The persistent sampling bias in developmental psychology: A call to action

Mark Nielsen a,b,*, Daniel Haun c, Joscha Kärtner d, Cristine H. Legare e

a School of Psychology, University of Queensland, Brisbane, QLD 4072, Australia
b Faculty of Humanities, University of Johannesburg, Auckland Park 2006, South Africa
c Department of Early Child Development and Culture, University of Leipzig, and Leipzig Research Center for Early Child Development, D-04109 Leipzig, Germany
d Department of Psychology, University of Münster, D-48149 Münster, Germany
e Department of Psychology, The University of Texas at Austin, Austin, TX 78712, USA

Article history:
Received 17 January 2017
Revised 18 April 2017

Keywords:
WEIRD data
Cross-cultural research
Generalizable data
Representative data
Developmental science
Diversity
Cultural psychology
Developmental psychology

Abstract
Psychology must confront the bias in its broad literature toward the study of participants developing in environments unrepresentative of the vast majority of the world’s population. Here, we focus on the implications of addressing this challenge, highlight the need to address overreliance on a narrow participant pool, and emphasize the value and necessity of conducting research with diverse populations. We show that high-impact-factor developmental journals are heavily skewed toward publishing articles with data from WEIRD (Western, educated, industrialized, rich, and democratic) populations. Most critically, despite calls for change and supposed widespread awareness of this problem, there is a habitual dependence on convenience sampling and little evidence that the discipline is making any meaningful movement toward drawing from diverse samples. Failure to confront the possibility that culturally specific findings are being misattributed as universal traits has broad implications for the construction of scientifically defensible theories and for the reliable public dissemination of study findings.

* Corresponding author at: School of Psychology, University of Queensland, Brisbane, QLD 4072, Australia.
E-mail address: nielsen@psy.uq.edu.au (M. Nielsen).

http://dx.doi.org/10.1016/j.jecp.2017.04.017
0022-0965/© 2017 Elsevier Inc. All rights reserved.
Introduction

Growing attention has been drawn to the lack of diversity in psychological testing, in particular to the fact that the vast majority of psychological research has been conducted on populations that are unrepresentative of human culture more globally—those from WEIRD (Western, educated, industrial, rich, and democratic) backgrounds (Henrich, Heine, & Norenzayan, 2010; Legare & Harris, 2016; Nielsen & Haun, 2016). The dearth of systematic research outside of Western cultural contexts is a major impediment to theoretical progress in the psychological sciences (Legare & Nielsen, 2015; Rowley & Camacho, 2015). Where psychological researchers assume that data are not specific to the sample of participants under direct test (i.e., that findings are generalizable), lack of attention to cultural variation and its psychological consequences risks yielding incomplete, and potentially inaccurate, conclusions (e.g., Apicella & Barrett, 2016; Evans & Schamberg, 2009; Mani, Mullainathan, Shafir, & Zhao, 2013; Votruba-Drzal, Miller, & Coley, 2016).

A complete understanding of the ontogeny and phylogeny of the developing human mind depends on sampling diversity (Clegg & Legare, 2016; Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007; Machluf & Bjorklund, 2015; Nielsen, 2012; van Schaik & Burkart, 2011). Where research efforts are focused on identifying core mechanisms or universal aspects of psychology, failure to acknowledge the possible impact of environment on the behavior of participants must be considered at best neglectful and at worst bad science. Our aim here is to show, presenting new data, that the influence of culture—"a set of meanings or information that is non-genetically transmitted from one individual to another, which is more or less shared within a population (or a group) and endures for some generations" (Kashima & Gelfand, 2012, p. 640)—is not afforded sufficient attention in the developmental literature. Moreover, despite growing awareness of a need to change, we show that there is little shift in research practices that overly rely on data from a markedly narrow sample pool and little acknowledgment that this is potentially problematic for interpreting data and arising theoretical assumptions.

From the limited research that exists, there is clear evidence of substantial differences between Western educated industrialized communities and non-Western populations in fundamental aspects of child development (Bornstein, 1991; Corsaro, 1996; Gaskins & Paradise, 2010; Kruger & Tomasello, 1996; Lave & Wenger, 1991; LeVine, LeVine, Schnell-Anzola, Rowe, & Dexter, 2012; Miller & Goodnow, 1995; Rogoff, 2003). This includes evidence for cross-cultural variation in child socialization and how parents interact with their infants (Keller, 2007; Keller & Kärtner, 2013; Kärtner, 2015), the kinds of tasks parents engage their infants in (Cole, 1996; Lancy, Bock, & Gaskins, 2010), and the amount of time children spend with nonparental caregivers and peers (Gaskins, 2006; LeVine, 1980). For example, there is cultural variation in the degree to which mothers focus on face-to-face interaction and object play when engaging their infants that leads to culture-specific maternal reactions to infants’ communicative signals (Keller et al., 2004; Kärtner, 2015; Little, Carver, & Legare, 2016).

Cultural variation has also been documented in other fundamental aspects of human cognition (Wang, 2017). For example, Haun, Rapold, Call, Janzen, and Levinson (2006) found that Hai/om children tend to employ a geocentric search strategy (where the position of relevant items is maintained relative to the larger surrounding environment) to find something hidden among an array of overturned cups, in contrast to the egocentric approach (where the position of relevant items is maintained relative to the viewpoint of the children) adopted by Western children. Our aim here is not to dwell on culturally determined differences in children’s developmental environments but rather to draw attention to interpretation and the assumptions that would be made without consideration of potential cultural influences. If Haun and colleagues had tested only Western children, then it could have easily been assumed, as is typically written, that “children” employ egocentric search strategies. But “children” generally do not do so; only children from specific cultural backgrounds do.

Critically, there is no universal developmental context in which children grow up, nor is there a universal environment for the human mind. To understand psychological processes, thus, it is necessary to exercise caution when generalizing beyond the specific sociocultural context at hand (Clegg, Wen, & Legare, 2017). To reiterate, if an underlying goal of any research endeavor is to identify globally relevant patterns of development—and not patterns that are specific to one population in isolation—then failure to acknowledge the possible influence of cultural factors on participants’ responses is...
either neglectful or bad science (an error that each of us, as authors, has made). It is the equivalent of missing a confound and assuming that the data at hand are unaffected. For example, it would be ill-advised to interpret children’s responses to questions relating to folk biological reasoning without acknowledging that such answers are population dependent and vary with culturally determined interactions with the natural world (Medin, Waxman, Woodring, & Washinawatok, 2010; Proffitt, Coley, & Medin, 2000; Ross, Medin, Coley, & Atran, 2003). It would be similarly ill-advised to make claims about human perceptual attention processes based on data collected with only Western or Asian participants (Nisbett & Miyamoto, 2005). For developmental science to be sure it is built on solid foundations, thus, it is critical that culturally variant and invariant patterns of development are identified and that it is acknowledged when reported data could be different if collected in a distinct sample. The alternative—that a sample lacking in cultural diversity is representative of all children—should no longer be treated as an acceptable default option.

Evidence for a persistent bias

Arguments that there is an inherent bias in what constitutes our participant pools are of course not new (Bornstein, 2002; Cole, 1996; Levine, Martinez, Brase, & Sorenson, 1994; Rogoff, 1990; Scribner & Cole, 1973; Serpell, 1976; Shweder, 1990; Super & Harkness, 1986; Whiting & Whiting, 1975). Building on these earlier endeavors, Henrich et al. (2010), in a prominent and highly cited article, drew attention to the disproportionate representation in psychology of what they coined “WEIRD” participants (i.e., those from Western, educated, industrialized, rich, and democratic backgrounds). This extended previous work by Arnett (2008), who analyzed articles published from 2003 to 2007 in peak journals from six subdisciplines of psychology, revealing that 96% of the participants were from Western industrialized countries (68% from the United States alone) and that 99% of first authors were at universities in Western countries (73% from the United States alone). Putting this in the context of population size, 96% of psychological samples came from countries with only 12% of the world’s population, and this skew in sampling was apparent in Arnett’s analysis of the journal Developmental Psychology. Perhaps a wider assessment of developmental journals would have yielded a more representative picture.

To evaluate this possibility, we surveyed every article published between 2006 and 2010 in the journals Child Development, Developmental Psychology, and Developmental Science (consistently the highest ranked experimental developmental psychology journals by impact factor) and recorded the geographical region of the participants, whether they were human or non-human primates, and the affiliation of the first author. Participant information was gleaned from information on where data were collected provided in “method” sections, and it was noted when such information was not provided. Meta-analyses based on previously published data were excluded to avoid artificial inflation. Participant region and author affiliation were classified as (1) the United States (coded separately given Arnett’s prior identification that individuals from the United States are overrepresented in psychology research), (2) countries with English as their first language (i.e., effectively the United Kingdom, Australia, New Zealand, and Canada), (3) European countries that do not have English as their first language, (4) Central and South America, (5) Africa, (6) Asia, and (7) the Middle East and Israel. Coding was not mutually exclusive in that an article featuring participants from different regions contributed data to all regions identified. The proportion of participant samples from each region from these journals across the 5 years assessed are presented in Table 1 (i.e., the percentage of the total number of articles surveyed featuring children from that region).

In terms of raw numbers, of the 1582 articles, 912 featured participants from the United States, 284 from English-speaking countries, and a further 236 from non-English-speaking Europe. Compare this with 112—the total number of articles featuring participants from all of Central and South America, Africa, Asia, and the Middle East and Israel combined. Perhaps more sobering, only 11 articles featured participants from Central and South America, and only 10 featured participants from Africa—fewer than the 23 articles devoted to non-human primates and fewer than the 15 articles that did not specify where their participants were from. In terms of total participant numbers, 633,775 were from the United States—well over double the entire number from the rest of the world combined (286,321, with a
further 1959 unspecified) after excluding data from one study involving all 654,707 births in Sweden from 1983 to 1991. Viewing this from another angle, less than 3% of the participants contributing to the expansion in our knowledge of children’s psychological development came from all of Central and South America, Africa, Asia, and the Middle East and Israel combined (which notably contain ~85% of the world’s population; http://esa.un.org/unpd/wpp/Download/Standard/Population/).

Of 1691 articles assessed for author affiliation, 1029 (61%) were first-authored by faculty members of institutions in the United States, 341 (20%) by those in English-speaking countries, and another 251 (15%) by those in non-English-speaking European countries—with the remaining 4% being first-authored by faculty members of institutions in Asia and Israel. Only 2 articles had the first author located in Central or South America, and none had the first author located in the Middle East or Africa. The critical point about author origin is that it emphasizes how developmental psychology, as a discipline, is characterized as one in which individuals in WEIRD institutions study WEIRD participants.

A wider assessment of developmental journals than that reported by Arnett (2008) did not, therefore, yield a more representative picture. But perhaps this skew in sampling is historical. It has been more than 8 years since Arnett (2008) was published and more than 6 years since Henrich et al. (2010) was published. Given the citation count for the latter (>3000 times according to Google Scholar), change could reasonably be expected. Is there any evidence that our research has become less biased in its underrepresentation of the majority of the world’s humans over the last 8 years?

Following the same method outlined previously, we collated data on participant origin for all articles published in 2008 (361 articles)—the year Arnett’s review was published—and in 2015 (383 articles). As is evident in Fig. 1, little has changed in the sampling region. In 2008, 91.67% of all articles published featured participants from the United States, English-speaking countries, or non-English-speaking Europe, leaving 8.33% representing all of Central and South America, Africa, Asia, and the Middle East and Israel combined. In 2015, the distribution between those two groups was remarkably similar to that in 2008 (92.37% and 7.63% respectively). This does not present a picture of a discipline thoughtfully contemplating its limitations and readily embracing change.

### Conclusion and a way forward

We have highlighted how the vast majority of the world’s population is underrepresented in high-impact developmental psychology research. There are many possible reasons for this. Opportunities for research with culturally heterogeneous samples are typically limited and depend on the commitment of unique, often substantive, temporal and fiscal resources and sometimes depend on years of investing in building trust among relevant communities with little immediate return. Authors from WEIRD backgrounds, hence, are not incentivized to pursue heterogeneous data collection practices. Furthermore, articles published in the journals sampled here are held to the highest standards of empirical rigor, and rejection rates are high. To meet standards for publication commonly requires extensive university training in scientific design, analysis, and writing. Research by staff at non-WEIRD institutions may be less consistent with Western-centric scientific practice or not done at

---

1 The difference between this number and the number analyzed for participant number is due to our inclusion here of meta-analyses, theoretical articles, and review articles.
Language is also a barrier. Whereas journals may encourage submission by authors from non-English-speaking backgrounds and offer to copyedit manuscripts, researchers without sufficient grasp of English to get to that level are forced to publish in local journals (or not at all). This runs another way as well. If reading in English is challenging, then identifying cutting-edge research may be elusive, leading to conducting research that is outside contemporary trends. Similarly, access to peak journals is expensive and might be beyond the budgets of many of the world’s universities. The obstacles may be many, but our preference is that these reasons be viewed as support for encouraging research featuring sampling diversity, not as excuses for perpetuating the status quo of failing to do so.

In certain situations, the exploration of possibly skewed findings as a result of restricted participant sampling might be redundant, but decisions about this need to be made through a lens of awareness and with appreciation of the potential impact of using homogeneous data. This latter point is key. Our suggestion is not that all developmental psychology studies must involve heterogeneous samples; this would be unreasonable and in many cases impractical. We are saying that where samples are from homogeneous groups, consideration should be given to the notion that whatever is being reported may be culturally specific, and hence possibly unrepresentative and not generalizable, and this should be openly acknowledged in print (e.g., in “discussion” sections).

There is a complementary need to acknowledge the implicit “othering” that can occur when issues of culture are referenced. Othering, a form of marginalization whereby individuals or groups are marked as distinct from oneself, is anchored in feminist theory and has been applied to the study of racism, identity, and difference (Ahmad, 1993; Fine, 1994; Hall, 1991; Tomaselli, 2003, 2005; Weis, 1995). Here, we use it to refer to situations where participants who are drawn from WEIRD cultures are considered to be the norm and those who are not are treated as exceptions to the norm. In this light, we borrow from the American Psychological Association (2003) publication of the guidelines on multicultural education, training, research, practice, and organizational change for psychologists and encourage researchers to recognize that, as cultural beings, they may hold attitudes and beliefs that can detrimentally influence their perceptions of research undertaken in populations that are economically, ethnically, and racially different from themselves.

If progress away from the sampling bias inherent in developmental psychology research identified here is to be made, then we need to shift away from this othering point of view. Positive steps forward include (a) encouraging publication of studies that feature non-WEIRD participants, (b) encouraging replication in a new population of a previously established finding, and (c) encouraging theoretically motivated cross-cultural comparisons that examine how children’s cultural environments might affect their development. Having members of editorial boards and grant-funding bodies with sufficient knowledge of the challenges encountered in collecting heterogeneous data will also help, especially when there is a need to distinguish reasonable from unreasonable reviewer critique.
We must be ever attentive to the possibility that where we think we are exploring human universals, we are rather exploring cultural specifics. A continued WEIRD-centric approach also has implications for the ways research is used. Where it forms the foundation for policy development, it is critical that the research match the target population. For example, in the United States there has been considerable political agitation and policy implementation aimed at bridging what has been termed the “word gap”—the disparity in the amount and quality of language that low-income children hear relative to their more affluent peers (Hart & Risley, 1995; Hirsh-Pasek et al., 2015). Many interventions focus on supporting change in the way caregivers interact with their children. Such interventions may be relevant in the United States, but they might not be relevant in other countries where children and their caregivers engage with each other in ways that are not commensurate with dyadic interaction (Little et al., 2016; Rogoff, Mistry, Göncü, & Mosier, 1993).

Reflecting the above, developmental research commonly makes its way into the public domain, and hence it is important to make apparent when there is no foundation for broad generalization of reported findings. For example, the impact of divorce on children has been shown to differ across cultures and economic strata (Fischer, 2007); if a parent is seeking to gain insight into the issues that might confront his or her children after divorce, then consulting literature that does not apply to the parent’s circumstances may lead to unsubstantiated concerns or misguided intervention. Similarly, early childhood development programs are frequently based on sensitivity and mind-mindedness (Meins et al., 2002; Slade, 2005) as the core elements of optimal parenting that should be supported. However, these ideals are highly cultural and may deviate wildly from the models followed by caregivers from other cultures. This might lead practitioners to misinterpret caregivers’ behavior as “problematic” where in fact it is just an expression of another developmental pathway (Otto & Keller, 2014).

There will be criticisms of the concerns laid out here. Haefel, Thiessen, Campbell, Kaschak, and McNeil (2009) suggested that “the problem of generalizability is often overstated. Studies using one sample of humans (e.g., Americans) often generalize to other samples of humans (e.g., Spaniards), particularly when basic processes are being studied (e.g., Anderson, Lindsay, & Bushman, 1999)” (p. 570). This assertion has several shortcomings. First, Haefel and colleagues defined basic processes as “those psychological or biological processes that are shared by all humans at appropriate developmental levels (e.g., cognition, perception, learning, brain organization, genome)” (p. 570). This may be true, but the universality of such basic processes is commonly assumed rather than empirically documented. Second, the article they used to support their claim aimed to provide external validity to laboratory-based research, not to identify human universals. Haefel and colleagues also argued, “It is not enough to show that American culture is different from other cultures. This fact is not disputed. The critical question is what these differences mean for human psychology” (p. 570). We agree, and this is the essence of our point. Theoretically driven, empirically falsifiable endeavors that involve participants across a range of environmental circumstances will enrich our understanding of psychology and help to clarify the validity of research findings.

A new path forward for developmental science is needed to meet this challenge to understand continuity and diversity in human cultural background. Although there may be widespread awareness of this challenge, what we highlight here is that this awareness is not translating into change in the approaches taken to publication strategy. Systematic comparisons across a wide variety of human environments are needed to enable examination of variation and stability in core domains of human psychology, and where convenience sampling is adhered to, the limitations of such an approach must be acknowledged. We need to accept the challenge posed by diversity, provide the explanations it requires, and harness this information to build an improved set of encompassing theories about the development of the human mind.

Acknowledgments

An Australian Research Council Discovery Project Grant (DP140101410) supported the writing of this manuscript. We also thank Karri Neldner and Kristyn Hensby for assistance in collecting the data reported here, and we thank Paul Harris, Bill von Hippel, and Yvette Miller for comments on an earlier draft.
References


