals in an organizational setting, there is considerable disagreement about how social science should be applied to the facts of a specific case, especially in legal contexts. Mitchell et al. (2011a,b) suggest that there are two ways to make this linkage: (1) relying on what they view as the flawed, personalistic judgment of experts, or (2) whenever possible, conducting case-specific research, grounded in a similar organizational context, to produce case-specific evidence: “If general social science principles are to be linked to a specific case by a social science expert, then those linkage opinions need to be based on a social fact study and not on the naked claim of ‘expert judgment’ that is the equivalent of *ipse dixit*” (p. 9). Mitchell and colleagues advocate the latter approach to linkage, claiming many ways to conduct empirical research in organizational settings to yield scientifically reliable case-specific opinions.

Moreover, they argue that organizations embroiled in discrimination litigation (or any form of litigation) should not be deterred from conducting case-specific empirical research because the barriers associated with ongoing litigation are not insurmountable, particularly in civil cases not involving discrimination claims. The obvious threats to internal validity, they argue, that are associated with conducting social science studies in an organizational setting where employees and managers alike are aware of ongoing litigation (e.g., reactivity of responding, ethical issues, use of in-house company records) are tempered by what they claim would be the greater external validity for the case at hand compared to simply applying aggregate-level, general-causation social science principles to understand the dynamics of the specific organizational context.

However, the difficulty of taking the necessary methodological and measurement steps to counter the numerous threats to internal validity in a litigation context should not be underestimated. As Mitchell et al. acknowledge, several important considerations (from choice of independent variables to selection of outcome measures to sample composition and choice of research setting) have the potential to undermine the fit between the case-specific study design and the features of the organizational setting one strives to model.

Thus, linkage between general-causation principles, grounded in quality science, and case-specific application, while desirable, may not be as feasible to achieve as Mitchell et al. maintain. The more common problem occurs when organizationally-specific research cannot be conducted or research access is denied by a company being sued for employment discrimination. Under those circumstances, quality, peer-reviewed science that is primarily experimental science or a mix of experimental science and research grounded in organizational contexts is a useful scientific foundation for generating scientifically-reliable inferences about the case-specific issues involved in a given context.

5. Parting advice

Principles must be translated for practical use. Even the best intended practitioners too-often rely on personal experience or a famous exemplar for guidance, rather than evidence-based management (Rousseau, 2006). The problem with one person’s experience is that it amounts to an unsubstantiated belief. “Everyone is entitled to his own opinion, but not his own facts” (attributed to Moynihan). What allows people to adopt an evidence-based approach?

Many professions have adopted evidence-based principles (Rousseau, 2006): policing, teaching, and medicine. Management may differ from these professions because managers expect more autonomy, experience more feedback lag, are not scientists, operate in the private sector, and do not live in a research context. An evidence-based Zeitgeist requires specific catalysts: (a) cause–effect principles, isolating moderator variables, (b) a receptive culture of social networks and organizational context that includes collaborative research activity, (c) structures for decision-making, and (d) access to a web of information and institutional support. Facts need both local knowledge (evidence) and general principles (Evidence). In effect, the established scientific consensus needs to be grounded in organizationally

relevant contexts, but there also needs to be a managerial recognition of the usefulness of relevant peer-reviewed science.

References

- American Sociological Association and the Law and Society Association Brief of amici curiae, in support of respondents, in the U.S. Supreme Court, *Wal-Mart Stores, Inc. v. Betty Dukes, et al.*, No. 10-277.
- Anderson, C. A., & Bushman, B. J. (1997). External validity of “trivial” experiments: The case of laboratory aggression. *Review of General Psychology*, 1(1), 19–41.
- Anderson, C. A., & Gentile, D. A. (2008). Media violence, aggression, and public policy. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 281–300). London: Blackwell.
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., et al. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173.
- Anderson, N., Ones, D. S., Sinangil, H. K., & Viswesvaran, C. (2001). *Handbook of industrial, work, and organizational psychology*. London: Sage.
- Aquino, K., & Thau, S. (2009). Workplace victimization: Aggression from the target’s perspective. *Annual Review of Psychology*, 60, 717–741.
- Argyris, C. (1974). *Theory in practice: Increasing professional effectiveness*. Washington: Jossey-Bass.
- Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method, and practice*. New York: Addison-Wesley.
- Bailar, J. C., & Patterson, K. (1985, March 7). Journal peer review: The need for a research agenda. *New England Journal of Medicine*, 312(10), 654–657.
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, 44(9), 1185–1193.
- Banks, R. R., Eberhardt, J. L., & Ross, L. (2008). Race, crime, and antidiscrimination. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 3–22). London: Blackwell.
- Bargh, J. A., & McKenna, K. Y. A. (2004). The Internet and social life. *Annual Review of Psychology*, 55, 573–590.
- Barling, J., Dupré, K. E., & Kelloway, E. K. (2009). Predicting workplace aggression and violence. *Annual Review of Psychology*, 60, 671–692.
- Bartunek, J. M., & Rynes, S. L. (2010). The construction and contributions of “Implications for Practice”: What’s in them and what might they offer? *Academy of Management Learning and Education*, 9(1), 100–117.
- Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science*, XX(X), 1–2.
- Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning & Education*, 3(2), 198–216.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407–425.
- Benard, S., Paik, I., & Correll, S. J. (2008). Cognitive bias and the motherhood penalty. *Hastings Law Journal*, 59, 1359–1387.
- Blair-Loy, M. (2003). *Competing devotions: Career and family among female executives*. Cambridge: Harvard University Press.
- Borgida, E., & Fiske, S. T. (Eds.). (2008). *Psychological science in the courtroom: Beyond common sense*. London: Blackwell.
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106, 265–289.
- Brief, A. P., & Dukerich, J. M. (1991). Theory in organizational behavior: Can it be useful? *Research in Organizational Behavior*, 13, 327–352.
- Brief, A. P., & Weiss, H. M. (2002). Organizational behavior: Affect in the workplace. *Annual Review of Psychology*, 53(1), 279–307.
- Brown v. Entertainment Merchants Ass’n*, 131 S.Ct. 2729 (2011).
- Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, 66(2), 204–225.
- Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. *American Psychologist*, 56(6–7), 477–489.
- Chubin, D. E., & Hackett, E. J. (1990). *Peerless science: Peer review and U.S. science policy*. Albany: SUNY Press.
- Cialdini, R. B. (1980). Full-cycle social psychology. In Bickman, L. (Ed.), *Applied social psychology annual*. Vol. 1 (pp.21–47). Beverly Hills, CA: Sage Publications.
- Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 86(2), 278–321.
- Cohen-Charash, Y., & Spector, P. E. (2002). The role of justice in organizations: A meta-analysis: Erratum. *Organizational Behavior and Human Decision Processes*, 89(2), 1215.
- Committee to Review the Scientific Evidence on the Polygraph. (2003). *The polygraph and lie detection*. Washington, DC: National Academies Press.
- Copus, D. (2005). Avoiding junk science: A lawyer’s view. In F. J. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 450–462). San Francisco: Jossey-Bass.
- Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, 112, 1297–1338.
- Crosby, F. J., & Dovidio, J. (2008). Discrimination in America and legal strategies for reducing it. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 3–22). London: Blackwell.
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn’t cut the ice. *Journal of Social Issues*, 60, 701–718.
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92, 631–648.

- Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J-Ph., et al. (2009). Is the stereotype content model culture-bound? A cross-cultural comparison reveals systematic similarities and differences. *British Journal of Social Psychology*, 48, 1–33.
- Cuddy, A. J. C., & Glick, P. (2011). Causes and consequences of warmth and competence judgments in organizations. *Research in Organizational Behavior*, doi:10.1016/j.riob.2011.10.004.
- Cumming, A. (2007, December 10). Polygraph Use by the Department of Energy: Issues for Congress. Congressional report. Library of Congress. Washington, DC: Congressional Research Service. Handle/proxy. Url: <http://handle.dtic.mil/100.2/ADA480718>, Vol. 20, p. 53.
- Cutler, B. L., & Kovera, M. B. (2011). Expert psychological testimony. *Current Directions in Psychological Science*, 20(1), 53–57.
- Diamond, S. S. (2008). Psychological contributions to evaluating witness testimony. In E. Borgida & S. T. Fiske (Eds.), *Beyond common sense: Psychological science in the courtroom*. Oxford: Blackwell Publishing.
- Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (Eds.). (2010). *The Sage Handbook of prejudice, stereotyping and discrimination*. Los Angeles, CA: Sage.
- Durante, F., Fiske, S. T., Cuddy, A. J. C., Kervyn, N., Akande, A., Adetoun, B. E., et al. (submitted for publication). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap.
- Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Boston: Harvard Business School Press.
- Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 100(3), 283–308.
- Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, 129(4), 569–591.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, 108(2), 233–256.
- Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, 60(5), 685–710.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598.
- Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, 117(1), 125–145.
- Eagly, A. H., Karau, S. J., Miner, J. B., & Johnson, B. T. (1994). Gender and motivation to manage in hierarchic organizations: A meta-analysis. *The Leadership Quarterly*, 5(2), 135–159.
- Eagly, A. H., & Koenig, A. M. (2008). Gender prejudice: On the risks of occupying incongruent roles. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 63–82). London: Blackwell.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111(1), 3–22.
- Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 100(3), 309–330.
- Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin*, 17(3), 306–315.
- Eckes, T. (2002). Paternalistic and Envious Gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles*, 47, 99–114.
- Eisenhart, M. (2002). The paradox of peer review: Admitting too much or allowing too little? *Research in Science Education*, 32, 241–255.
- Faigman, D. L. (2010). Evidentiary incommensurability: A preliminary exploration of the problem of reasoning from general scientific data to individualized legal decision-making. *Brooklyn Law Review*, 75(4), 1–22.
- Faigman, D. L., Dasgupta, N., & Ridgeway, C. L. (2008). A matter of fit: The law of discrimination and the science of implicit bias. *Hastings Law Journal*, 59, 1389–1434.
- Fiske, D. W., & Fogg, L. F. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist*, 45(5), 591–598.
- Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48, 621–628.
- Fiske, S. T. (2011). *Envy up scorn down: How status divides us*. New York: Russell Sage Foundation.
- Fiske, S. T. (in press). Managing ambivalent prejudices: The smart-but-cold, and the warm-but-dumb stereotypes. *Annals of the American Academy of Political and Social Sciences*.
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (1991). Social science research on trial: The use of sex stereotyping research in *Price Waterhouse v. Hopkins*. *American Psychologist*, 46, 1049–1060.
- Fiske, S. T., & Borgida, E. (2008). Providing expert knowledge in an adversarial context: Social cognitive science in employment discrimination cases. *Annual Review of Law and Social Science*, 4, 123–148.
- Fiske, S. T., & Borgida, E. (2011). Standards for using social psychological evidence in employment discrimination cases. *Temple Law Review*, 83(4), 101–111.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social perception: Warmth and competence. *Trends in Cognitive Science*, 11, 77–83.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82, 878–902.
- Fiske, S. T., & Taylor, S. E. (2008). *Social cognition: From brains to culture*. New York: McGraw-Hill.
- Fitzgerald, L. F., & Collingsworth, L. L. (2008). (Un)common knowledge: The legal viability of sexual harassment research. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 103–126). London: Blackwell.
- Fogg, L., & Fiske, D. W. (1993). Foretelling the judgments of reviewers and editors. *American Psychologist*, 48(3), 293–294.
- Gilovich, T. D., & Griffin, D. W. (2010). Judgment and decision-making. In Fiske, S. T., Gilbert, D. T., & Lindzey, G. Eds. *The handbook of social psychology*. Vol. 1 (pp.542–588). New York: Wiley.

- Guest, D. E. (2007). Don't shoot the messenger: A wake-up call for academics. *Academy of Management Journal*, 50, 1020–1026.
- Harris, L. T., McClure, S. M., van den Bos, W., Cohen, J. D., & Fiske, S. T. (2007). Regions of MPFC differentially tuned to social and non-social affective evaluation. *Cognitive and Behavioral Neuroscience*, 7, 309–316.
- Hart, M., & Secunda, P. M. (2009). A matter of context: Social framework evidence in employment discrimination class actions. *Fordham Law Review*, 78, 37–70.
- Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist*, 42, 443–455.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*, 5, 269–298.
- Heilman, M. E., & Haynes, M. C. (2008). Subjectivity in the appraisal process: A facilitator of gender bias in work settings. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 127–156). London: Blackwell.
- Heilman, M. E., & Okimoto, T. G. (2008). Motherhood: A potential source of bias in employment decisions. *Journal of Applied Psychology*, 93(1), 189–198.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). The weirdest people in the world? *Behavioral and brain sciences*, 33, 61–135.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). *Nature*, 466, 29.
- Henry, P. J. (2010). Institutional bias. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The Sage Handbook of prejudice, stereotyping, and discrimination* (pp. 426–440). Newburg Park, CA: Sage.
- Hodgkinson, G. P., & Healey, M. P. (2008). Cognition in organizations. *Annual Review of Psychology*, 59, 387–417.
- Hoobler, J. M., Wayne, S. J., & Lemmon, G. (2009). Bosses' perceptions of family-work conflict and women's promotability: Glass ceiling effects. *Academy of Management Journal*, 52(5), 939–957.
- Huesmann, L. R. (2010). Nailing the coffin shut on doubts that violent video games stimulate aggression: Comment on Anderson et al. *Psychological Bulletin*, 136(2), 179–181.
- Iacono, W. G. (2008). Polygraph testing. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 219–236). London: Blackwell.
- Iacono, W. G., & Lykken, D. T. (2005). The case against polygraph tests. In Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. Eds. *Modern scientific evidence: The law and science of expert testimony*. Vol. 4 (pp.605–655). Eagan MN: West.
- Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., et al. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, 29, 39–69.
- Kahan, D. M., Hoffman, D. A., & Braman, D. (2009). Whose eyes are you going to believe? *Scott v. Harris* and the perils of cognitive illiberalism. *Harvard Law Review*, 122(3), 837–906.
- Kahneman, D. (2007). Daniel Kahneman. In L. Gardner & W. M. Runyan (Eds.), *A history of psychology in autobiography* (pp. 155–197). Washington, DC: American Psychological Association.
- Kang, S. K., Inzlicht, M., & Derks, B. (2010). Social neuroscience and public policy on intergroup relations: A Hegelian analysis. *Journal of Social Issues*, 66(3), 585–601.
- Kassin, S. M. (2008). Expert testimony on the psychology of confessions: A pyramidal framework of the relevant science. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 195–218). London: Blackwell.
- Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior*, 34, 3–38.
- Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The “general acceptance” of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*, 44, 1089–1098.
- Kervyn, N., Fiske, S. T., & Malone, C. (in press). Do we perceive brands like we perceive people? Applying the Stereotype Content Model to brand perception. *Journal of Consumer Psychology*.
- King, E. B. (2008). The effect of bias on the advancement of working mothers: Disentangling legitimate concerns from inaccurate stereotypes as predictors of advancement in academe. *Human Relations*, 61(12), 1677–1711.
- King, E. B., Dunleavy, D. G., Jaffer, S., Morgan, W. B., Elder, K., & Graebner, R. (2011). Discrimination in the 21st century: Are science and the law aligned? *Psychology, Public Policy, and Law*, 17(1), 54–75.
- Klenowski, P. M., Bell, K. J., & Dodson, K. D. (2010). An empirical evaluation of juvenile awareness programs in the United States: Can juveniles be “Scared Straight?”. *Journal of Offender Rehabilitation*, 49, 254–272.
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642.
- Kovera, M. B., & Borgida, E. (2010). Social psychology and the law. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 1343–1385). New York: Wiley.
- Kovera, M. B., & McAuliff, B. D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: Are judges effective gatekeepers? *Journal of Applied Psychology*, 85, 574–586.
- Krieger, L. H., & Fiske, S. T. (2006). Behavioral realism in employment discrimination law: Implicit bias and disparate treatment. *California Law Review*, 94, 997–1062.
- Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53, 636–647.
- Landy, F. J. (2008). The tenuous bridge between research and reality: The importance of research design in inferences regarding work behavior. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 341–352). London: Blackwell.
- Lewin, K., & Grabbe, P. (1945). Conduct, knowledge, and the acceptance of new values. *Journal of Social Issues*, 1(3), 53–64.
- Lipsey, M.W. (1974). Research and relevance: A survey of graduate students and faculty in psychology. *American Psychologist*, 29(7), 541–553.

- Locke, E. A. (1986). Generalizing from laboratory to field: Ecological validity or abstraction of essential elements? In E. A. Locke (Ed.), *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management* (pp. 3–10). Lexington, MA: Lexington Books.
- Locke, E. A. (2009). Editor's Introduction. In E. A. Locke (Ed.), *Handbook of principles of organizational behavior: Indispensable knowledge for evidence-based management* (2nd ed., pp. xxv–xxx). New York: Wiley.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717.
- Loftus, E. F., Garry, M., & Hayne, H. (2008). Repressed and recovered memory. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 177–194). London: Blackwell.
- McAuliff, B. D., Kovera, M. B., & Nunez, G. (2009). Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? *Law and Human Behavior*, 33, 247–257.
- Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, 12, 269–275.
- Miller, C. C. (2006). From the editors: Peer review in the organizational and management sciences: Prevalence and effects of reviewer hostility, bias, and dissensus. *Academy of Management Journal*, 49(3), 425–431.
- Miner, J. B. (1984). The validity and usefulness of theories in an emerging organizational science. *Academy of Management Review*, 9, 306–326.
- Mitchell, G., Monahan, J., & Walker, L. (2011a, April). The ASA's missed opportunity to promote sound science in court. SSRN Electronic paper Collection: <http://ssrn.com/abstract=1816752>.
- Mitchell, G., Walker, L., & Monahan, J. (2011b). Beyond context: Social facts as case-specific evidence. *Emory Law Journal*, 60, 1109–1155.
- Monahan, J., Walker, L., & Mitchell, G. (2008). Contextual evidence of gender discrimination: The ascendance of “social frameworks”. *Virginia Law Review*, 94, 1715–1749.
- Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, 22(2), 103–122.
- Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, 115(2), 210–227.
- Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy Checklist scores in sexual violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior*, 32, 352–362.
- Nelson, T.D. (2008). The young science of prejudice against older adults: Established answers and open questions about ageism. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 45–62). London: Blackwell.
- Nelson, T. D. (Ed.). (2009). *Handbook of prejudice, stereotyping, and discrimination*. New York: Psychology Press.
- Nisbett, R. E., & Ross, L. (1980). *Human inferences: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87(6), 817–831.
- Olson, G. M., & Olson, J. S. (2003). Human-computer interaction: Psychological aspects of the human use of computing. *Annual Review of Psychology*, 54, 491–516.
- Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *The Academy of Management Perspectives*, 23(1), 6–16.
- Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, 112(3), 536–552.
- Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, 60, 339–367.
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.
- Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, 38(6), 922–934.
- Pfeffer, J., & Sutton, R. I. (2006, January 1–13). Evidence-based management. *Harvard Business Review: Decision Making*. OnPoint Article. Retrieved April 3, 2011, from www.hbr.org
- Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry*, 21, 77–99.
- Price Waterhouse v. Hopkins*, 490 U.S. 228 (1989).
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111(3), 781–799.
- Raelin, J. A. (2008). Refereeing the game of peer review. *Academy of Management Learning and Education*, 7(1), 124–129.
- Ridgeway, C. L. (2001). Social status and group structure. In M. A. Hogg & S. Tindale (Eds.), *Blackwell handbook of social psychology: Group processes* (pp. 352–375). Malden, MA: Blackwell.
- Ridgeway, C. L., & Correll, S. J. (2004). Motherhood as a status characteristic. *Journal of Social Issues*, 60, 683–700.
- Robinson v. Jacksonville Shipyards, Inc.* (M.D. Fla. 1989; Case No. 86-927).
- Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279–301.
- Rousseau, D. M. (2006). Is there such a thing as “evidence-based management”? *The Academy of Management Review*, 31(2), 256–269.
- Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology*, 31(6), 493–517.
- Rudman, L. A., Glick, P., & Phelan, J. E. (2008). From the laboratory to the bench: Gender stereotyping research in the courtroom. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 83–102). London: Blackwell.

- Rynes, S. L. (2006). "Getting on board" with AMJ: Balancing quality and innovation in the review process. *Academy of Management Journal*, 49(6), 1097–1102.
- Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, 50(5), 987–1008.
- Schacter, D. L., Dawes, R., Jacoby, L. L., Kahneman, D., Lempert, R., Roediger, H. L., et al. (2008). Policy forum: Studying eyewitness investigations in the field. *Law and Human Behavior*, 32, 3–5.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51(3), 515–530.
- Smith, A. N., Brief, A. P., & Colella, A. (2010). Bias in organizations. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The Sage Handbook of prejudice, stereotyping, and discrimination* (pp. 441–456). Newburg Park, CA: Sage.
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, 111(1), 42–61.
- Staw, B. M. (2010). The trouble with JDM: Some limitations to the influence of JDM on organizational research. *Industrial and Organizational Psychology*, 3, 411–416.
- Staw, B. M., & Epstein, L. D. (2000). What bandwagons bring: Effects of popular management techniques on corporate performance, reputation, and CEO pay. *Administrative Science Quarterly*, 45(3), 523–556.
- Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., et al. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, 62(4), 271–286.
- Tetlock. (2010, January 22). Deposition testimony in *EEOC v. Bloomberg L.P.* U.S. District Court, Southern District of New York, Case #07-CV-8383.
- Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior*, 29, 3–38.
- Tetlock, P. E., Mitchell, G., & Murray, T. L. (2008). The challenge of debiasing personnel decisions: Avoiding both under- and overcorrection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1(4), 439–443.
- Tropp, L. R., & Mallett, R. K. (2011). *Moving beyond prejudice reduction: Pathways to positive intergroup relations*. Washington, DC: American Psychological Association.
- Tropp, L. R., & Pettigrew, T. F. (2005). Differential relationships between intergroup contact and affective and cognitive dimensions of prejudice. *Personality and Social Psychology Bulletin*, 31(8), 1145–1158.
- Tsang, E. W. K., & Frey, B. S. (2007). The as-is journal review process: Let authors own their ideas. *Academy of Management Learning and Education*, 6(1), 128–136.
- Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, 49, 577–585.
- van den Bos, W., McClure, S. M., Harris, L. T., Fiske, S. T., & Cohen, J. D. (2007). Dissociating affective evaluation and social cognitive processes in ventral medial prefrontal cortex. *Cognitive and Behavioral Neuroscience*, 7, 337–346.
- van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, 58, 515–541.
- Vinchur, A. J., & Koppes, L. L. (2011). A historical survey of research and practice in industrial and organizational psychology. In Zedeck, S. (Ed.). *APA handbook of industrial and organizational psychology*. Vol. 1 (pp. 3–36). Washington, DC: American Psychological Association.
- Walsh, J. P., Tushman, M. L., Kimberly, J. R., Starbuck, B., & Ashford, S. (2007). On the relationship between research and practice: Debate and reflections. *Journal of Management Inquiry*, 16, 128–154.
- Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, 60, 53–85.
- Welch, J., & Byrne, J. A. (2001). *Jack: Straight from the gut*. New York: Warner.
- Weller, A. C. (2001). *Editorial peer review: Its strengths and weaknesses*. Medford, NJ: Information Today, American Society for Information Science and Technology.
- Wells, G. L., & Hasel, L. E. (2008). Eyewitness identification: Issues in common knowledge and generalization. In E. Borgida & S. T. Fiske (Eds.), *Psychological science in the courtroom: Beyond common sense* (pp. 159–176). London: Blackwell.
- Williams, J. C. (2010). *Reshaping the work-family debate: Why men and class matter*. Cambridge, MA: Harvard University Press.
- Wojciszke, B. (2005). Morality and competence in person and self-perception. *European Review of Social Psychology*, 16, 155–188.
- Zedeck, S. (2011). Introduction. In Zedeck, S. (Ed.). *APA handbook of industrial and organizational psychology*. Vol. 1 (pp. xxi–xxxiv). Washington, DC: American Psychological Association.