revised March, 2017
Medin, D.L., ojalehto, b., Marin, A., & Bang, M. (2017). Systems of (non-)diversity. *Nature Human Behaviour,* 1, 0088. doi:10.1038/s41562-017-0088

Douglas Medin, bethany ojalehto, Ananda Marin, Megan Bang

Systems of (Non) Diversity

Abstract. Intrinsic to the social, educational and behavioral sciences is the aim of addressing patterned variation in human thought and action across settings. Surprisingly, however, empirical work in these sciences continues to be limited by a lack of diversity in study populations, research methodology and the researchers themselves. This commentary analyzes these dimensions of diversity as they are situated in and affected by the larger organizational systems for publication, grants and academic advancement. This complex system appears to operate in a mutually reinforcing manner to discourage diversity. This analysis suggests that diversity goals central to our sciences will require systems level action rather a focus on any one component in isolation.

We take social, educational, and behavioral sciences as having the fundamental goal of identifying and understanding the range of human potential in forms of interaction with physical, biological, and social environments. Yet, almost paradoxically, current normative practices in these sciences appear to be directly antithetical to this aim. We consider three inter-related dimensions of diversity (or lack thereof) and suggest that they conspire to create something of a crisis for the science of human behavior. The three dimensions are 1. who gets studied (sample non-diversity), 2. the theory and methods used (methodological non-diversity), and 3. who directs and controls the research (researcher non-diversity). These facets of diversity operate within a larger system that regulates the publication process, research grants and standards for promotion and tenure, and this system can serve not only to reflect but also to amplify diversity

or nondiversity.

The central thesis of this commentary is that these various dimensions of diversity, including that of both the researchers and the researched, are inter-related and mutually reinforcing. That is to say, the same forces that lead to a lack of diversity in research methods and narrow sampling of research participants also tend to undermine researcher diversity. Conversely, encouraging and supporting researcher diversity works to encourage sample and methodological variability. In short, these factors work together to produce either strong diversity or extreme narrowness as two "attractor states."

First a few disclaimers. Disciplines that strongly value field studies or work in schools (that are themselves) becoming increasingly diverse will dominate psychology when it comes to study population diversity. At the same time behavioral economics and education are not immune to the use of convenience samples, and fields that attend to institutions may favor the powerful over the poor and may neglect phenomena of potential interest, precisely because of a bias towards institutionalized spaces. There is also considerable cross-discipline variation in researcher gender and ethnic diversity (here, economics is more of an extreme than psychology or education). We are less able to judge theoretical and methodological diversity in each discipline (we write as psychological and educational scientists), but if science reflects who does it, odds are that limitations are widespread. In summary, although some of the specific criticisms and examples we use may not apply with equal force across disciplines, lack of diversity can manifest in different ways with equally profound consequences.

It is also important to note that by researcher diversity we mean scholars whose socio-cultural history and experiences allow them to bring distinct perspectives to the research enterprise. This form of diversity almost surely is correlated with social class, race, gender and

ethnicity but it is not guaranteed to be, especially in fields or subfields with strong training models (and/or culturally-narrow epistemologies) that tend to emphasize a single perspective or in fields and subfields plagued by culturally-narrow epistemologies[1].

Right now, the social sciences in general and psychology in particular seem to be trapped on the narrow side of diversity, in direct contradiction to the overarching goal of seeking to understand patterned variations in human cognition and behavior. The solution—increasing the diversity of participants, researchers and methods—is no easy task, in part because the entire complex system of conducting, supporting, and reporting research has evolved to encourage and reinforce norms that are often narrow. The rest of this commentary elaborates on this argument.

Sample Non-Diversity. Extreme biased sampling of research participants and the neglect of their cultural context are increasingly recognized as threats to the generalizability of much of what we know about human thought and behavior. For example, by some estimates, well over 90% of the participants in published psychology studies are from what Henrich, Heine and Norenzayan [2] have called WEIRD samples (Western, Educated, Industrialized, Rich, Democratic). They reviewed evidence suggesting that, true to the acronym, these participants are particularly unrepresentative of the world at large. Henrich et al. also suggested that sampling bias reflected the WEIRDness of the people conducting this research (see also[3]). This paper already has been cited thousands of times, its implications acknowledged, but there is little if any evidence that research sampling practices and associated researcher training has changed in response to this acknowledgement (not to mention earlier critiques which have been met with the same apparent indifference).

Why has change been so elusive? Practical reasons of time and personnel costs associated with research have led to convenience sampling of participants as a default, dominant practice.

Even within this extremely narrow sampling, samples appear to matter. On the rare occasions where major fields of undergraduate study have been examined in psychological studies, the results suggest that major affects cognition (e.g.,[4-7].) Overall, maximizing nondiversity may not be the explicit goal, but current practices often have that consequence.

This practice is harmful in a number of ways, most centrally in treating middle-class European American experiences as "the" norm, especially for purposes of theory building. This default makes deviations from this standard "abnormal" and findings from "abnormal" groups of questionable generality. When data are reported from non-WEIRD samples, there is a burden to compare the results to WEIRD populations and explain any differences. In fact, however, any such burden should be fully symmetrical.

The increasing use of internet studies (e.g., with Amazon's Mechanical Turk) is a potential palliative for sample diversity, but again, internet studies have been conducted primarily for convenience and only rarely as a tool to explore sample variation. Narrow sampling in medical research is transparently immoral and inexcusable--is psychological science any different? [We hasten to add that there is a long, deplorable history of medical studies where under-represented groups were included in research without their informed consent. Although substantial progress has been made in eliminating blatantly unethical studies, there are residual issues of deficit orientations and research interpretation that might be best addressed with minority scientists at the table.]

The Case for Sample Diversity. Theories of cognition and behavior may not just be limited by limited samples, they may be inaccurate even for that selfsame sample. There is increasing evidence of the fundamental role of culture and patterns of experience in human learning and development (for a review see[8].) For example, a dominant perspective on cognitive

development is that it is best analyzed in terms of at least three distinct domains[9,10] commonly defined in terms of the ontological kinds they treat:[11] intentional agents are the proper subjects of *folk psychology*; nonhuman animals and plants fall under *folk biology*; and natural inanimates under *naïve or folk physics*. It has been argued that young children's biology is initially organized in terms of a naïve psychology where human beings are the prototype[12], but more recently research supports the view that this trajectory varies across culture [13,14] and that a human-centered biology is a learned cultural model[15].

Recently there have been signs that this partitioning of domains (folk physics, folk biology, folk psychology) may itself constitute a particularly Western way of parsing knowledge that fails to capture central features of knowledge organization in other cultural settings[8]. If so, this is doubly problematic because what are often taken as natural domains may affect both what research is done and how it is conducted. For example, Atran and Medin [16] report a body of work on culture and folkbiology, focused on people's understandings of plants and animals. Note, however, that if the initial research framing had been in terms of ecosystems (folk ecology), the researchers likely would have included natural inanimates such as the sun and moon, rocks, soil, and water in their probes. They did not. The category, folkbiology, may have biased the focus away from natural inanimates and consequently may have limited the researchers' understanding of how knowledge varies within and across communities. Atran and Medin did include human beings in their studies of plant-animal interactions, a species one might overlook from the cultural perspective that humans are not part of nature [1,17].

As another example consider the inter-disciplinary field of decisionmaking. Researchers have identified a number of biases such as attending to relative rather than absolute saving (e.g.

Medin, D.L., ojalehto, b., Marin, A., & Bang, M. (2017). Systems of (non-)diversity. *Nature Human Behaviour,* 1, 0088. doi:10.1038/s41562-017-0088
being willing to make an extended trip to save $20 on a $40 purchase but unwilling to do so for that same $20 savings on a $120 purchase). That and related context effects are quite robust, but more recent research shows that these effects do not extend to poor people (or those placed under artificial conditions of scarcity) who behave "more rationally." [18,19]

Another important domain where study sample diversity has refined (or upended) prior understanding is that of human sensory experience. Recent interdisciplinary research shows that diverse languages facilitate differing modes and degrees of sensory discrimination (see special issue, introduced by Majid and Levinson [20]). For example, it was long thought that humans (represented by speakers of English and closely related languages) are poor at discriminating odors, leading researchers to conclude that odor was not a coherent semantic domain or even to propose that smell representations were inaccessible to language centers of the brain. Yet recent work with non-Western languages undermines this generalization [21,22] Compared with English, for instance, the Maniq and Jahai languages have many olfactory names, and Jahai speakers are substantially better than English speakers at discriminating unfamiliar odors. Contrary to prior received wisdom based on Western samples, olfactory abstraction is possible if you speak the right language.

The Henrich et al. review[2] made important points about generalizability. Probing for generalizability is one very useful perspective, but searching for systematic, patterned variation with other samples is also intrinsic to good science and may have the additional virtue of steering researchers away from the questionable practice of treating Western samples as the default or standard by which other samples are evaluated. One quite literal example of "one size fits all" concerns the universal anthropometric standards for assessing obesity and under-nutrition which focus on body mass index (BMI) or weight-for-height z-scores (WHZ). As noted by Hruschka

and Hadley [23], these metrics fail to take into account human cultural and biological variation in body shapes, which can be quite large. Hruschka and Hadley estimate that up to 500 million obese people around the world could be misclassified if these measures are not adjusted to use population-sensitive cutoffs. For some of the other negative consequences of this Western, middle class default (such as deficit models) see [24].

The Case for Researcher Diversity. Researcher diversity is typically treated within a framework that focuses on fairness and representational equity. In the United States, numerous reports by the National Research Council (e.g., [25]), have noted that minority scholars are under-represented in the sciences and have made constructive recommendations for addressing this gap. These recommendations have focused on equal opportunity but do not connect researcher diversity – and diverse ways of knowing, attending and acting– with science itself. In other words, the call has been to broaden participation rather than to broaden conceptions of science and scientific methods.

Before defending the claim that researcher diversity makes for better science, we need to clarify that we do not subscribe to a "box model" of diversity in which gender or ethnicity are essentialized or reduced to a list of internal traits. Instead, we focus on the extent to which diversity of life practices, perspectives, values, and motivations are correlated with these groupings[26].

Validity in the sciences involves choices about what problems to study, what populations to study, and what procedures and measures should be used. In making these choices, diverse perspectives and values are important. Consider the strong correlation between social-science researchers and the people they study. This mostly White middle-class group of scientists

focuses their research programs primarily on White, middle-class populations. We suspect that this focus is driven by convenience (such as undergraduates from introductory psychology courses or organizations and institutions such as museums, where middle class samples are very-well represented) rather than purposeful neglect of other potential samples. Although undergraduate samples increasingly include scholars of color, their numbers remain small enough that, even when they are described in method sections of papers, they are not analyzed for potential differences. And unless researchers are specifically interested in gender differences, a breakdown of results by gender is rare. It is as if these researchers assume that they are studying "what people do" and any conveniently available people will do.

Diverse researcher perspectives often are associated with diverse research foci and the generation of new findings. For example, when female scientists began to study primate social behavior, new insights into both female and male behaviors were uncovered [27]. Similarly, in psychological science, minority scholars (and culturally oriented majority scholars) have expanded previously accepted pathways of identity development, motivation, and resilience [28,29]. For example, theories arguing for the primacy of an internal locus of control have been contested, pointing to the efficacy of an external locus of control when populations face stigmatization and forms of racism that they do not control [30].

Yet another example comes for developmental research on moral cognition. Work in the United States by Lawrence Kohlberg (e.g. [31,32],) modeled after Piaget, identified distinct stages in moral development by using moral dilemmas and analyzing child and adult reasoning underlying judgments. This stage theory represented a strong metric but was soon challenged by Carol Gilligan [33] who argued Kohlberg's model was male-oriented and that, in contrast to men, women

attended more to issues of relationships and care and less to abstract principles of justice or equality. Kohlberg's model was further undermined in the form of a cultural critique by Snarey [34] who reviewed the application of Kohlberg's stage theory in 27 cultures and found that the suitability of the model varied substantially across cultures. Snarey also cited work by non-US scholars in India identifying the unity of life and the human relationships with plants and animals as an important moral principle that was outside the scope of Kohlberg's (and Gilligan's) theory. Note how in these cases sample diversity and theoretical diversity follow on the footsteps of researcher diversity.

Researcher diversity (see [35] for arguments on the importance of political diversity) may also carry with it new approaches to the task of research itself. One example is evidenced by the slow scholarship movement, which argues that quality scholarship is informed by a holistic engagement with work that requires longer timescales than the standard publishing race would admit.[36] (It is telling that this Mountz et al. article was the collaborative endeavor of 11 authors—all of them women.) Slow theory offers an expanded definition of scholarly work that includes care, community building, and advocacy as well as productive critique of theoretical orthodoxy. These values directly align with the research practices required for sample and methodological diversity outlined above, and may also enable novel perspectives. In the field of pedagogy for instance, an experiential environmental/outdoor education program has successfully mobilized slow theory principles in order to provide experiential, place-based learning as an alternative to the acceleration of speedy pedagogy in higher education [37].

While modes of engagement with research may not seem directly relevant to questions of substantive diversity in psychological science, it is an open question to what extent this attitude reflects a White middle-class male intuition that contributes to niche construction in social

science. These niches discourage potential researchers who do not share the dominant mode of engaging with research. Whether or not one finds the slow scholarship movement appealing, giving up convenience samples and adapting and moving beyond standard, convenient measures will necessarily lead to slower (and, we suggest, more thoughtful) research scholarship.

The Case for Methodological Diversity. Practices across the social sciences are cultural practices, and the norms that influence such practices emerge from across diverse sites, from diverse practitioners, addressing diverse problems. This includes variability of theory, materials, methods, measures, environments and study contexts. What we are glossing as methodological diversity can be understood as multi-layered and multi-faceted. The weak, but still critically important, stance would be to recognize that considerations of efficiency and convenience often push research in the direction of unhealthy narrowness (e,g, neglecting practices such as seeking converging measures). Stronger forms of methodological diversity include encouraging and supporting the different perspectives that researcher diversity might afford.

There have been periodic critiques of psychological science's lack of diversity in choice of research problems, methods, measures, settings and theoretical orientations; [38-41] see also [42], for an additional perspective). Coming from multiple areas and subfields of social science, these calls underscore the scope of the problem: for these very forms of variation that go underexplored are critical to understand human cognition in context. Developing contextualized theories requires culturally-informed measures that account for the unique histories and demographics of groups (see [43].) This goal, in turn, necessitates building relationships with groups and communities in a way that takes time and includes power-sharing.

Multiple generations of social scientists have been brought up on Campbell and Stanley's "Experimental and quasi-experimental designs for research," first published in 1963. Among its nuggets of wisdom is the discussion of the importance of convergent (and sometimes divergent) measures. It is ironic that the current so-called "replicability crisis" favors exact replications over the kind of robustness that can only be reinforced by converging measures. And perhaps it is doubly ironic that the replication project's involvement of many labs has led to increased diversity in the (college and university) sites for research and the perhaps grudging acceptance that effects implicitly assumed to hold across sites may be contextual and cultural [44].

The advent of the personal computer dramatically increased the range of materials and stimulus presentation methods that could be employed in research and the ease with which this could be done, but it also has led to greater methodological conformity, with a participant interacting on a computer in a cubicle as the modal "social context." This is an extremely narrow sample of contexts and situations, excluding the outdoors, other people, plants, animals, food, noncomputer artifacts and so on. Baumeister, Vohs, and Funder [45] summarized this concern for at least one of the social sciences by noting that psychology is increasingly limited to self-reports and finger movements (on computer keyboards). Some might argue that the shift to computer-based contexts is not so dissimilar from contexts where a single researcher is sitting across from a single participant in a university lab. Considering that for much of human history, interactions have occurred in intergenerational contexts where land and mobility play a central role in activity, both of these research modalities (the use of computer-based stimuli and dyadic research interactions) could be described as weird or strange.

A positive example of employing technology to assess external validity is a study by Hofmann, et al. [46] They used smartphones to text participants randomly five times a day

(between 9 a.m. and 9 p.m.) to probe whether they had witnessed, heard, read about, or experienced a moral or immoral incident within the last hour. They found evidence for moral self-licensing (engaging in a moral act made a subsequent immoral act more likely and a subsequent moral act less likely), consistent with laboratory studies and reinforcing their robustness. Hofmann et al. also found that being the recipient of a moral act increased the likelihood of engaging in a later moral act. Just a few years ago this type of study would have been either technologically impossible or prohibitively expensive to conduct (see also [47].) We need more such positive examples combined with distinct samples of research participants.

The inter-disciplinary field of decision making provides other good examples of both breadth and narrowness in methods and procedures. The prototypical decision task involves participants making choices between bets varying in probabilities and payoffs. Baruch Fischhoff [48] has criticized this kind of work (which he himself has substantially contributed to) as risking "implosion" as laboratory simplifications of real world situations take on a life of their own and exclude any aspects of decision making that cannot or do not conform to those task constraints. A salient example of this practice is when participants are cautioned only to use the information provided in the scenarios and nothing more (such as coming up with new options). These strong closed world assumptions may prevent relevant decision processes from being evidenced and may more or less guarantee that a model of the task will fit the data (see [49] for an extended critique).

Fischhoff also notes that mapping from narrow tasks back to real world situations requires considerable extrapolation and may miss other important phenomena. For example, his own studies of adolescent decision making show that a very common form of choice involves not selecting among options of the same kind, but rather whether or not to do something [48].

There's nothing in principle that would keep these types of situations from being brought into the lab, but if young scholars base their research on what has gone before, they may gravitate toward the safe precedent of tasks that pose bets varying in probability and amount.

In other cases generalizability of laboratory results to real world contexts is assumed and used to guide policy recommendations without a firm empirical basis for doing so. For example, there is considerable evidence that external (e.g., monetary) rewards can work to undermine intrinsic motivation (see original work by Lepper and Greene [50] and a recent review by Goswami and Urminsky [51]; see also [52], for a more review of incentive effects). These effects have been often replicated, but their boundary conditions not systematically probed. Therefore, it may not be surprising that a field study where students were given monetary incentives for taking and passing advanced placement courses resulted in positive and enduring effects [53]. Taking note of these findings, Goswami and Urminsky [51] have identified laboratory conditions that produce and remove undermining effects.

A Systems Level Perspective and Niche Construction. Given that the behavioral and social sciences constitute complex systems interacting with other complex systems, they almost surely have evolved in a way that is adaptive for practitioners. When the telling of this history includes a lack of diversity (e.g., when the field of Black psychology is disregarded), these sorts of niche construction processes work to reinforce narrow (White middle-class male) intuitions, narrow (White middle-class) values and narrow (White middle-class male) research practices. Each of these factors may be a source of discord for would-be researchers who do not fall into this privileged (including privileged by precedent) group. Privilege includes who gets to decide what research questions are important, who gets to become a research participant, which methods and measures are condoned, and what dissemination looks like.

This niche construction process may be well underway by the time youth enter college. College entrance exams are correlated with academic success but Sternberg [54,55] notes that they focus on analytic skills at the expense of creative skills and practical skills. When admissions exams include measures of these other skills, creative and practical skills also predict academic success and notably, using these measures reduces or eliminates the achievement gap for minorities (see also [56]).

The training of minority graduate students is another diagnostic example. When minority and female graduate students attending predominantly white institutions, recruited especially because of their diverse backgrounds, move to broaden the methodological frameworks they use in their own work, they at a minimum face challenges that are epistemological (questioning from faculty about the benefits of employing more diverse approaches and study populations), contextual (the cultural geography of the university campus and surrounding neighborhoods) and practical (time required to develop research relationships). Such challenges are diagnostic of the systems-level changes that are needed if our field is to uncover and understand the patterned variation so central to our enterprise and mobilize this knowledge to address central challenges facing humankind.

It will take commitment and energy to encourage and support study sample diversity, methodological diversity, and researcher diversity. Perhaps the quickest route to study population diversity and expanded methods, contexts and theoretical perspectives is through researcher diversity, supported by encouraging diverse research perspectives. When women and underrepresented minorities see their own orientations and practices recognized and supported as relevant to the practices of science, the field of science should seem much more attractive to them. And our sciences will be all the better for the perspectives that diverse scholars can bring

to them and in so doing accomplishing more equitable forms of knowledge production as well as more robust foundational knowledge.

References

1 Medin, D. L. & Bang, M. *Who's asking? Native Science, Western Science and Science Education*. (The MIT Press, 2014).
2 Henrich, J., Heine, S. J. & Norenzayan, A. Most people are not WEIRD. *Nature* **466**, 29 (2010).
3 Arnett, J. J. The Neglected 95%, a Challenge to Psychology's Philosophy of Science. *Am Psychol***64**, 571-574, doi:10.1037/a0016723 (2008).
4 Lehman, D. R. & Nisbett, R. E. A longitudinal study of the effects of undergraduate training on reasoning. *Devl Psychol* **26**, 952 (1990).
5 Frank, R. H., Gilovich, T. D. & Regan, D. T. Do economists make bad citizens? *J Econ Perspec* **10**, 187-192 (1996).
6 Frank, R. H., Gilovich, T. & Regan, D. T. Does studying economics inhibit cooperation? *J Econ Perspec* **7**, 159-171 (1993).
7 Wang, L., Malhotra, D. & Murnighan, J. K. Economics education and greed. *Acad Manag Learn Educ* **10**, 643-660 (2011).
8 ojalehto, b. & Medin, D. in *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (eds R. Scott & S. Kosslyn) (John Wiley & Sons, 2015).
9 Carey, S. *The origin of concepts*. (Oxford University Press, USA, 2009).
10 Gelman, S. A. & Legare, C. H. Concepts and folk theories. *Annu Rev Anthropol* **40**, 379 (2011).
11 Shweder, R. A. *et al.* Likeness and Likelihood in Everyday Thought: Magical Thinking in Judgments About Personality [and Comments and Reply]. *Curr Anthropol*, 637-658 (1977).
12 Carey, S. *Conceptual change in childhood*. (The MIT Press, 1985).
13 Medin, D. L., Waxman, S. R., Woodring, J. & Washinawatok, K. Human-centeredness is not a universal feature of young children's reasoning: Culture and experience matter when reasoning about biological entities. *Cognit Dev* **25**, 197-207 (2010).
14 Waxman, S., Medin, D. & Ross, N. Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Dev Psychol***43**, 294-308 (2007).
15 Herrmann, P., Waxman, S. R. & Medin, D. L. Anthropocentrism is not the first step in children's reasoning about the natural world. *Proc Natl Acad Sci Unit States Am* **107**, 9979-9984 (2010).
16 Atran, S. & Medin, D. *The Native Mind and the Cultural Construction of Nature*. (MIT Press, 2008).
17 Bang, M., Medin, D. L. & Atran, S. Cultural mosaics and mental models of nature. *Proc Natl Acad Sci Unit States Am* **104**, 13868-13874 (2007).

18 Mullainathan, S. & Shafir, E. *Scarcity: Why having too little means so much*. (Henry Holt and Company, LLC, 2013).
19 Shah, A. K., Shafir, E. & Mullainathan, S. Scarcity frames value. *Psychol Sci* **26**, 402-412 (2015).
20 Majid, A. & Levinson, S. C. The senses in language and culture. *The Senses and Society* **6**, 5-18 (2011).
21 Wnuk, E. & Majid, A. Revisiting the limits of language: The odor lexicon of Maniq. *Cognit* **131**, 125-138 (2014).
22 Majid, A. & Burenhult, N. Odors are expressible in language, as long as you speak the right language. *Cognit* **130**, 266-270 (2014).
23 Hruschka, D. & Hadley, C. How much do universal anthropometric standards bias the global monitoring of obesity and undernutrition? *Obes Rev* **17**, 1030-1039 (2016).
24 Medin, D. L., Bennis, W. & Chandler, M. Culture and the home-field disadvantage. *Perspect Psychol Sci* **5**, 708-713 (2010).
25 National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. (Washington, DC, 2011).
26 Gutiérrez, K. D. & Rogoff, B. Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher* **32**, 19-25 (2003).
27 Small, M. F. & Hrdy, S. B. Secondary sex ratios by maternal rank, parity, and age in captive rhesus macaques (Macaca mulatta). *Am J Primatol***11**, 359-365 (1986).
28 Graham, S. & Hudley, C. Race and ethnicity in the study of motivation and competence. *Handbook of Competence and Motivation*, 392-413 (2005).
29 Spencer, S. *Race and ethnicity: culture, identity and representation*. (Routledge, 2006).
30 Crocker, J. & Major, B. Social stigma and self-esteem: The self-protective properties of stigma. *Psychol Rev***96**, 608 (1989).
31 Kohlberg, L. The claim to moral adequacy of a highest stage of moral judgment. *J Philos* **70**, 630-646 (1973).
32 Kohlberg, L. *The psychology of moral development: The nature and validity of moral stages*. (Harper & Row, 1984).
33 Gilligan, C. *In a different voice: Psychological theory and women's development*. (Harvard University Press, 1982).
34 Snarey, J. R. Cross-Cultural Universality of Social-Moral Development:: A Critical Review of Kohlbergian Research. *Psychol Bull* **97**, 202-232 (1985).
35 Duarte, J. L. *et al.* Political diversity will improve social psychological science. *Behavioral and Brain Sciences* **38**, 1-58 (2015).
36 Mountz, A. *et al.* For slow scholarship: A feminist politics of resistance through collective action in the neoliberal university. *ACME: An International Journal for Critical Geographies* **14**, 1235-1259 (2015).
37 Payne, P. G. & Wattchow, B. Phenomenological Deconstruction, Slow Pedagogy, and the Corporeal Turn in Wild Environmental/Outdoor Education. *Canadian Journal of Environmental Education* **14**, 15-32 (2009).
38 Fiedler, K. What Constitutes Strong Psychological Science? The (Neglected) Role of Diagnosticity and A Priori Theorizing. *Perspect Psychol Sci* **12**, 46-61 (2017).

39 Neisser, U. in *Practical Aspects of Memory* (eds M.M Gruneberg, P.F. Morris, & R.N. Sykes) 3-24 (Academic Press, 1978).
40 Rozin, P. Social psychology and science: Some lessons from Solomon Asch. *Pers Socl Psychol Rev* **5**, 2-14 (2001).
41 Rozin, P. What kind of empirical research should we publish, fund, and reward?: A different perspective. *Perspect Psychol Sci* **4**, 435-439 (2009).
42 Lloyd, G. E. R. *Cognitive variations: Reflections on the unity and diversity of the human mind*. (Oxford University Press, USA, 2007).
43 Yoshikawa, H., Mistry, R. & Wang, Y. Advancing methods in research on Asian American children and youth. *Child Dev* **87**, 1033-1050 (2016).
44 Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J. & Reinero, D. A. Contextual sensitivity in scientific reproducibility. *Proc Natl Acad Sci Unit States Am*, 6454-6459 (2016).
45 Baumeister, R. F., Vohs, K. D. & Funder, D. C. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspect Psychol Sci* **2**, 396-403 (2007).
46 Hofmann, W., Wisneski, D. C., Brandt, M. J. & Skitka, L. J. Morality in everyday life. *Science* **345**, 1340-1343 (2014).
47 Hofmann, W. & Patel, P. V. SurveySignal A Convenient Solution for Experience Sampling Research Using Participants' Own Smartphones. *Soc Sci Comput Rev* **33**, 235-253 (2015).
48 Fischhoff, B. The Real World: What Good Is It? *Organ Behav Hum Decis Process* **65** (1996).
49 Bennis, W. M., Medin, D. L. & Bartels, D. M. The costs and benefits of calculation and moral rules. *Perspect Psychol Sci* **5**, 187-202 (2010).
50 Lepper, M. & Greene, D. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology* **31**, 479-486 (1975).
51 Goswami, I. & Urminsky, O. The dynamic effect of incentives on postreward task engagement. *J ExpPsychol Gen* **146**, 1 (2017).
52 Gneezy, U., Meier, S. & Rey-Biel, P. When and why incentives (don't) work to modify behavior. *J Econ Perspect* **25**, 191-209 (2011).
53 Jackson, C. K. A little now for a lot later a look at a texas advanced placement incentive program. *J Hum Resour* **45**, 591-639 (2010).
54 Sternberg, R. J. Who are the bright children? The cultural context of being and acting intelligent. *Educational Researcher* **36**, 148-155 (2007).
55 Sternberg, R. J. & Collaborators, T. R. P. The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence* **34**, 321-350 (2006).
56 Sternberg, R. J., Bonney, C. R., Gabora, L. & Merrifield, M. WICS: A model for college and university admissions. *Educ Psychol* **47**, 30-41 (2012).