Go WILD, Not WEIRD

Martha Newson
University of Oxford
martha.newson@anthro.ox.ac.uk

Michael Buhrmester
University of Oxford
buhrmester@gmail.com

Dimitris Xygalatas
University of Connecticut
xygalatas@uconn.edu

Harvey Whitehouse
University of Oxford
harvey.whitehouse@anthro.ox.ac.uk

Reliance on convenience samples for psychological experiments has led to the oversampling of Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations (Henrich *et al.* 2010a). Our analysis of academic articles from six leading psychology journals revealed a significantly lower but still very high percentage of studies from European and English-speaking nations (92%), compared to a decade ago (95%), largely due to more studies from Asia (6%). Further analysis of four cognitive science of religion (CSR) journals showed how a more representative field is possible (67% from the Western and Other region), with proportionately more studies in Latin America (4%) and Africa (7%) than psychology (<1% each). Thanks to its interdisciplinary nature, CSR is in a good position to address "WEIRD" problems and may be able to offer psychology methodological and epistemological tools that involve diversifying sample populations, increasing ecological validity, capturing the causes and consequences of cultural variation, and developing novel methodologies. Despite the challenges, we encourage more researchers to embrace the lessons offered by CSR's history of global and interdisciplinary research. Where WEIRD identifies the populations we need to stop privileging, conducting work that is not just *Worldwide*, but also *In Situ*, *Local*, and *Diverse* (WILD) is what researchers themselves can aspire to. Just as nineteenth century "armchair anthropologists" were replaced by generations of ethnographers who went out into the real world to study human variation, so modern day psychologists need to conduct experiments outside the lab with suitably heterogeneous populations.

equinoxonline

Overreliance on *Western, Educated, Industrialized, Rich, and Democratic* (WEIRD) participants in psychological experiments is not merely biased but samples one of the most atypical slices of humanity, both historically and cross-culturally (Heine 2008; *Henrich et al.* 2010a; Henrich *et al.* 2010b). WEIRD participants appear to be outliers in numerous cognitive and behavioral domains, from visual perception, spatial reasoning, and inferential induction, to moral reasoning, decision-making, and the heritability of IQ. While researchers can quite simply refrain from over-generalizing their results, ongoing conversations around sampling and recruitment speak to a deeper issue, which is yet to be properly addressed (Boas 1930; Sears 1986). We argue that shifting to a new framework, moving beyond the traditional confines of psychological research will be the next revolution in the social sciences, one which we may already be seeing signs of (Kagitcibasi 2017; Serpell 2017). With inspiration from Edwin Hutchins' *Cognition in the Wild* (1995), we propose that researchers need to stop focusing on the WEIRD and get WILD. WILD however, is not the opposite of WEIRD – the WILD approach takes researchers out of the confines of comfortable laboratory walls to work with all kinds of populations in the real world (including WEIRD ones).

The psychology of religion has been no exception, not only being biased toward WEIRD populations, but having a tendency to treat "religion" almost exclusively based on a Judeo-Christian understanding of the concept (Friedman 2009; Hill and Pargament 2008; Hood Jr. *et al.* 2018). CSR (the cognitive science of religion) is an interdisciplinary approach to understanding human thought and behaviour, which draws not only on psychology but anthropology, evolutionary sciences, and archaeology (Barrett 2007; Geertz 2014). With a diverse range of methodologies available, and collaborators with expertise in fieldwork, we might expect samples to be more representative of the global population. But is this the case?

We endorse Henrich *et al.*'s position and seek to expand on it, offering solutions rooted in the experiences of researchers who have been working across multiple disciplines including experimental psychology, cognitive anthropology, and social anthropology. These solutions come, at least in part, from lessons learned while conducting research in the field with diverse religious groups and at mass rituals across six continents. Further, we argue that CSR is in a unique position to offer lessons to researchers in other fields, thanks to its interdisciplinary character and an established history of truly cross-cultural samples and keeping both cognition *and* culture in focus (Barrett 1998; Boyer and Ramble 2001; Knight *et al.* 2004).

In recent years, many researchers have developed strategies to avoid the WEIRD problem. These researchers tend to be interdisciplinary in focus

and come from a wide variety of research backgrounds that include training in methods outside of the tradition of laboratory or online experiments and surveys. From examining these researchers' strategies, we have identified four fundamental principles that together may aid in solving the WEIRD problem, which we have dubbed the WILD approach. This approach is:

- *Worldwide:* At a discipline level, participants should come from all regions of the world (as critics of WEIRD oversampling have emphasized).
- *In situ:* Research conducted outside the laboratory is likely to be less prone to the "white room" effect (Cicourel 1996), encouraging participants to behave independently, ideally in their own environments, in turn producing more ecologically valid research (Cole *et al.* 1994).
- *Local:* Cross-cultural research should be informed by local values, attitudes, and belief systems, further contributing to ecological validity. Research uses bottom-up designs rather than carbon-copy protocols that are developed in a lab and implemented in other cultures.
- *Diverse:* Samples reflect human diversity and are not limited to student populations. This may include a broad demographic spectrum or specific foci, and special or atypical populations.

A WILD approach encourages a diverse range of participants, reality-grounded research, increased ecological validity, more information on variation, and a push for novel methodologies. Triangulation and contextualization are essential for thorough research; as a part of this effort, the WILD approach bolsters multi-method, cross-cultural research. Although not *all* researchers necessarily need to adopt the WILD approach (e.g., those seeking proof of concept might do better working with simpler recruitment strategies), in many fields of cognitive and behavioral research, tools sensitive to cultural complexity and diversity are urgently required (Boster 2011).

We begin with a brief examination of the WEIRD problem's prevalence in psychology and CSR in recent years and discuss how solving this problem requires more than simply selecting participants from non-WEIRD countries. This is supported by a further investigation of the extent to which these bodies of research have employed WILD methodologies. Last, we outline the challenges of conducting WILD research and propose a series of steps to overcome them.

Has the WEIRD problem decreased?

Recent methodological approaches to combat the WEIRD problem

Popular approaches to cross-cultural methods have often involved collaborators at international institutions or national surveys (see Klein *et al.* (2014) and Van de Vijver *et al.* (1997) respectively for good examples of each). The upside of

involving international collaborators for a "many labs" approach is that research extends beyond some WEIRD confines by diversifying samples. However, samples still tend to be limited to students who, by merit of attending university, are likely to be Educated, Industrialized, and Rich (and potentially Westernized and Democratic) (Beller *et al.* 2012; Whitehouse and Cohen 2012).

In recent years, a number of solutions to the WEIRD problem have been posed that go beyond classic student recruitment via international colleagues or national polling. Each address certain issues, but none have yet encompassed both samples and the research itself, in a framework as holistic as that of the WILD approach. Here we summarize these recent methodological innovations. Some of these methods are problematic in a wide array of research (e.g., online surveys) and some are more pertinent to cross-cultural research specifically (e.g., ethnographic databases), but all of them have been used in some way to attempt to alleviate the WEIRD bias in recent years.

Online participant recruitment

The simple act of recruiting online, rather than within student-subject pools (often incentivizing would-be participants with course credits) appears to improve external validity, either via surveys or online laboratories. For instance, in one study on gambling behaviors and attitudes, undergraduate psychology students who were recruited from psychology subject-pools behaved differently from other adults, and even from other students who were recruited via websites (Gainsbury *et al.* 2014). Recruiting via online surveys (e.g., advertising on sample-specific websites) offers unprecedented global reach. However, until this reach is maximized, we are left with the predicament that most online participants are still relatively WEIRD.

Mechanical Turk (MTurk) is currently one of the most popular platforms for crowdsourcing participants (Buhrmester *et al.* 2018; Keith and Harms 2016; Lowry *et al.* 2016; Sheehan 2017; Stewart *et al.* 2017). The platform offers relatively low-cost access to a participant pool that is more ethnically and socio-economically diverse than most traditional student samples (Casler 2013). However, online survey-takers are necessarily Internet users living in relatively industrialized contexts – some of whom are likely to become non-naïve to research (Briones and Benham 2017) or even professional survey-takers. In fact, the majority of MTurk samples consist of US workers (Keith and Harms 2016). What's more, MTurk participants are still likely to be more educated and be disproportionately young compared to a truly representative sample, though they tend to have lower household incomes than the rest of the US population, which is where MTurk samples differ from traditional samples (Keith and Harms 2016; see Buhrmester *et al.* 2018 for a comprehensive evaluation of MTurk). While crowdsourcing offers some practical advantages, it offers, at best, only a partial solution to WEIRD problems.

One example of a successful online laboratory is Volunteer Science, which facilitates large-scale, multi-person, and synchronous work (Pilny *et al.* 2016). As with online surveys, research is still required to determine whether being online or offline affects the behaviors of individuals and biases research. And, much like the online marketplace, it is difficult to assess the quality of the data. Again, such techniques rely on biased (Internet-using) samples, so even if the Internet's reach was maximized, a significant proportion of the world's populations will be left out. Nonetheless, these techniques open up possibilities for venturing away from student-based sample populations.

Utilizing "big data"

Web-scraping obtains big data via the automated collection of information from webpages, generating databases with tens to tens of thousands of variables, and thousands of cases, in just a matter of hours (Landers *et al.* 2016). The data are generally behavioral (as opposed to the survey data typically obtained from online studies), involve large sample sizes, are quick to obtain, low in experimenter effects, and given the Internet's increasing global prevalence, provide at least some access to the behavioral patterns of those in nearly every part of the world (Landers *et al.* 2016). However, samples may be less representative than the potential global reach of the Internet. For instance, mining Twitter will result in data that is less WEIRD than psychology samples overall, but is still biased toward North Americans / Western Europeans (34.5% in 2017) and Asian-Pacific users (38.8% in 2017; an increasing annual percentage of users) (Statista 2018). Ultimately, online interactions are qualitatively different to real-life interactions, and the available behavioral data are limited to digital monitoring such as "clicks" or activity cycles, so they are useful for answering a limited set of questions.

Pre-existing ethnographic databases, such as the HRAF World Cultures database (Human Relations Area Files) or the SCCS (Standard Cross-Cultural Sample) (Murdock and White 1969), are another way to test hypotheses in a broad array of cultures. This includes hypotheses resulting from collaborations between psychologists and anthropologists (e.g., Atkinson and Whitehouse 2011). Benefits include the ability to generalize from a representative sample and obtain quantitative measures from ethnographic work (Ember and Cunnar 2015). For instance, the persistent sampling bias in developmental psychology (i.e., toward children from WEIRD environments, see Nielsen *et al.* 2017) could be offset by the considerable progress that has been made into variation in children's play and care-giving using ethnographic records (Ember and Cunnar 2015). Though cheap and benefiting from the established knowledge of a diverse array of well-researched cultures, an obvious downside of this approach is that the data is limited to answering only certain questions. These databases are also extremely fragmentary,

as the data was not recorded systematically to begin with. As such, what is reported is selective and subjective, and analyzing those data involves further subjective decisions. In effect, researchers thus rely on third- or fourth-hand accounts of behavior: the investigator's account of the coder's assessment of the author's interpretation of the informant's report.

Quantifying the problem

Three past investigations have sought to assess the scale of the WEIRD problem in the last decade. First, Arnett (2008) analyzed six APA journals over a 20 year period, largely concentrating on 2003-2007. Samples were heavily skewed, with 95% based in the US, Europe and other English-speaking countries. In a second analysis, Kurzban (2013) found that 90% of studies in the *Journal of Personality and Social Psychology* utilized WEIRD participants, and even in a highly interdisciplinary journal, *Evolution and Human Behavior* (EHB)*,* over 60% of articles were "WEIRD." In a third analysis focused on major developmental journals, Nielsen *et al.* (2017) found that fewer than 10% of samples came from outside North America and Europe.

Given that the vast majority of these analyses focused on articles conducted in the 2000s to early 2010s, we conducted an examination of our own on more recent studies from 2016 and 2018 to see whether the situation is finally changing. Using independent coders, we investigated the extent to which six leading psychology and four CSR journals reported WEIRD, partially, or non-WEIRD research (whether they include Worldwide participants) and examined how these participants were recruited. All articles published in 2016 and 2018 in the following psychology journals were coded: *Psychological Science* (PsySci); *Journal of Personality and Social Psychology* (JPSP); *Developmental Psychology* (DP); *Journal of Abnormal Psychology* (JAP); *Health Psychology* (HP); and *Journal of Educational Psychology* (JEP). The six journals were selected as leading research journals, representing diverse fields of psychology, replicating Arnett's 2008 study (we replaced his *Journal of Family Psychology* for PsySci as we felt DP and EP already represented children and PsySci was the number one research journal by several metrics). We excluded 104 studies that did not use human participants (e.g., animal studies, meta studies, solely statistical modeling) at this stage, leaving a total of 1239 articles. In addition, we coded 2016 and 2018 articles from the following CSR journals: *Religion, Brain and Behavior* (RBB); *Journal for the Cognitive Science of Religion* (JCSR); *Journal of Cognition and Culture* (JCC); and *International Journal for the Psychology of Religion* (IJPR). Thirty were excluded, leaving ninety-six CSR articles and a total sample of 1335. *Journal for the American Academy of Religion* (JAAR) did not have codeable articles for our purposes. See SM1 for methods.

Are WEIRD participants still over-sampled?

Articles in psychology journals were significantly more likely to be WEIRD or partially WEIRD (96.6%) than CSR articles (74.0%), $\chi^2(1) = 95.90$, $p <$ .0005. There was no evidence of change in how WEIRD articles tended to be over the two-year timespan for either field, $p > .423$, nor were changes detected over the two-year period for individual journals. Articles in psychology journals were significantly more likely to report studies with Western & Other participants, compared to articles in CSR journals (Figure 1). This region includes North America, Western Europe, and other English-speaking countries (see SM 2). At the same time, CSR journals reported a greater proportion of studies from every other region than the psychology journals.
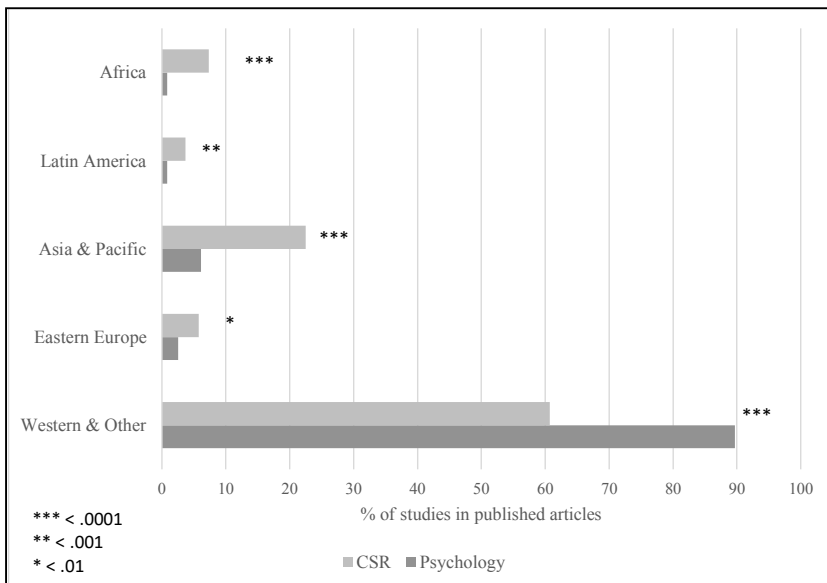

Figure 1: Percentage of CSR and psychology articles reporting regionally diverse samples

Despite journal-specific variation, all ten journals favored Western and Other participants (Table 1). Generally, CSR journals were more likely to be "Worldwide," with three of the journals reporting 30–55% of studies from areas outside the Western and Other region. However, the only journal to report 100% of its participants from this region was *JCSR. RBB* had the most variation in participant recruitment (47% Western and Other; 4% Eastern Europe; 29% Asia and Pacific; 9% Latin America; 11% Africa). Of the six psychology journals, two included around 15% of studies with participants outside of the Western and Other region (*JPSP* and *EP*). The rest reported >90% of studies from the Western and Other region. Latin America and

	PS	JPSP	DP	JAP	HP	JEP	RBB	JCSR	JCC	IJPR
Western and Other	90.1	85.5	91.1	94.7	93.9	86.1	46.7	100	60.3	68.8
Eastern Europe	2.5	4.9	1.3	1.8	1.1	0.7	4.4	0	2.6	10.9
Asia	6.3	8.1	5.3	2.3	3.0	10.4	28.9	0	23.1	18.8
Latin	0.4	0.7	0.9	0.6	1.1	2.1	8.9	0	3.9	0
Africa	0.7	0.7	1.3	0.6	0.7	0.7	11.1	0	10.3	1.6

Table 1: Total studies (%) by region and journal.

Notes: Psychology journals (*n* = 1239): *Psychological Science* (PsySci); *Journal of Personality and Social Psychology* (JPSP); *Developmental Psychology* (DP); *Journal of Abnormal Psychology* (JAP); *Health Psychology* (HP); and *Journal of Educational Psychology* (JEP). CSR journals (*n* = 96): *Religion, Brain and Behavior* (RBB); *Journal for the Cognitive Science of Religion* (JCSR); *Journal of Cognition and Culture* (JCC); and *International Journal for the Psychology of Religion* (IJPR).

Africa constituted around 1% of studies each for psychology journals, compared to 4% and 7% in CSR journals.

Nonetheless, in comparison with Arnett's analysis of six leading APA journals (1988−2007), there has been a slight improvement in diversity for psychology journals. There were significantly fewer studies from the United States, English-speaking countries, Israel, or Europe in 2016 and 2018 (92.2%, including Eastern Europe) compared to 1988-2007 (95%), $\chi^2(1) = 19.52$, $p < .0001$. Asia is now significantly more represented (6.1%), compared to the previous decade (3%), $\chi^2(1) = 34.37$, $p < .0001$. However, there was no difference in how Latin-American or African nations were represented across decades (<1%).

Has the surge of online participant recruitment solved the WEIRD problem?

Online platforms have promised international participant reach. One such platform, Amazon's Mechanical Turk, has been particularly popular amongst social scientists. In our dataset, 9.5% of articles contained MTurk samples (range: 43.4% in *JPSP* – 1.2% in *JAP;* see Table 2). However, the majority of MTurk participants are from the U.S. (Buhrmester *et al.* 2018; Stewart *et al.* 2017). In addition, MTurk is conducted on an English-only platform and requires an internet-capable device, so its use, along with many other similar online platforms, has significant limitations in terms of reaching non-WEIRD participants. Perhaps in response to these criticisms and the proliferation of alternative platforms for online data collection, 2018 psychology articles were significantly less likely to feature MTurk studies (6%) than 2016

articles (11.7%), $\chi^2(1) = 11.29$, $p = .001$. As we only coded for MTurk, the use of online recruitment is likely to be more widespread.

	PS	JPSP	DP	JAP	HP	EP	RBB	JCSR	JCC	IJPR
WEIRD	91.9	88.7	93.7	97.6	95.2	91.7	45.5	100	58.6	72.1
Partially	4.7	11.3	2.6	0.6	0.8	2.3	13.6	0	27.6	0
Non-WEIRD	3.4	0	3.7	1.8	4.0	6.0	40.9	0	13.8	27.9
MTurk sampling	22.2	43.4	1.7	1.2	2.8	1.5	9.1	100	10.3	11.6

Table 2: Percentage of articles rated WEIRD, partially WEIRD, or non-WEIRD, and inclusion of MTurk recruitment, by journal.

Going WILD

Clearly, the WEIRD problem runs deep in the psychological sciences, and technological solutions such as online participant platforms have had little impact on the problem. In contrast, CSR seems to have fared better, likely due to its interdisciplinary nature with access to diverse populations and the pivotal role of anthropologists in shaping the discipline (Atran 2002; Boyer 1992; Geertz 2004; Guthrie *et al.* 1980; Sperber 1996; Whitehouse 1992; Whitehouse 2004). Currently, anthropologists working in CSR are often at the forefront of non-WEIRD research not only in CSR itself but in the psychology of religion more broadly. These anthropologists incorporate psychological measurements into their ethnographic work (Power 2017; Purzycki 2016; Soler 2012; Willard 2018; Xygalatas 2013b), regularly collaborate with colleagues who stem from psychology and other disciplines to conduct research in the field (Kavanagh *et al.* 2018; Shaver *et al.* 2018; Willard and McNamara 2019; Xygalatas *et al.* 2016; Xygalatas *et al.* 2018), and lead or join larger teams to design cross-cultural projects (Gervais *et al.* 2017; Purzycki *et al.* 2016).

We propose that the psychological sciences would benefit from more researchers going "WILD" – that is, conducting research that is *Worldwide, In situ, Local, and Diverse*. Henrich *et al.* (2010) captured the key traits associated with the bulk of psychology participants. In support of this, WILD researchers do not unjustly privilege a culturally-specific cognitive developmental environment. Considering the lack of progress in recruiting beyond WEIRD participants in the last decade, however, we feel there is a need to frame progressive research for what it *is*, rather than what it is *not* (i.e., the WEIRD/non-WEIRD dichotomy).

Worldwide

First, the WILD acronym encapsulates non-WEIRD participants with the use of *Worldwide* (non-Western/industrialized) samples. Going *Worldwide* entails the inclusion of countries that may not necessarily have well-funded universities or psychology departments, so goes beyond replicating studies across continents' richest countries with trained collaborators. Nonetheless, rather than seeking a Rousseauian "noble savage," WILD approaches may still recruit from Western populations, albeit samples aimed toward representative cross-sections of the West (e.g., including broader age ranges and educational backgrounds than typical undergraduate cohorts).

In situ

In situ research tackles the fact that context tends to be downplayed as part of the Western analytic framework (de Oliveira and Nisbett 2017). Psychology as a discipline has both internal variation and the resources with which to develop *In situ* approaches. For instance, social psychology currently accesses *In situ* research methods more than laboratory-based sub-disciplines, such as cognitive psychology (Inbar 2016). This emphasis on contextually sensitive issues helps the discipline solve its ecological validity problem, which is reflected in its relatively poor replication results (Inbar 2016). Rather than chasing the academic holy grail of universality via duplication, the role of replication should be to reveal the behavioral effects of cultural and sociodemographic variation (Greenfield 2017). In other words, lack of replicability could indicate cultural differences across space or cultural/sociodemographic changes across time, rather than methodological weakness.

Local

Local research also improves ecological validity. As Arnett (2008) points out, it would be a mistake to take American-based questionnaires and re-use them in ten different countries. Cultural comparisons reveal a distinctively Western analytic framework, forming the foundations of research questions and approaches in psychology (de Oliveira and Nisbett 2017), as well as interpretation (Keller 2017). As research questions and the surveys that accompany them are laden with cultural assumptions, adaptations to each cultural context (or preferably bottom-up research designs) are essential to avoid imposed etic biases (Berry 1989). Not only does a consideration of Local mean deriving culturally-relevant measures, but also extracting meaningful interpretations of the data and being cautious of generalizations about the human species. For instance, a study across ten cultures with little consideration as to how the environment might produce the results might not necessarily provide much more insight into what it means to be human in *all societies* than a well-considered study across three diverse cultures.

Diverse

WILD participants may original from a single society but at the same time reflect radically different characteristics, as countries host multiple cultures (Keller 2017). For example, within the US there is a diverse array of sub-cultures and communities (Fischer 1984; Woodard 2011) in terms of religion, social ecology, kinship structures, economic markets, and moral behaviors (Graham *et al.* 2016). Recruiting via international colleagues is a relatively cheap and easy way to access non-WEIRD samples, at least in the common usage of the term (i.e., non-Westernized). However, such forms of recruitment may also be heavily reliant on student participant pools or online samples, meaning that although such samples are international they are still markedly WEIRD. Across fifty-nine countries, student samples have been found to be problematic when generalizing to broader populations – both within and between populations (Hanel *et al.* 2016), an ongoing problem (Sears 1986).

Multiple student samples may result in only marginal cultural differences from the student samples in the researcher's native university, making any claims of universals in human behavior dubious. Similarly, urban less-WEIRD populations (which is where most student samples will be found) may behave differently from rural less-WEIRD populations (Talhelm *et al.* 2014). Such diversity speaks of massive variation, which is now empirically well evidenced. But how best to access this variation in less WEIRD, more WILD populations?

WILD

Going truly WILD means more than sampling from "non-WEIRD" populations. Psychological experiments involving cross-cultural replications tend to be carried out by research groups based in different countries or by compiling meta-studies, and university students are still likely to be used as participants (Beller *et al.* 2012; Henrich *et al.* 2010a; Whitehouse and Cohen 2012). Though these methods can increase the size and diversity of samples, there are serious limitations (Beller *et al.* 2012). Often, cross cultural research is designed with WEIRD demographics in mind and is unlikely to utilize methodologies or research questions that reflect the nuances of the specific cultures under study, running the risk of long-term errors in the data due to differences in construct validity across cultural groups (Pollet *et al.* 2014).

Going WILD?

To assess the extent to which recent high-quality research in psychology and CSR reflects WILD principles, we conducted our own investigation. Coders were instructed to score each of the 1,332 articles described above for how *"Worldwide," "In situ," "Local,"* and *"Diverse"* they were. *Worldwide* describes whether the article incorporated studies from less industrialized regions. *In*

situ was assessed according to how independently the participants could behave in the study, such as the incorporation of field studies (ecological validity). *Local* scores were determined by whether studies were informed by the local culture, e.g., cultural appropriateness of materials, a bottom-up approach to research questions, or the use of ethnography to contextualize research. Finally, *Diversity* was assessed according to whether samples were gleaned beyond student subject-pools. This measure tells us when an article in a typically non-WEIRD context uses exclusively student populations. Each variable was coded as "not at all," "partially" (e.g., one study may meet the criteria but not the other studies within the same article), or "fully." Journal-based metrics for WILD can be found in Table 3.

Worldwide

Of the articles in psychology journals, only 0.3% were coded as fully *Worldwide*, 3.1% as partially (i.e., included at least one non-industrialized region), and 96.4% were not *Worldwide* in scope. A further 0.2% did not provide enough information to quantify this variable. Reflecting the pattern to be less WEIRD, CSR articles were significantly more *Worldwide* in scope, $\chi^2(1) = 88.54$, $p < .0001$: 12.5% fully *Worldwide*, a further 12.5% partially *Worldwide*, with 75% recruiting solely from industrialized contexts.

	PS	JPSP	DP	JAP	HP	JEP	RBB	JCSR	JCC	IJPR
Worldwide	3.4	11.3	3.4	0.6	2.4	2.3	45.5	0.0	27.6	14.0
In Situ	27.8	27.4	46.3	24.2	43.4	63.9	68.2	100.0	58.6	74.4
Local	71.4	100.0	72.0	0.8	80.5	60.2	77.3	0.0	75.9	72.1
Distinctive	55.6	60.4	55.9	43.0	49.0	79.7	77.3	100.0	65.5	74.4

Table 3. WILD metrics: % reporting partially or fully WILD results by journal.

In Situ

Fully *In Situ* research comprised 9% of psychology journal articles were fully *In Situ*, 30.5% were partially in the field, and 59.5% were exclusively in the laboratory or online. CSR journals were significantly more likely to be fully or partially *In Situ* than psychology journals, $\chi^2(1) = 31.26$, $p < .0001$: 8.3% were fully In Situ, 60.4% were partially in the field, and 20.8% were exclusively in the laboratory or online. 1% and 10.4% of articles did not provide enough information to code in psychology and CSR journals respectively. Taken together, articles including Eastern European, Asian-Pacific, Latin-American, or African samples were significantly more likely (54.5%) to include *In Situ* methodologies, compared to samples from the Western and Others region (40.2%), $\chi^2(1) = 10.03$, $p = .002$.

Local

Local culture fully informed 68.8% of psychology articles, 6.4% were partially informed, 24.5% did not have evidence of being informed by local culture, and 0.2% lacked sufficient information to code. Perhaps surprisingly, given the influence of anthropology, CSR was not significantly more *Local* than psychology (p = .615) and tallied 46.9% fully informed by local culture, 6.4% partially, and 24.5% with no evidence of being informed by local culture. *Local* did not reflect which region a sample was recruited from (p = .093).

Diverse

Across the psychology articles, 42% were fully diverse, 13.6%. partially diverse, and 42.8% relied on student samples. CSR articles were significantly more diverse than psychology: 44.8% fully diverse, 28.1% partially, and 26% relied on students, $\chi^2(1)$ = 10.88, p = .001. Taken together, Eastern European, Asian-Pacific, Latin-American, or African samples were equally as likely to comprise student samples (61.4%) as samples from the Western and Others region (56.4%), p = .270. Diverse populations were more likely to be involved in articles reporting *In situ* studies (61.7%) than primarily laboratory and online-based articles (53.4%), $\chi^2(1)$ = 9.09, p = .003.

WILD

We computed a WILD score by summing the 4 WILD variables, such that a score of four would indicate that an article at least partially represented a sample that was from beyond the West, was at least partially ecologically valid, at least partially informed by local culture, *and* included at least one study with non-students. A score of zero would indicate no evidence of WILD traits. CSR articles were significantly more WILD (M = 2.40, SD = 1.08) than psychology articles (M = 1.74, SD = .83), $t(103.87)$ = -5.84, p < .0001. This difference largely related to the proportion of articles achieving partial or full WILDness in 4 categories: just 0.6% of psychology articles compared to 16.7% of CSR articles. Both fields had around 3% of articles who were not WILD by any of our metrics. These data support previous research into the WEIRD phenomenon and extend the existing WEIRD / non-WEIRD dichotomy by offering a more nuanced framework with which to assess psychological research. By considering an article's WILD achievements, not only is WEIRDness (*Worldwide* and *Diverse*) of the research considered, but also its ecological validity (*In Situ*) and the import of local culture (*Local*).

Implications of MTurk in WILD research

Limitations with MTurk generated research were revealed in that psychology articles featuring MTurk tended to be less *Diverse* (Spearman's $r(1,235)$ =

-.11, *p* <.0001), while CSR articles using MTurk were more likely to be constrained to the Western & Others region (*r*(96) = -.22, *p* = .035). Using the composite WILD score, articles featuring MTurk were significantly less likely to score highly for WILD (*M* = 1.76, *SD* = .87) than articles which sought participants from elsewhere (*M* = 2.01, *SD* = .77), *t*(1330) = -3.05, *p* = .002, though this effect was only true for psychology journals.

The WILD approach in action

Each WILD sample is likely to be unique in some way. This goes beyond differences in nationality, and cognitive anthropologists have already started to use this variation as an advantageous background from which to test their hypotheses. Cross-cultural research will not deliver high-quality outcomes by having five identical studies, following a single protocol developed in a Western lab, conducted across five countries using online student samples. Rather, a set of five nuanced, WILD studies tackling the same question and using similar but non-identical, locally adapted methodologies, will generate the strongest inferences on the human condition. A good example comes from Gelfand and colleagues who have shown how social traits vary across nations in correspondence to "tightness and looseness" – research achieved via cross-cultural, field, experimental, computational, and neuroscientific methods (Gelfand *et al.* 2017). Thus we advocate the use of multiple methods in cross-cultural research, including the WILD approach.

Numerous studies using anthropological and psychological techniques show that anchoring research questions in close observations of the real world can move psychology beyond WEIRD limitations and assumptions (Atran *et al.* 2005; Barrett and Behne 2005; Cohen 2007; Fessler 2004; Gibson and Mace 2007). This approach has recently been implemented in CSR by a large team of researchers, who systematically adapted experimental designs to 15 different cultures to study religious morality. Applying techniques from cognitive anthropology and experimental economics, the team combined emic data used to reconstruct site-specific cultural models of morality with directly comparable behavioral and survey data. Those data have revealed a wealth of information on both universal and culturally variable aspects of religious morality (Lang *et al.* 2019; Purzycki *et al.* 2016; Purzycki *et al.* 2018a; Purzycki *et al.* 2018b; Purzycki *et al.* 2018c).

Topics that would be impossible to study in laboratory environments have been tackled with WILD methodologies, such as collective rituals (Xygalatas 2016). To study high arousal rituals in Mauritius and Spain, researchers have moved the lab into the field (Xygalatas 2013a), as the cultural salience of such rituals cannot be matched in any laboratory conditions. By adapting experimental methodologies ranging from biometric to psychometric and behavioral measures to a variety of local contexts, this work has, for instance,

shown that these rituals can produce affective alignment and social bonding between participants, as well as how these effects can be modulated based on participants' specific roles in the ceremony (Bulbulia *et al.* 2013; Fischer *et al.* 2014; Konvalinka *et al.* 2011; Xygalatas *et al.* 2011; Xygalatas *et al.* 2013). Other studies have explored health outcomes of participation in mass religious events in India and Mauritius, finding positive effects on wellbeing – studies that only WILD methods, focusing on broad demographics and with high ecological validity in mind, could achieve (Snodgrass *et al.* 2017; Tewari *et al.* 2012; Xygalatas *et al.* 2019).

Another example of the WILD approach in action relates to ritual participation and delayed gratification. Based on substantial fieldwork in Vanuatu, carefully controlled experiments designed for the target samples were conducted in the field (Rybanska *et al.* 2018). These experiments then directly compared European (Slovakia) and Melanesian (Vanuatu) populations, showing that in both populations children who engaged in ritualistic behaviours via a "circle time" game over a three-month period improved their ability to delay gratification.

WEIRD research still dominates psychology and, to a lesser extent, CSR. So how can psychologists capitalize on the WILD framework and improve future research? Given that many psychologists may not know researchers in other fields, or even from other sub-disciplines of psychology, what does truly WILD research look like? What lessons can be gleaned from CSR for researchers in other fields wishing to go WILD? And how does one make WILD research happen?

WILD hurdles

There are three main challenges faced by WILD researchers: finding specialist collaborators, balancing career demands and tenure, and funding (see Wilson and Whitehouse (2016) for a detailed view on establishing and maintaining field sites). First, initiating collaborations with researchers equipped with the tools for WILD research is not always easy: not all fieldworkers are open to collaborating with experimental psychologists; global outreach is both financially costly (e.g., shipping materials or sourcing them locally and the cost of international flights) and time-consuming: e.g., designs must be contextualized; measures must be translated and back-translated; researchers must often access participants in remote locations; and bureaucratic requirements become even more onerous when multiple institutions are involved. Yet their array of international collaborations, novel approaches, tradition of working with non-WEIRD populations and experience with the challenges of fieldwork, and above all their specialist knowledge of religious traditions, make cognitive scientists of religion well-positioned to assist disciplines such as psychology to "go WILD." In turn, psychologists bring fresh constructs,

measures, experimental designs, and quantitative methodologies as well as access to wider audiences and funding opportunities.

There is a special challenge in going WILD as an early career scientist. Psychologists, for instance, are driven to publish as many articles as possible before applying for tenure, whereas anthropologists and scholars of religion are expected to produce a single detailed monograph as a final thesis. While anthropological journals privilege ethnographic detail and context-specific knowledge during the peer-review process, psychological journals tend to focus more on methodological and analytical rigor, but, as illustrated in this paper, place little emphasis on ecological validity. The threat of "publish or perish" tends to push psychological scientists toward the quick and easy, but creating long-standing relationships with fieldworkers (and their field sites) can be a great asset to researchers in all stages of their careers.

Finally, one of the biggest challenges faced by modern researchers is funding. WILD research is often likely to be uniquely positioned among grant proposals, as it is inherently inter-disciplinary, with global outreach, and high impact due to its ecological validity and unusual populations. To investigate funding options, we coded the funding bodies reported in the analysed articles. A total of 539 funding bodies were listed across the 1,239 psychology articles and 62 across the 96 CSR articles. No funder was listed for 28.3% and 34.4% of psychology and CSR articles respectively. Within psychology, the three main funders were university of departmental sources (14.2%), the National Institute of Health (9.6%), and the National Science Foundation (7.2%). Within CSR, the three biggest funders were the Templeton Foundation (15.6%), the Social Sciences and Humanitites Research Council (12.5%), and the Cultural Evolution of Religion Research Consortium (CERC). See SM2 for a journal breakdown of most common funders.

Heterogeneous samples entail clear advantages for scientific research. There may be occasions when the pursuit of such data results in reduced sample sizes or compromised controls (which can be ironed out in future replications). If editorial boards and grant funding bodies could incorporate an appreciation of the practicalities associated with WILD research, it is likely that more of this much called for research could be conducted. In this vein, critiques of research that makes an effort to extend into WILD domains could be better distinguished as reasonable or unreasonable (Nielsen *et al.* 2017), e.g., replicating laboratory standards of control in the field and matching the increasingly common large sample sizes associated with online recruitment in remote regions.

Conclusions

Many of the currently proposed solutions to the replication crisis, such as pre-registration, increased sample sizes, and rigid replications, only place

more pressure on researchers to deliver Western-centric research. Science is a cumulative endeavor, but the accumulation of knowledge need not come merely through ever-increasing sample sizes, whether within or across populations. This alone adds very little to the ecological validity of the research, as any design errors are only likely to be amplified. A culturally uninformed study that is conducted with a larger sample or in more locations may provide a false sense of greater confidence in the results while in fact simply reproducing erroneous measures and constructs that are of little relevance to the populations studied (Xygalatas 2019).

By diversifying samples, the WILD approach tackles those issues ignored by focusing on replicability beyond improving external validity. WILD approaches will increase internal validity, construct validity, consequentiality, and cumulativeness (Finkel *et al.* 2017). This is not to say that WEIRD research is no longer important. Indeed, to address some research questions, WEIRD samples constitute an especially suitable starting point. Furthermore, with a process of globalized Gesellschaft-directed social change, much of the world – or at least the technologically accessible regions – will become increasingly WEIRD (Greenfield 2017). This is something that WILD researchers must continue to pay close attention to.

By teaming up with anthropologists, psychologists are better placed to collect data among non-WEIRD populations – with the possibility of reducing levels of laboratory artifice, yet retaining the explanatory power of carefully controlled experiments. In addition, there are non-academic fieldworkers, with extensive networks and local knowledge, who may welcome the experimental and statistical rigor that psychologists can provide. These include, among others, humanitarian and aid workers, and researchers from international corporations.

The last decade has seen a slight but significant decrease in WEIRD sampling bias in psychological research, mainly due to improvements in sampling participants from the Asian-Pacific region (largely in the field of education research). Nonetheless, WEIRD samples are still vastly over-represented, with 92% of research coming from European, Israeli, or English-speaking nations in psychology journals and 67% in CSR journals. In their efforts to increase the representativeness of their samples, psychologists not only need to capture *Worldwide* populations but conduct research that is *In Situ, Local and Diverse* – in a word, *WILD*. Drawing on the interdisciplinary tools and collaborative networks commonly used by researchers in CSR would be one way to achieve this. This is a call for psychologists to knock on the doors of anthropologists, sociologists, and other fieldworkers, to make new friends, and utilize one another's tools and methods. There is no quick fix to the WEIRD paradigm: it is expensive, risky, and time-consuming, but engaging with WILD collaborators and approaching WILD research is the first step.

Acknowledgements

References

Arnett, J. Jeffrey. 2008. "The Neglected 95%: Why American Psychology Needs to Become Less American." *American Psychologist* 63(7): 602−614. https://doi.org/10.1037/0003-066X.63.7.602

Atkinson, D. Quentin and Harvey Whitehouse. 2011. "The Cultural Morphospace of Ritual Form: Examining Modes of Religiosity Cross-Culturally." *Evolution and Human Behavior* 32(1)**:** 50−62. https://doi.org/10.1016/j.evolhumbehav.2010.09.002

Atran, Scott. 2002. *In Gods We Trust: The Evolutionary Landscape of Religion*. Oxford: Oxford University Press.

Atran, Scott, Douglas L. Medin and Norbert O. Ross. 2005. "The Cultural Mind: Environmental Decision Making and Cultural Modeling Within and Across Populations." *Psychological Review* 112(4): 744−776. https://doi.org/10.1037/0033-295X.112.4.744

Barrett, H. Clark, and Tanya Behne. 2005. "Children's Understanding of Death as the Cessation of Agency: A Test Using Sleep Versus Death." *Cognition* 96(2): 93−108. https://doi.org/10.1016/j.cognition.2004.05.004

Barrett, L. Justin. 1998. "Cognitive Constraints on Hindu Concepts of the Divine." *Journal for the Scientific Study of Religion* 37(4): 608−619. https://doi.org/10.2307/1388144

———. 2007. "Cognitive Science of Religion: What is it and Why is it?" *Religion Compass* 1(6): 768−786. https://doi.org/10.1111/j.1749-8171.2007.00042.x

Beller, Sieghard, Andrea Bender and Douglas L. Medin. 2012. "Should Anthropology Be Part of Cognitive Science?" *Topics in Cognitive Science* 4(3): 342-353. https://doi.org/10.1111/j.1756-8765.2012.01196.x

Berry, W. John. 1989. "Imposed Etics − Emics − Derived Etics: The Operationalization of a Compelling Idea." *International Journal of Psychology* 24(2-6): 721−735. https://doi.org/10.1080/00207598908247841

Boas, Franz. 1982 [1930]. "Some Problems of Methodology in the Social Sciences." In *Race, Language, and Culture*, edited by Franz Boas, 260−269. Chicago: Chicago Press.

Boster, James. 2011. "Data, Method and Interpretation in Cognitive Anthropology." In *A Companion to Cognitive Anthropology*, edited by David B. Kronenfeld, Giovanni Bennardo, Victor C. de Munck and Michael D. Fisher, 131−152. Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444394931.ch8

Boyer, Pascal. 1992. "Explaining Religious Ideas: Elements of a Cognitive Approach." *Numen* 39(1): 27–57. https://doi.org/10.1163/156852792X00159

Boyer, Pascal and Charles Ramble. 2001. "Cognitive Templates for Religious Concepts: Cross-Cultural Evidence for Recall of Counter-Intuitive Representations." *Cognitive Science* 25(4): 535–564. https://doi.org/10.1207/s15516709cog2504_2

Briones, M. Elizabeth and Grant Benham. 2017. "An Examination of the Equivalency of Self-Report Measures Obtained from Crowdsourced Versus Undergraduate Student Samples." *Behavior Research Methods* 49(1): 320–334. https://doi.org/10.3758/s13428-016-0710-8

Buhrmester, D. Michael, Sanaz Talaifar and Samuel D. Gosling. 2018. "An Evaluation of the Rapid Rise of MTurk and Recommendations for its Effective Use." *Perspectives on Psychological Science* 13(2): 149–154. https://doi.org/10.1177/1745691617706516

Bulbulia, A. Joseph, Dimitris Xygalatas, Uffe Schjoedt, Sabela Fondevila, Chris G. Sibley *et al.* 2013. "Images from a Jointly-Arousing Collective Ritual Reveal Affective Polarization." *Frontiers in Psychology* 4: 1–11. https://doi.org/10.3389/fpsyg.2013.00960

Casler, Krista, Lydia Bickel and Elizabeth Hackett. 2013. "Separate But Equal? A Comparison of Participants and Data Gathered via Amazon's MTurk, Social Media, and Face-to-Face Behavioral Testing." *Computers in Human Behavior* 29(6): 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Cicourel, V. Aaron. 1996. "Ecological Validity and 'White Room Effects': The Interaction of Cognitive and Cultural Models in the Pragmatic Analysis of Elicited Narratives From Children." *Pragmatics & Cognition* 4(2): 221–264. https://doi.org/10.1075/pc.4.2.04cic

Cohen, Emma. 2007. *The Mind Possessed: The Cognition of Spirit Possession in an Afro-Brazilian Religious Tradition*. Oxford: Oxford University Press.

Cole, Michael, Lois Hood and Ray McDermott. 1994. *Ecological Niche Picking: Ecological Invalidity as an Axiom of Experimental Cognitive Psychology*. New York: Rockefeller University.

de Oliveira, Stephanie and Richard E. Nisbett. 2017. "Culture Changes How We Think about Thinking: From 'Human Inference' to 'Geography of Thought'." *Perspectives on Psychological Science* 12(5): 782–790. https://doi.org/10.1177/1745691617702718

Ember, R. Carol, and Christine M. Cunnar. 2015. "Children's Play and Work: The Relevance of Cross-Cultural Ethnographic Research for Archaeologists." *Childhood in the Past* 8(2): 87–103. https://doi.org/10.1179/1758571615Z.00000000031

Fessler, Daniel. 2004. "Shame in Two Cultures: Implications for Evolutionary Approaches." *Journal of Cognition and Culture* 4(2): 207–262. https://doi.org/10.1163/1568537041725097

Finkel, J. Eli, Paul W. Eastwick and Harry T. Reis. 2017. "Replicability and Other Features of a Highquality Science: Toward a Balanced and Empirical Approach." *Journal of Personality and Social Psychology* 113(2): 244–253. https://doi.org/10.1037/pspi0000075

Fischer, S. Claude. 1984. *The Urban Experience.* 2nd ed. San Diego, CA: Harcourt Brace Jovanovich.

Fischer, Ronald, Dimitris Xygalatas, Panagiotis Mitkidis, Paul Reddish, Penny Tok *et al.* 2014. "The Fire-Walker's High: Affect and Physiological Responses in an Extreme Collective Ritual." *PloS ONE* 9(2): e88355. https://doi.org/10.1371/journal.pone.0088355

Friedman, Harris. 2009. "Xenophilia as a Cultural Trap: Bridging the Gap between Transpersonal Psychology and Religious/Spiritual Traditions." *International Journal of Transpersonal Studies* 28(1): 107–111. https://doi.org/10.24972/ijts.2009.28.1.107

Gainsbury, M. Sally, Alex Russell and Alex Blaszczynski. 2014. "Are Psychology University Student Gamblers Representative of Non-University Students and General Gamblers? A Comparative Analysis." *Journal of Gambling Studies* 30(1): 11–25. https://doi.org/10.1007/s10899-012-9334-9

Geertz, W. Armin. 2004. "Cognitive Approaches to the Study of Religion". In *New Approaches to the Study of Religion. Vol 2, Textual, Comparative, Sociological, and Cognitive Approaches,* edited by Peter Antes, Armin W. Geertz and Randi R. Warne, 347–399. Berlin: Walter deGruyter.

Geertz, W. Armin, ed. 2014. *Origins of Religion, Cognition and Culture*. London; Abingdon: Routledge. https://doi.org/10.4324/9781315728988

Gelfand, J. Michele, Jesse R. Harrington and Joshua C. Jackson. 2017. "The Strength of Social Norms across Human Groups." *Perspectives on Psychological Science* 12(5): 800–809. https://doi.org/10.1177/1745691617708631

Gervais, M. Will, Dimitris Xygalatas, Ryan T. McKay, Michiel van Elk, Emma E. Buchtel *et al.* 2017. "Global Evidence of Extreme Intuitive Moral Prejudice against Atheists." *Nature Human Behaviour* 1(8): 1–6. https://doi.org/10.1038/s41562-017-0151

Gibson, A. Mhairi, and Ruth Mace. 2007. "Polygyny, Reproductive Success and Child Health in Rural Ethiopia: Why Marry a Married Man?" *Journal of Biosocial Science* 39(2): 287–300. https://doi.org/10.1017/S0021932006001441

Graham, Jesse, Peter Meindl, Erica Beall, Kate M. Johnson and Li Zhang. 2016. "Cultural Differences in Moral Judgment and Behavior, across and within Societies." *Current Opinion in Psychology* 8: 125–130. https://doi.org/10.1016/j.copsyc.2015.09.007

Greenfield, M. Patricia. 2017. "Cultural Change Over Time: Why Replicability Should Not Be the Gold Standard in Psychological Science". *Perspectives on Psychological Science* 12(5): 762–771. https://doi.org/10.1177/1745691617707314

Guthrie, Stewart, Joseph Agassi, Karin R. Andriolo, David Buchdahl, H. Byron Earhart *et al.* 1980. "A Cognitive Theory of Religion [and Comments and Reply]." *Current Anthropology* 21(2): 181–203. https://doi.org/10.1086/202429

Hanel, H. Paul, Katia C. Vione and Martin Voracek. 2016. "Do Student Samples Provide an Accurate Estimate of the General Public?" *PloS one* 11(12): e0168354. https://doi.org/10.1371/journal.pone.0168354

Heine, J. Steven. 2008. *Cultural Psychology*. New York: W. W. Norton.

Henrich, Joseph, Steven J. Heine and Ara Norenzayan. 2010a. "Beyond WEIRD: Towards a Broad-Based Behavioral Science." *Behavioral and Brain Sciences* 33(2-3): 111–135. https://doi.org/10.1017/S0140525X10000725

———. 2010b. "The Weirdest People in the World?" *Behavioral and Brain Sciences,* 33(2-3): 61–83. https://doi.org/10.1017/S0140525X0999152X

Hill, C. Peter, and Kenneth Pargament. 2008. "Advances in the Conceptualization and Measurement of Religion and Spirituality: Implications for Physical and Mental Health Research." *Psychology of Religion and Spirituality* S(1): 3–17. https://doi.org/10.1037/1941-1022.S.1.3

Hood, W. Ralph Jr., Peter C. Hill and Bernard Spilka. 2018. *The Psychology of Religion: An Empirical Approach.* Fifth edition. New York; London: Guilford.

Hutchins, Edwin. 1995. *Cognition in the Wild.* Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/1881.001.0001

Inbar, Yoel. 2016. "Association Between Contextual Dependence and Replicability in Psychology May Be Spurious." *Proceedings of the National Academy of Sciences of the United States of America* 113(34): E4933-4. https://doi.org/10.1073/pnas.1608676113

Kagitcibasi, Cigdem. 2017. "Doing Psychology with a Cultural Lens: A Half-Century Journey." *Perspectives on Psychological Science* 12(5): 824–832. https://doi.org/10.1177/1745691617700932

Kavanagh, M. Christopher, Jonathan Jong, Ryan McKay and Harvey Whitehouse. 2018. "Positive Experiences of High Arousal Martial Arts Rituals Are Linked to Identity Fusion and Costly Pro-Group Actions." *European Journal of Social Psychology* 49(3): 461–481. https://doi.org/10.1002/ejsp.2514

Keith, G. Melissa and Peter D. Harms. 2016. "Is Mechanical Turk the Answer to Our Sampling Woes?" *Industrial and Organizational Psychology* 9(1): 162–167. https://doi.org/10.1017/iop.2015.130

Keller, Heidi. 2017. "Culture and Development: A Systematic Relationship." *Perspectives on Psychological Science* 12(5): 833–840. https://doi.org/10.1177/1745691617704097

Klein, A. Richard, Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams, Jr., Stepán Bahník *et al.* 2014. "Investigating Variation in Replicability: A 'Many Labs' Replication Project." *Social Psychology* 45(3): 142–152. https://doi.org/10.1027/1864-9335/a000178

Knight, Nicola, Paulo Sousa, Justin L. Barrett and Scott Atran. 2004. "Children's Attributions of Beliefs to Humans and God: Cross-Cultural Evidence." *Cognitive Science* 28(1): 117−126. https://doi.org/10.1207/s15516709cog2801_6

Konvalinka, Ivana, Dimitris Xygalatas, Joseph Bulbulia, Uffe Schjødt, Else-Marie Jegindø *et al.* 2011. "Synchronized Arousal Between Performers and Related Spectators in a Fire-Walking Ritual." *Proceedings of the National Academy of Sciences* 108(20): 8514−8519. https://doi.org/10.1073/pnas.1016955108

Kurzban, Rob. 2013. "Is Evolutionary Psychology WEIRD or NORMAL?" *The Evolutionary Psychology Blog*. Available from: http://epjournal.net/blog/2013/09/is-evolutionary-psychology-weird-or-normal/

Landers, N. Richard, Robert C. Brusso, Katelyn J. Cavanaugh and Andrew B. Collmus. 2016. "A Primer on Theory-Driven Web Scraping: Automatic Extraction of Big Data from the Internet for Use in Psychological Research." *Psychological Methods* 21(4): 475−492. https://doi.org/10.1037/met0000081

Lang, Martin, Benjamin G. Purzycki, Coren L. Apicella, Quentin D. Atkinson, Alexander Bolyanatz *et al.* 2019. "Moralizing Gods, Impartiality, and Religious Parochialism Across 15 Societies." *Proceedings of the Royal Society B* 286(1898): 1−10. https://doi.org/10.1098/rspb.2019.0202

Lowry, B. Paul, John D'Arcy, Bryan Hammer and Gregory D. Moody. 2016. "'Cargo Cult' Science in Traditional Organization and Information Systems Survey Research: A Case for Using Nontraditional Methods of Data Collection, Including Mechanical Turk and Online Panels." *The Journal of Strategic Information Systems* 25(3): 232−240. https://doi.org/10.1016/j.jsis.2016.06.002

Mesoudi, Alex, Kesson Magid, Delwar Hussain and Christine A. Caldwell. 2016. "How Do People Become WEIRD? Migration Reveals the Cultural Transmission Mechanisms Underlying Variation in Psychological Processes." *PloS ONE* 11(1): e0147162. https://doi.org/10.1371/journal.pone.0147162

Murdock, P. George, and Douglas R. White. 1969. "Standard Cross-Cultural Sample." *Ethnology* 8(4): 329−369. https://doi.org/10.1371/journal.pone.0147162

Nielson, Mark, Daniel Haun, Joscha Kärtner and Christine H. Legare. 2017. "The Persistent Sampling Bias in Developmental Psychology: A Call to Action." *Journal of Experimental Child Psychology* 162: 31−38. https://doi.org/10.1016/j.jecp.2017.04.017

Pilny, Andy, Brian Keegan, Brooke Wells, Chris Riedl, David Lazer *et al.* 2016. "Designing Online Experiments: Citizen Science Approaches to Research." In *Proceedings of the 19ᵗʰ ACM Conference on Computer Supported Cooperative Work and Social Computing Companion,* pp. 498−502. New York: Association for Computing Machinary. https://doi.org/10.1145/2818052.2855516

Pollet, V. Thomas, Joshua M. Tybur, Willem E. Frankenhuis and Ian J. Rickard. 2014. "What Can Cross-Cultural Correlations Teach Us about Human Nature?" *Human Nature* 25(3): 410−429. https://doi.org/10.1007/s12110-014-9206-3

Power, A. Eleanor. 2017. "Discerning Devotion: Testing the Signaling Theory of Religion." *Evolution and Human Behavior* 38(1): 82–91. https://doi.org/10.1016/j.evolhumbehav.2016.07.003

Purzycki, G. Benjamin. 2016. "The Evolution of Gods' Minds in the Tyva Republic." *Current Anthropology* 57(S13): S88-S104. https://doi.org/10.1086/685729

Purzycki, G. Benjamin, Coren Apicella, Quentin D. Atkinson, Emma Cohen, Rita Anne McNamara, Aiyana K. Willard, Dimitris Xygalatas, Ara Norenzayan, and Joseph Henrich. 2016. "Moralistic Gods, Supernatural Punishment and the Expansion of Human Sociality." *Nature* 530(7590): 327–330. https://doi.org/10.1038/nature16980

Purzycki, G. Benjamin, Joseph Henrich, Coren Apicella, Quentin D. Atkinson, and Adam Baimel, Emma Cohen, Rita Anne McNamara, Aiyana K. Willard, Dimitris Xygalatas, and Ara Norenzayan. 2018a. "The Evolution of Religion and Morality: A Synthesis of Ethnographic and Experimental Evidence from Eight Societies." *Religion, Brain & Behavior* 8(2): 101–132. https://doi.org/10.1080/2153599X.2016.1267027

Purzycki, G. Benjamin, Anne C.Pisor, Coren Apicella, QuentinAtkinson, EmmaCohen, Joseph Henrich, Richard McElreath, Rita A.McNamara, Ara Norenzayan, Aiyana K.Willard, Dimitris Xygalatas. 2018b. "The Cognitive and Cultural Foundations of Moral Behavior." *Evolution and Human Behavior* 39(5): 490–501. https://doi.org/10.1016/j.evolhumbehav.2018.04.004

Purzycki G. Benjamin, Cody T. Ross, Coren Apicella, Quentin D. Atkinson, Emma Cohen, Rita Anne McNamara, Aiyana K. Willard, Dimitris Xygalatas, Ara Norenzayan, Joseph Henrich. 2018c. "Material Security, Life History, and Moralistic Religions: A Cross-Cultural Examination." *PloS ONE* 13(3): e0193856. https://doi.org/10.1371/journal.pone.0193856

Rybanska, Veronika, Ryan McKay, Jonathan Jong and Harvey Whitehouse. 2018. "Rituals Improve Children's Ability to Delay Gratification." *Child Development* 89(2): 349–359. https://doi.org/10.1111/cdev.12762

Sears, O. David. 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. https://doi.org/10.1037/0022-3514.51.3.515

Serpell, Robert. 2017. "How the Study of Cognitive Growth Can Benefit from a Cultural Lens." *Perspectives on Psychological Science* 12(5): 889–899. https://doi.org/10.1177/1745691617704419

Shaver, H. John, Martin Lang, Jan Krátký, Eva Kundtová Klocová, Radek Kundt *et al.* 2018. "The Boundaries of Trust: Cross-Religious and Cross-Ethnic Field Experiments in Mauritius." *Evolutionary Psychology* 16(4): 17 December. https://doi.org/10.1177/1474704918817644

Sheehan, B. Kim. 2017. "Crowdsourcing Research: Data Collection with Amazon's Mechanical Turk." *Communication Monographs* 85(1): 140–156. https://doi.org/10.1080/03637751.2017.1342043

Snodgrass, G. Jeffrey, David E. Most and Chakrapani Upadhyay. 2017. "Religious Ritual Is Good Medicine for Indigenous Indian Conservation Refugees: Implications for Global Mental Health." *Current Anthropology* 58(2): 257−284. https://doi.org/10.1086/691212

Soler, Montserrat. 2012. "Costly Signaling, Ritual and Cooperation: Evidence from Candomblé, an Afro-Brazilian Religion." *Evolution and Human Behavior* 33(4): 346−356. https://doi.org/10.1016/j.evolhumbehav.2011.11.004

Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

Statista. 2018. *Distribution of Twitter Users Worldwide from 2012 to 2018, by Region* [Online]. Available from: https://www.statista.com/statistics/303684/regional-twitter-user-distribution/

Stewart, Neil, Jesse Chandler and Gabriele Paolacci. 2017. "Crowdsourcing Samples in Cognitive Science." *Trends in Cognitive Sciences* 21(10): 736−748. https://doi.org/10.1016/j.tics.2017.06.007

Talhelm, Thomas, Jonathan Haidt, Shigehiro Oishi, Xuemin Zhang, Felicity F. Miao *et al.* 2014. "Liberals Think More Analytically (More "WEIRD") than Conservatives." *Personality and Social Psychology Bulletin* 41(2): 250−267. https://doi.org/10.1177/0146167214563672

Tewari, Shruti, Sammyh Khan, Nick Hopkins, Narayanan Srinivasan, Stephen Reicher *et al.* 2012. "Participation in Mass Gatherings Can Benefit Well-Being: Longitudinal and Control Data from a North Indian Hindu Pilgrimage Event." *PLoS ONE* 7(10): e47291. https://doi.org/10.1371/journal.pone.0047291

van de Vijver, J. R. Fons, and Kwonk Leung. 1997. *Methods and Data Analysis for Cross-Cultural Research*. London: Sage.

Whitehouse, Harvey. 1992. "Memorable Religions: Transmission, Codification and Change in Divergent Melanesian Contexts." *Man* 27(4): 777−797. https://doi.org/10.2307/2804174

———. 2004. *Modes of Religiosity: A Cognitive Theory of Religious Transmission*. Walnut Creek: Altamira Press.

Whitehouse, Harvey and Emma Cohen. 2012. "Seeking a Rapprochement Between Anthropology and the Cognitive Sciences: A Problem-Driven Approach." *Topics in Cognitive Science* 4(3): 404−412. https://doi.org/10.1111/j.1756-8765.2012.01203.x

Willard, K. Aiyana. 2018. "Religion and Prosocial Behavior Among the Indo-Fijians." *Religion, Brain & Behavior* 8(2): 227−242. https://doi.org/10.1080/2153599X.2016.1267032

Willard, K. Aiyana and Rita A. McNamara. 2019. "The Minds of God(s) and Humans: Differences in Mind Perception in Fiji and North America." *Cognitive Science* 43(1): e12703. https://doi.org/10.1111/cogs.12703

Wilson, S. David and Harvey Whitehouse. 2016. "Developing the Field Site Concept for the Study of Cultural Evolution." *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* 7(2): 228–287. https://doi.org/10.21237/C7CLIO7233542

Woodard, Colin. 2011. *American Nations: A History of the Eleven Rival Regional Cultures of North America*. New York: Penguin Random House.

Xygalatas, Dimitris. 2013a. "Bringing the Lab into the Field: Using Mixed Methods to Study Religion in the Wild." *Sociální Studia* 10(2): 15–25. https://doi.org/10.5817/SOC2013-2-15

———. 2013b. "Effects of Religious Setting on Cooperative Behavior: A Case Study from Mauritius." *Religion, Brain & Behavior* 3(2): 91–102. https://doi.org/10.1080/2153599X.2012.724547

———. 2016. "Bridging the Gap Between Laboratory and Field. Commentary on David Sloan Wilson and Harvey Whitehouse, 'Developing the Field Site Concept for the Study of Cultural Evolution'". *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* 7(2): 267–270. https://doi.org/10.21237/C7CLIO7233542

———. 2019. "Strong Interdisciplinarity and Explanatory Pluralism in Social Scientific Research." *Social Science Research Council, Insights from the Social Sciences* [Online]. Available from: https://items.ssrc.org/insights/strong-interdisciplinarity-and-explanatory-pluralism-in-social-scientific-research/

Xygalatas, Dimitris, Ivana Konvalinka, Joseph Bulbulia and Andreas Roepstorff. 2011. "Quantifying Collective Effervescence: Heart-Rate Dynamics at a Fire-Walking Ritual." *Communicative & Integrative Biology* 4(6): 735–738. https://doi.org/10.4161/cib.17609

Xygalatas, Dimitris, Sammyh Khan, Martin Lang, Radek Kundt, Eva Kundtová Klocová, Jan Krátký, and John Shaver. 2019. "Effects of Extreme Ritual Practices on Health and Well-Being." *Current Anthropology* 60(5): 699–707. https://doi.org/10.1086/705665

Xygalatas, Dimitris, Eva Kundtová Klocová, Jakub Cigán, Radek Kundt, Peter Maňo, Silvie Kotherová, Panagiotis Mitkidis, Sebastian Wallot, and Martin Kanovsky . 2016. "Location, Location, Location: Effects of Cross-Religious Primes on Prosocial Behavior." *The International Journal for the Psychology of Religion* 26(4): 304–319. https://doi.org/10.1080/10508619.2015.1097287

Xygalatas, Dimitris, Panagiotis Mitkidis, Ronald Fischer, Paul Reddish, Joshua Skewes, Armin W. Geertz, Andreas Roepstorff, Joseph Bulbulia. 2013. "Extreme Rituals Promote Prosociality." *Psychological Science* 24(8): 1602–1605. https://doi.org/10.1177/0956797612472910

Xygalatas, Dimitris, Silvie Kotherová, Peter Mano, Radek Kundt, Jakub Cigán *et al.* 2018. "Big Gods in Small Places: The Random Allocation Game in Mauritius." *Religion, Brain & Behavior* 8(2): 243–261. https://doi.org/10.1080/2153599X.2016.1267033

SM1: How to decrease reliance on WEIRD populations and foster WILD approaches

Like previous researchers, we approach WEIRDness on a continuum rather than a WEIRD vs. non-WEIRD dichotomy, which tends to only reflect developed vs. less developed countries (rather than diversity of samples) (Mesoudi *et al.* 2016; Henrich *et al.* 2010b). However, this dichotomy is useful for analytical purposes.

Coders were also asked to tally the number of times articles sourced participants from different world regions. These regions were determined by the UN's regional groups, for which the "Western and Other" region broadly maps on to our concept of "WEIRD" regions (e.g., North America, North/Western Europe, Australia, New Zealand, and Israel are included). This differs from Arnett's 2008 classifications, as we grouped "Western" nations together and included Israel; distinguished Eastern and Western Europe; and separated Africa from the Middle East.

The coders were familiar with Henrich *et al.*'s paradigm and were instructed to code articles as non-WEIRD, partially WEIRD (e.g., one study was not WEIRD but other studies in the same article were WEIRD), or fully WEIRD by reading each article's abstract and methodology. The extent to which an article could be considered as using a sample from a non-industrialized context was also coded (non-industrialized, partially industrialized, industrialized), and this variable was expected to mirror WEIRDness. There was strong agreement between the coders as to whether articles reported WEIRD studies ($\kappa = .85$, $p < .0005$). In cases of disagreement, the two coders resolved the issue via discussion. If they could still not agree, a third coder was included to resolve the case leading to a single final dataset. There was also strong agreement as to how many times each region was represented, so mean values were computed for each region in the final dataset ($\kappa > .82$, $p < .0005$).

	PS	JPSP	DP	JAP	HP	JEP	RBB	JCC	IJPR
1	University / department	University / department	University / department	National Institute of Mental Health	National Institute of Health	National Science Foundation	Social Sciences and Humanities Research Council		Templeton Foundation
2	National Institute of Health	National Science Foundation	EKSNI for Child Health and Human Development	National Institute of Health	U.S. National Institute on Aging	U.S. Dept. Education, Institute of Education Sciences	Templeton Foundation		*Remaining funders listed once*
3	National Science Foundation	Social Sciences and Humanities Research Council of Canada	National Science Foundation	University / department	National Cancer Institute	German Research Foundation	CERC	*Each funder listed once*	
4	European Research Council	National Institute of Health	National Institute of Health	National Institute on Drug Abuse	University / department	EKSNI for Child Health and Human Development	Canadian Institute for Advanced Research		
5	German Research Foundation	John Templeton Foundation	National Institute of Child Health and Human Development	National Institute on Alcohol Abuse and Alcoholism (US)	Canadian Institutes of Health Research	University / departmental specific	Rutherford Discovery Fellowship		

Figure 3: SM2 Funders (EKSNI: Eunice Kennedy Shriver National Institute; CERC: Cultural Evolution of Religion Research Consortium)