

# What Can Cross-Cultural Correlations Teach Us about Human Nature?

Thomas V. Pollet · Joshua M. Tybur ·  
Willem E. Frankenhuis · Ian J. Rickard

Published online: 5 August 2014  
© Springer Science+Business Media New York 2014

**Abstract** Many recent evolutionary psychology and human behavioral ecology studies have tested hypotheses by examining correlations between variables measured at a group level (e.g., state, country, continent). In such analyses, variables collected for each aggregation are often taken to be representative of the individuals present within them, and relationships between such variables are presumed to reflect individual-level processes. There are multiple reasons to exercise caution when doing so, including: (1) the ecological fallacy, whereby relationships observed at the aggregate level do not accurately represent individual-level processes; (2) non-independence of data points, which violates assumptions of the inferential techniques used in null hypothesis testing; and (3) cross-cultural non-equivalence of measurement (differences in construct validity between groups). We provide examples of how each of these gives rise to problems in the context of testing evolutionary hypotheses about human behavior, and we offer some suggestions for future research.

**Keywords** Ecological fallacy · Cross-cultural research · Research methods · Simpson's paradox · Non-independence · Measurement equivalence

Many hypothesis-testing techniques used by evolutionary behavioral scientists involve examining degrees of covariation between two or more variables and making inferences based on such relationships (e.g., analysis of variance, Pearson correlation). This approach may be applied to a number of different research designs that examine variation at different, hierarchically organized levels (e.g., individuals versus groups;

---

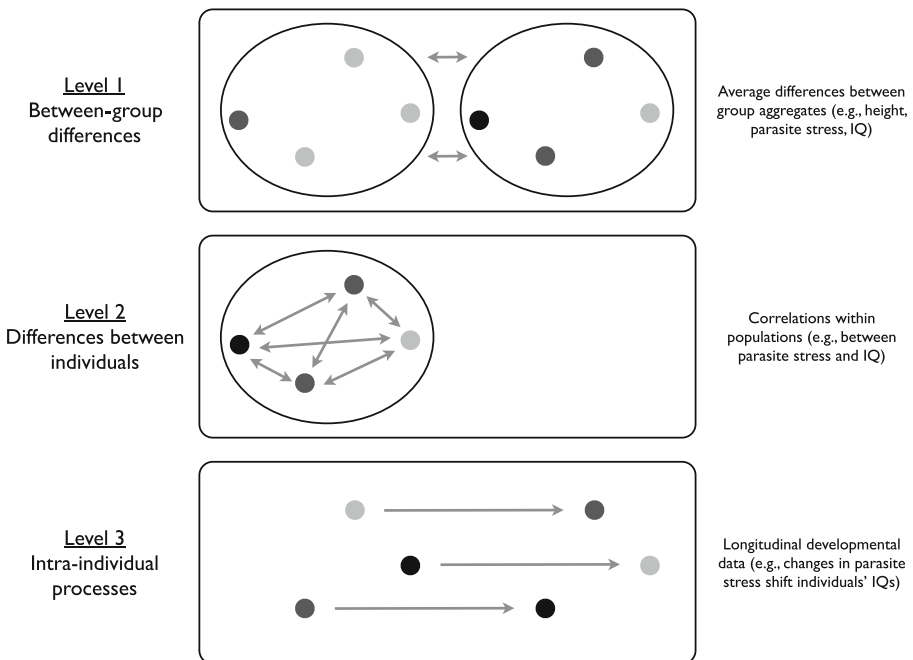
T. V. Pollet (✉) · J. M. Tybur  
Department of Social and Organizational Psychology, VU University Amsterdam, van der  
Boechorststraat 1, 1081BT Amsterdam, The Netherlands  
e-mail: t.v.pollet@vu.nl

W. E. Frankenhuis  
Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands

I. J. Rickard  
Department of Anthropology, Durham University, Durham, UK

see Fig. 1), and hypothesis tests at different levels often require different analytic approaches. For example, the comparative method involves examining variation between species in order to learn about the phylogenetic history of traits within groups of animals, e.g., among great apes (Hare et al. 2001; Wasserman 1993). Within the same species, between-population differences can be used to study the evolution of genetic differences between populations (Burger et al. 2007; Powell et al. 2009) and developmental plasticity (West-Eberhard 2003). Within a given population, between-individual differences may also illuminate the maintenance of genetic variation (Cagliani et al. 2012) and developmental plasticity (Frankenhuis and Panchanathan 2011).

Over the past few years, evolutionary behavioral scientists have increasingly used macro-level data (e.g., region, state or country) to test hypotheses about inter- and intra-individual processes (Fig. 1). Investigations using these techniques have addressed topics such as the effects of ecological pathogen stress on crime, values, and cognitive ability (e.g., Cashdan and Steele 2013; Eppig et al. 2010; Hackman and Hruschka 2013; Shrira et al. 2013; Thornhill and Fincher 2011), parent-offspring conflict over mate choice (e.g., Apostolou 2010), sex ratios (both operational and offspring sex ratios: e.g., Barber 2000; Dama 2011, 2012, 2013; Kruger and Nesse 2006; Kruger and Schlemmer 2009; Thomas et al. 2013), mate preferences (e.g., DeBruine et al. 2010; DeBruine et al. 2011; Moore et al. 2013), pair bonding (e.g., Quinlan and Quinlan 2008), parental investment (e.g., Barber 2003; Marlowe 2003), homosexual preferences (Barthes et al. 2013), personality (e.g., Schmitt et al. 2008), economic decision-making (e.g., Marlowe et al. 2008, 2011), and sexual dimorphism (Wells 2012).



**Fig. 1** Three levels at which hypotheses can be analyzed: between groups, and within individuals over time. Note: Interactions between levels might exist as well; for instance, if the differences between individuals, or the developmental trajectories of individuals, differ between groups

Analyses of cross-cultural correlations have several merits, including interdisciplinary appeal, intersecting the social and biological sciences, and investigating populations outside of the “Western, educated, industrialized, rich, educated, and democratic” (WEIRD; Henrich et al. 2010) populations typically sampled by psychologists. Despite these merits, macro-level data are characterized by inherent limitations in what they can tell us about individual-level processes. Problems inherent in extrapolation across levels of analyses are not limited to cross-cultural correlations on humans but also pertain to studies of life history. For example, researchers may generate predictions based on individual-level optimization but test these predictions by observing variation between populations in life-history traits (e.g., Anderson 2010). Given that patterns of variation at different levels of analysis (species, population, individual) are influenced by multiple mechanisms (e.g., genetic variations, facultative trade-offs, phenotypic plasticity), covariation at one level cannot be straightforwardly interpreted as reflecting a specific process at another level. Similarity between patterns of covariation at different levels may be driven by fundamentally different mechanisms operating at these different levels. For instance, animals such as birds can be grouped according to various levels of taxonomic organization that are hierarchical with respect to one another, ranging from “upper” (e.g., across families and orders) to “lower” (e.g., genera and species). Even though population size among British birds is negatively related to body size across families and orders, the relationship between these same variables is reversed across species and genera (Harvey et al. 1995; Nee et al. 1991, 1996).

Challenges associated with drawing inferences about individuals from macro-level relationships have received substantial attention in the social and biological sciences at large (Cavusgil and Das 1997; Gravelle 1998; Hui and Triandis 1985; Kievit et al. 2013; Mace and Pagel 1994; Nasif et al. 1991; Van de Vijver and Leung 1997), but such discussions remain to be fully integrated into the human evolutionary behavioral sciences (for exceptions, see Hawkes 1994; Nettle 2009). Our goal here is to facilitate progress in this direction by discussing limitations of macro-level data and offering constructive suggestions for interpreting results using macro-level data.

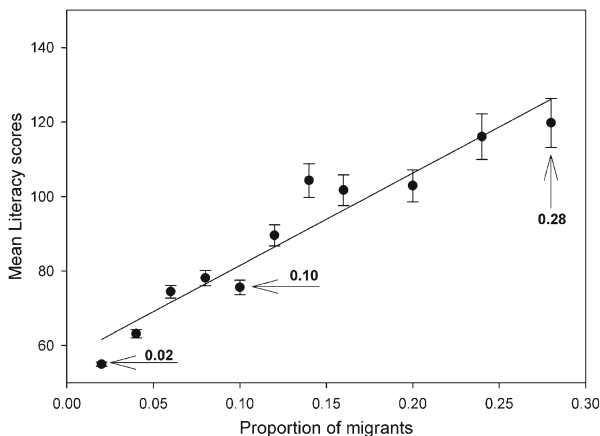
First, we describe the ecological fallacy, which can occur when statistical patterns observed at one level of analysis (e.g., country) are assumed to generalize to other levels of analyses (e.g., covariation between individual-level personality traits). Next, we describe two other issues pertinent to between-population (including cross-cultural) analyses. Thus, we cover three issues relevant to cross-cultural correlations, starting with the ecological fallacy; then, non-independence of observations; and finally, cross-cultural non-equivalence in measurement.

## The Ecological Fallacy

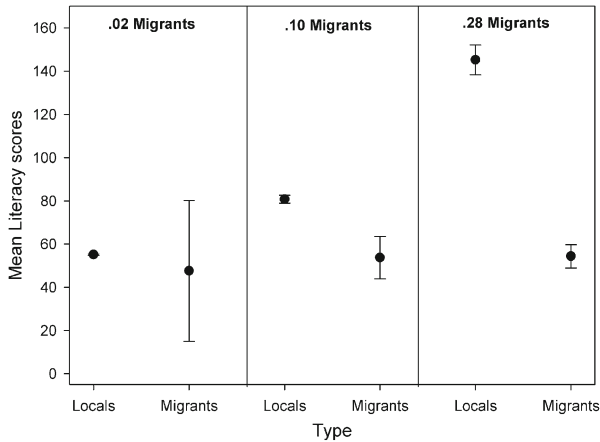
The term “ecological fallacy” became well known after William Robinson (1950) used U.S. census data to test hypotheses related to immigration and literacy (see also Menzel 1950; Thorndike 1939). Robinson considered what might be described as a “brain-drain” hypothesis: those who are literate are more likely to migrate, and therefore proportions of immigrants within a state will be positively related to literacy rates in those states. Consistent with this hypothesis, Robinson (1950) found evidence for such

a positive relationship between the average literacy of U.S. states and the proportion of immigrants living in those states. However, at the individual level, immigrants were *less* likely to be literate than native individuals (see also Schwartz 1994; Subramanian et al. 2009; Te Grotenhuis et al. 2011 for discussions). The positive state-level relationship between proportion of immigrants and literacy rates might have arisen because immigrants tended to settle in states with higher literacy levels, perhaps because these states afforded better economic opportunities or were otherwise more tolerant of immigrants. Thus, literacy levels are higher in some states *despite*, rather than because of, lower literacy among immigrants. The state-level literacy statistics at the aggregate level did not accurately reflect the literacy of immigrants (and, indeed, portrayed a pattern that was opposite to the individual-level pattern). In sum, then, the ecological fallacy is committed when group-level relationships are assumed to reflect individual-level relationships. The fallacy can occur when group aggregates are incorrectly assumed to be representative of individuals within those groups, or when macro-level relationships are governed by processes that are unrelated to those hypothesized to operate at the individual level. In the Robinson (1950) study on literacy, for example, the scores at state level were assumed to represent literacy of immigrants and non-immigrants equally, whereas at the individual level, immigrants were less likely to be literate than non-immigrants.

We will illustrate this pattern using a hypothetical example in Figs. 2 and 3. We simulated data for 11 hypothetical states (with 400 inhabitants each). Next, we generated a pattern whereby immigrants have a randomly assigned literacy score between 0 and 100 and locals have a random value between 50 and  $x$ , where  $x$  ranges between 60 and 220 across states (i.e., generally higher than that of immigrants). The value of  $x$ , however, varies with the proportion of migrants, with states with a larger proportion of migrants having a greater  $x$ . Figure 2 shows a positive relationship between the proportion of migrants and literacy rates at a state level. This pattern might seem to support the “brain drain” hypothesis. However, within each state, locals have a higher literacy score than immigrants. The gap between migrants and locals in literacy scores is most pronounced in the state with the highest proportion of migrants, for which the



**Fig. 2** Example of aggregated relationship between proportion of migrants and literacy rates at state level with simulated data. Error bars represent 95% confidence intervals



**Fig. 3** Relationship between migrant status and literacy scores *within* three hypothetical states (simulated data; proportions of 0.02, 0.10, and 0.28 from Fig. 2). Error bars represent 95% confidence intervals

aggregate data would suggest the highest literacy rate among migrants. Figure 3 shows that the positive relationship between state literacy level and proportion of migrants exists despite there being no change in the literacy of migrants between states.

Equivalent issues might pertain to recent investigations in the human evolutionary sciences. For instance, Thornhill and Fincher (2011) show that U.S. states with higher estimated levels of parasite stress have higher rates of romantic partner murder, homicide, and male-honor homicide. From these patterns, the authors infer that exposure to parasites during development (i.e., parasite stress) has a causal effect on individuals' propensity for violence. Parasite stress is argued to affect individuals' in-group sociality, which in turn affects a broader set of interpersonal relationship dynamics, including violence. However, the critical individual-level patterns *within states*—arguably, the relationship relevant to the research question of whether parasite stress affects values and violence—could be absent, weaker, or (as in the literacy example above) reversed, regardless of the state-level pattern. Hence, group-level correlations such as these are not particularly relevant to individual-level hypotheses.

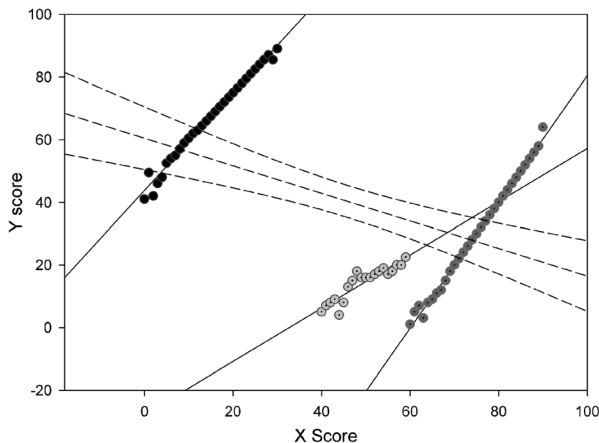
The degree to which the ecological fallacy has affected conclusions drawn in the field of evolution and human behavior is difficult to estimate. It seems to be the case, however, that even in instances where the aggregate measures of central tendency provide a reasonably good descriptor of the underlying level, the ecological fallacy can occur. Firebaugh (1978) argued that bias in regression models with aggregated data is absent *only* in the case when the group mean of the independent variable ( $x$ ) has no effect on the (lower-level) dependent variable  $y$ , with (lower-level)  $x$  controlled (also see Sheppard 2003). However, in many cases in the field of evolution and human behavior (e.g., Eppig et al. 2010), this “rule” is difficult to test because data at the lower level are either absent or difficult to obtain. Nevertheless, it is clear that drawing inferences from the incorrect level of analysis can lead to inaccurate conclusions. Indeed,

as in our example and several published others (e.g., Connolly 2006; Dutton 1994; Yip and Liu 2006), the relationship can be reversed entirely. We note that, in some cases, effects at the underlying level can actually also be stronger than those reported at the macro level (Piantadosi et al. 1988). Even in cases where there are ample data at different levels (e.g., on suicide and socioeconomic status: Rehkopf and Buka 2006), it is not always clear which factors determine the reversal of the relationships between levels and how often these reversals occur. The only solution in this case would be to collect more data and empirically assess how often the aggregate pattern corresponds to the lower-level pattern.

### Simpson's Paradox

When data are aggregated and a change in a statistical relationship can occur owing to the influence of an unknown variable, this situation is known as Simpson's Paradox, a specific form of the ecological fallacy (Simpson 1951; see Scheiner et al. 2000 for example; Kievit et al. 2013 for review). At its most extreme, Simpson's Paradox is manifested as a reversal in the sign of a relationship between analyses at the aggregate and individual levels. Figure 4 illustrates such a case with hypothetical data, where within three groups there is a moderate to strong positive relationship between the two variables  $x$  and  $y$ . However, at the aggregate level, the relationship is reversed, and  $x$  and  $y$  are negatively associated.

Several cases of Simpson's Paradox have been documented in social science literatures. For example, a study on voting patterns in recent US presidential elections found a positive relationship between a state's average wealth and support for the Democratic Party (Gelman et al. 2007). However, this association is reversed at the level of individuals: wealthier individuals tend to support the Republican Party, with this relationship being moderated by state income. The apparent paradox, in this case, arises because poorer states are more likely to vote Republican for reasons other than income.



**Fig. 4** Illustration of Simpson's Paradox. Within each group the association between  $x$  and  $y$  is positive. However, at the aggregate level the relationship is reversed. Solid lines are Ordinary Least Squares (OLS) regression for each subgroup; dashed lines represent OLS fit and 95% confidence intervals

## Artificial Inflation of Effect Sizes

In addition to the complete reversal of the purported effect when a lower level is examined rather than the aggregate level, the effect size at the aggregate level can be dramatically overestimated. In psychology, Brand and colleagues (2010) demonstrated that averaging over multiple trials can strongly inflate effect size estimates (also see Brand and Bradley 2012; McCormick 2013 for discussion of this issue). A related issue has been documented in social neuroscience, whereby aggregating across voxels, volumetric pixels used in brain imaging, can inflate effect size (Vul et al. 2009). Similar issues of inflation can occur in the context of cross-cultural correlations, where aggregation can inflate effect sizes.

As discussed above for the ecological fallacy, the degree to which the process of aggregation influences the effect sizes of cross-cultural correlations is largely unknown since most of the time data at lower levels are either not available or not reported. However, we can provide one specific example from the evolutionary behavioral sciences. Researchers have suggested that, because higher levels of some Big Five personality traits (namely, extraversion and openness to experience) might increase exposure to pathogens, pathogen exposure might partially contribute to the maintenance of variability in these personality traits (e.g., Schaller and Murray 2008). This hypothesis has been tested at a group level, where covariation between national averages of personality and national averages of parasite prevalence have been examined, and at an individual level, where covariation between individual differences in personality and variables posited to reflect investment in pathogen avoidance have been examined. At the group level, Schaller and Murray (2008) report that the correlations between national averages of parasite prevalence and national averages of extraversion and openness both equal  $-0.59$ . Using an individual differences approach, though, Duncan et al. (2009) report that the correlations between individual differences in perceptions of infectability (e.g., self-reports of infection frequency and severity) and extraversion and openness are  $-0.06$  and  $-0.03$ , respectively. Similarly, Tybur and colleagues (2011) report that individual differences in sensitivity to pathogen disgust correlate weakly with individual differences in extraversion and openness ( $-0.05$  and  $-0.24$ , respectively; see also Tybur and De Vries 2013). The effect sizes of relationships between other variables measured at the group level in this literature might similarly overestimate the magnitude of relationships at the individual level, and effect sizes at the group level should be interpreted with caution.

## Cross-Level Interactions

Examinations that focus exclusively on cross-cultural correlations are also unable to test for possible cross-level interactions, which arise when individuals only behave in particular ways in particular environments (Piantadosi et al. 1988; Subramanian et al. 2009). To illustrate, data indicate a positive correlation between per capita meat consumption at a national level and Gross Domestic Product (GDP) per capita (York and Gossard 2004). From this correlation, we may be tempted to infer that meat consumption should be relatively higher among wealthier regions. However, the true pattern may vary across regions depending on region wealth. Consider meat consumption in Belgium as an example. Meat consumption is higher in Wallonia (a poorer

region based on GDP) than it is in Flanders (a more affluent region) (Vleesbarometer 2011). From the correlation between meat consumption and GDP, we might also be tempted to infer that, *within* a given group (e.g., region or nation), richer individuals would eat more meat. However, this inference is also unwarranted: the relationship could be positive, negative, or null. Further, the relationship could be different within different groups. Indeed, this seems to be the case. In some developing countries, individual income positively relates to (bush)meat consumption (e.g., in Gabon; Wilkie et al. 2005). Meat consumption in developing countries can also vary depending on local markets (Brashares et al. 2004); moreover, within the same ecosystem the relationship between bushmeat consumption and wealth can differ between different populations (Mgawe et al. 2012). In other countries, especially developed countries, income is negatively associated with meat consumption: wealthier individuals tend to consume less meat than poorer individuals. In the UK, for instance, the top 20% richest households consume 986 g of meat per person (on average), whereas the 20% poorest households consume 1,029 g of meat per person (on average) (DEFRA 2011). In our example, the ecological fallacy would involve extrapolating the relationship between wealth and meat consumption from macro-level data to the level of regions or individuals. Such a straightforward extrapolation would also ignore the possibility of other factors playing a role, such as the possibility that the relationship between meat consumption and wealth can vary as a function of other factors, such as inequality. Cross-level interactions might allow for a more accurate description of, and tests of predictions of, human behavior, allowing for individual attenuation as a function of characteristics of local environments. The relevance of cross-level interactions has long been recognized in the medical sciences whereby neighborhood or region effects might attenuate individual predispositions toward certain risks (e.g., Dahl et al. 2006; Von Korff et al. 1992; Winkleby et al. 2006). Similarly, in the education sciences, attributes of the school can interact with traits at the pupil level to determine learning outcomes (e.g., Bosker et al. 1990). Cross-level interactions may have been largely underexplored in the field of evolution and human behavior in favor of a more narrow focus on cross-cultural correlations.

### How Can the Ecological Fallacy Be Dealt with?

The examples we described above illustrate some reasons why data patterns at one level of analyses (e.g., between nations) are not necessarily informative about processes occurring at other levels (e.g., between individuals). Similarities in the direction of effects between studies carried out at different levels might, in the worst case, be nothing more than coincidence, and differences between patterns at different levels of analysis may be so extreme that relationships at one level are opposite to that observed at another level. The fundamental problem that the ecological fallacy illustrates is the loss of vital information concerning individual-level processes when aggregates are used as a substitute. Aggregate data can be examined in order to investigate questions pertinent to that level of variation, but similarity between patterns observed or expected at one level and those observed at another level may be entirely coincidental.

Researchers in the field of evolution and human behavior could address such challenges by developing new longitudinal projects or exploring existing databases that cover data on multiple levels (e.g., National Longitudinal Survey of Youth 2007;



National Child Development Study 2012; Avon Longitudinal Study of Parents and Children [Golding et al. 2001], Millennium Cohort Study [Plewis et al. 2006], and Tracking Adolescents' Individual Lives Survey [Huisman et al. 2008]). For example, as argued by Nettle (2009), recent hypotheses concerning relationships between parasite stress and other variables, such as intelligence, collectivism, religiosity, and ingroup investment (e.g., Eppig et al. 2010; Eppig et al. 2011; Fincher and Thornhill 2012; Fincher et al. 2008), could be tested in a more robust manner using such datasets. Although these measures typically do not assess parasite stress directly, relevant proxies—such as self-reported health, number of doctor visits, number of infections, and outcome measures (e.g., intelligence, values)—are available. Importantly, these variables are measured at the appropriate level: for each individual in the sample. An additional benefit is that they are measured in a consistent way across different points in time, and the chronology of events could suggest causal associations in a manner that cross-cultural correlations cannot. Moreover, many of the datasets have detailed information on the environment in which individuals developed. This can then be meaningfully linked to existing indices of historical pathogen stress (for example, children from different regions in the NLSY can be compared via geocoding). We therefore suggest that stronger tests of many hypotheses derived from the pathogen stress model will examine how pathogen stress predicts outcomes such as intelligence, sociopolitical values, or criminal behavior using individual longitudinal designs, such as the NLSY, rather than aggregated data from a single point in time (also see Nettle 2009). Other opportunities lie in using individual-level data from multiple societies to make cross-population comparisons (e.g., Borgerhoff Mulder et al. 2009; Hill et al. 2011) or mathematical simulation (e.g., for the evolution of complex societies: Turchin et al. 2013; for developmental processes: Frankenhuis and Panchanathan 2011).

In the field of evolution and human behavior, many predictions are tested that involve *interactions* between higher and lower levels of analyses (i.e., cross-level interactions: some individuals behave in specific ways in particular ecological contexts). For example, individuals who were exposed to deprivation during childhood might only speed up their reproductive strategy when relevant cues are present in their current environment (Grisevicius et al. 2011). Multilevel modeling strategies not only allow for the correct specification of multiple levels but can be used to test predictions on interactions between higher and lower levels of analyses (for reviews, see Gelman and Hill 2007; Hox 2010; Raudenbush and Bryk 2002; Snijders and Bosker 1999). Although multilevel modeling has occasionally been used for cross-cultural correlational data (e.g., Eppig et al. 2010), in those models countries have been nested within regions rather than individuals being nested within countries. It is difficult to locate data at both the individual and regional levels, but several resources do exist for testing cross-level hypotheses. For example, data on child development (e.g., in the NLSY) include descriptions of where children grow up, which can be linked to existing databases of historical pathogen stress, as well as data on individual-level pathogen exposure, which can be combined to predict intelligence. Moreover, such an approach can differentiate between effects of individual exposure to parasites and historical parasite prevalence in a region, which can inform underlying processes (e.g., cultural evolution via facultative shifts in response to ecological conditions versus developmental trade-offs). An additional advantage is that these multilevel models can account for the reality that meaningful variation exists at many different levels. These models

can nest intra-individual variation in individuals who are in turn nested into meaningful units such as households, provinces, regions, countries, and so forth.

To be clear: we are not arguing that macro-level patterns—at group, societal, or cultural levels—are uninformative about, or inconsequential for, the study of human behavior. On the contrary, cultural evolutionists have developed detailed models examining the conditions in which specific individual-level strategies result in particular group-level traits, which in turn affect the adaptive value of individual-level strategies (e.g., Boyd and Richerson 1985). We specifically take aim at inferences drawn from group-level data about individual-level processes. We also do not argue that group-level data are irrelevant (see Lubinski and Humphreys 1996; Schwartz 1994) or cannot be used to test hypotheses at the *group level*. Rather, we stress that statistical relationships at a macro level should not be assumed to reflect equivalent relationships at a micro level. Finally, the ecological fallacy does not imply that statistical models using individual-level data (or multilevel models) are inherently superior to those using group-level data (Piantadosi et al. 1988; Schwartz 1994). Instead, the ecological fallacy shows that mismatches between those levels at which empirical data are gathered and analyzed, and those at which hypotheses are formulated, are problematic.

## Secondary Issues with Cross-Cultural Correlations

We will now highlight two issues that are not specific to extrapolation from group-level data to individual-level processes but are common to all interpretations of analyses involving cross-cultural correlations. Like the ecological fallacy, non-independence of observations and lack of cross-cultural equivalence can affect the degree to which inferences can be drawn from observations of cross-cultural correlations.

### Non-Independence of Observations

Cross-cultural correlations often suffer from statistical non-independence (also known as Galton's Problem; Mace and Pagel 1994; Nettle 2009). This issue has been well-documented and widely discussed in anthropology (e.g., Eff 2004; Naroll 1965; Ross and Homer 1976) and sociology (e.g., Herkenrath 2002) but has received minimal attention in work in the field of evolution and human behavior, at least in research relying heavily on cross-cultural correlations. Notably, Currie and Mace (2012) and Nettle (2009) have raised the issue of non-independence in reference to tests of the parasite stress model with cross-cultural correlations. Even widely used samples designed to partially counter this problem, such as the standard cross-cultural sample (SCCS; Murdock and White 1969), can suffer from this problem (Dow 2007; Dow and Eff 2008; Eff 2004). This problem is not limited to studies using aggregated data or studies examining country-level data, in which neighboring nations (e.g., Germany and Austria) may have many similarities not specified in the statistical model (e.g., latitude, language, diet). Non-independence of units can lead to erroneous inferences; for example, rather than an effect being driven by the identified variable of interest (e.g., pathogen stress, GDP, inequality), the effect could be due to an unspecified third variable (e.g., latitude, language, diet, religion).

Most inferential statistical techniques rely on the premise of random sampling of independent observations from some distribution (see Howell 2010). Violations of this assumption can threaten the validity of statistical inference; in other words, researchers might infer that a statistical relationship exists, or does not exist, in the population even though this inference is unwarranted because the sample is biased (Type I and II errors, respectively). Country- or region-level data do not generally meet the assumption of random sampling since neighboring countries or regions are similar to one another in many ways.

Researchers who are conscious of this issue may attempt to account for non-independence by statistically controlling for region or continent, or by analyzing at higher macro-levels (e.g., Murdock and White 1969 cultural regions; Fincher et al. 2008). However, non-independence also exists *within* larger geographical clusters. For example, within Europe, Austria is more similar in many respects to Germany (shared language, cuisine, medieval and recent history) than it is to France. Also, these regional clusters are not independent. For example, in terms of values, because of their shared history, a geographical cluster of North America may be more similar to Europe than to sub-Saharan Africa.

At least three solutions exist to combat this issue. One solution is to employ methods that take into account the shared variance between neighboring countries by modeling the covariance between countries, as is done in “random effects models” (e.g., McCulloch and Neuhaus 2001). These models explicitly incorporate shared variance between countries (modeled as a random intercept) as well as the shared variance as a function of shared traits between countries (e.g., social inequality; modeled as a random slope). It is also possible to model the effects of autocorrelation at different clusters (e.g., as done for the SCCS: Dow and Eff 2008). A second approach is to construct cultural phylogenies (e.g., Mace, Jordan, and Holden 2003; Mace and Pagel 1994 for a review; Currie et al. 2010). Thus far we focused our description of the problem of non-independence predominantly on lack of statistical independence, but lack of “cultural” independence might be at play as well. We acknowledge that many evolutionary behavioral scientists have avoided the issues relating to cross-cultural correlations that we raise here by explicitly considering that the correlation between traits can be a result of their shared cultural history. Analogous to accounting for shared variance due to phylogenetic history in biology (e.g., Nee et al. 1996), researchers examining the coevolution between cultural traits have used phylogenetic methods to account for shared cultural history (Mace et al. 2003; Mace and Pagel 1994). This use of cultural phylogenies has allowed researchers to model shared variance between cultures based on these cultures’ histories (inferred from a language phylogeny, for example). If the relationship between two traits remains after taking into account shared (cultural) variance, this constitutes evidence that these traits are related through some other process. This phylogenetic approach has also allowed for the establishment of a chronology of cultural traits. A potential drawback, however, is that this phylogenetic method assumes that cultural traits are transmitted faithfully (e.g., Borgerhoff Mulder 2001; Borgerhoff Mulder et al. 2006; see Towner et al. 2012 for a hybrid approach).

A third possibility, largely unexplored by evolutionary behavioral scientists, is to use spatial statistics (e.g., based on GIS mapping, Chang 2003; Waller and Gotway 2004). This approach takes into account the physical distance between regions, assuming that neighboring regions are more alike. Using the spatial coordinates, researchers can

correctly account for the non-independence in spatial data. Spatial statistics have been used widely in the field of animal behavior and ecology (e.g., Boyce and McDonald 1999; Brown et al. 2009; Rushton et al. 1997; Wilkin et al. 2006). It is possible, however, that the assumption that neighboring regions are similar to each other and that distance is therefore a good proxy for similarity is not always upheld. For example, although Austria borders both Germany and Italy, it has more in common with the former than the latter in terms of diet and language. A spatial approach might treat the difference between Austria and Germany and between Austria and Italy as equal, whereas a phylogenetic approach based on language would not do so. When the assumption that distance is a reasonable proxy for similarity is not upheld, a phylogenetic approach could be preferred, which perhaps explains why spatial statistics have not been commonly used in the field of evolution and human behavior.

### Cross-Cultural Non-equivalence of Measurement

When psychologists test hypotheses relevant to group differences on psychological constructs, they often compare scores on particular instruments across groups (e.g., intelligence: Wechsler 1981 Adult Intelligence Scale; Raven 2000 Progressive Matrices; and personality: the Big Five Inventory [John and Srivastava 1999]; the NEO-PI-R [Costa and MacCrae 1992], the HEXACO model [Ashton and Lee 2007, 2009; Lee and Ashton 2004]). However, the validity of inferences based on these instruments entails multiple, non-trivial assumptions that are not always tested (see Byrne and Campbell 1999; Poortinga 1989; Van de Vijver and Leung 1997; Van de Vijver and Tanzer 2004). These issues include the degree to which group-level data are acquired using similar sampling techniques (e.g., if estimates of some nations' scores are inferred from samples of college students, whereas estimates of other nations' scores are inferred from foragers), the degree to which methods of data collection vary across groups (e.g., paper-pencil questionnaire versus interview), the degree to which the same assessment method may produce different demand characteristics across groups, and the degree to which factor structures, factor loadings, and item intercepts are equivalent across groups.

Heine et al. (2002) highlight one example of the fallibility of cross-cultural comparisons based on self-reports: the “reference group effect.” Put simply, the validity of group-level data (i.e., means of individuals within those groups) can be attenuated by the degree to which self-report responses reflect comparisons to norms within a group. For example, self-reported height (a seven-point scale of “How tall are you?” with 1 = not tall and 7 = very tall) might yield the same group mean in India and the Netherlands despite large group differences in actual height. The same issue applies to individual difference variables commonly used by psychologists (e.g., collectivism; see Heine et al. 2002). Simple differences in the interpretation of a concept within an item (or set of items) may influence inferences in a manner independent from the construct of interest. Indeed, Heine et al. (2008) find that, whereas group-level aggregates of self-reported conscientiousness correlate poorly with theoretically relevant behavioral measures (e.g., group-average walking speed, clock accuracy), *perceptions* of national averages of conscientiousness (so, the degree to which an individual rates their culture as conscientious, relative to the U.S.; Terracciano et al. 2005) related strongly to these indices.

Moving beyond self-report methodology can reduce some issues, though not all, concerning cross-group equivalency in measurement. For example, comparisons of murder rates across groups are influenced by multiple factors, including whether or not groups differ in the number of murders, whether the same act is classified as “murder” in one group versus “manslaughter” or “negligence” in another group, and whether groups differ in the proportion of murders that are reported and/or detected by authorities. Findings of group differences in murder rates could indicate actual differences in violence between groups, or they could reflect other factors, such as group differences in funding of police units investigating murders or differences in the criteria by which a death is categorized as a murder. As another example, Wolff et al. (2011) recently showed that, owing to a variety of errors, some estimates—such as Gini coefficients—based on values of the Human Development Index deviate from the true value by up to 100%. They also found that as much as 34% of the countries were misclassified in this widely used UN source (Anand and Sen 1994; United Nations Development Programme 2011). Finally, in anthropology there have been long-standing debates on whether, when increasing the scale of comparison (e.g., moving from local comparisons within a region to comparisons across regions), some traits can still be compared or whether this amounts to making invalid comparisons (e.g., Holý 1987; Moore 1994; White et al. 1988). For example, bridewealth payments might occur in both New Guinea and East Africa, but they can serve vastly different functions (Comaroff 1980) and differ to such an extent that they should not be interpreted as measuring the same phenomenon. Therefore, comparisons *within* a region might be necessary and more valuable in understanding the coevolution of traits than global comparisons (see Eggan 1954; Hammel 1980), simply because otherwise one could run the risk of “comparing apples with oranges.” Although such focused analyses may have limited generalizability, they might provide a better test for the coevolution of certain traits.

In conclusion, many studies examining cross-cultural correlations draw upon secondary data (data collected by another research team). To attenuate challenges to inference caused by non-equivalence in measurement, researchers should be careful to evaluate (and report) evidence for the validity of measures before assuming cross-cultural equivalence. Such evidence can include techniques modeling response differences (e.g., King et al. 2003) or measurement invariance (e.g., via Structural Equation Modeling: MacCallum and Austin 2000) across groups. For example via Structural Equation Modeling we can compare a model in which a latent variable is considered to be invariant across different cultures with a model in which it is allowed to vary (e.g., Byrne 2009; Byrne and Campbell 1999). For a more in-depth review, we refer to other work on cross-cultural measurement (e.g., Hui and Triandis 1985; Van de Vijver and Leung 1997, 2000).

## Conclusion

There is great merit in cross-cultural work in evolution and human behavior. In addition to recent work focusing on comparing individuals from different populations (e.g., Little et al. 2007; Stephen et al. 2012; Tovée et al. 2006), there has been a surge in work testing hypotheses via cross-cultural correlations. These studies have several merits

(e.g., moving beyond unrepresentative, Western populations; Henrich et al. 2010), but they also have several limitations. Many recent examples in the human evolutionary behavioral sciences, described in the introduction, demonstrate the temptation to interpret these group-level data as informative about individual-level adaptation (e.g., the effect of parasite stress on life-history development). However, such inferences are inherently limited in part by the issues discussed here: the ecological fallacy, non-independence of observations, and cross-cultural non-equivalence of measurement. Many of the issues we discussed apply to a broad range of research areas where cross-cultural correlations are presented as evidence for individual-level hypotheses—for instance, research on sex ratios (Dama 2011), parent-offspring conflict (Apostolou 2010), or intelligence (Kanazawa 2006). We hope that novel hypotheses derived from evolutionary theory continue to be tested with increasingly sophisticated and robust methods, for example by comparing individuals from different cultures (e.g., Stephen et al. 2012), by combining many individual-level samples from small-scale societies (e.g., Borgerhoff Mulder et al. 2009), or by modeling techniques (e.g., Frankenhuis et al. 2013). Such approaches could allow researchers to make better inferences on human nature than those based on cross-cultural correlations.

**Acknowledgments** We thank Daniel Nettle, two anonymous reviewers, and Monique Borgerhoff Mulder for providing us with helpful and extensive comments to improve earlier versions of this manuscript. Thomas Pollet is supported by The Netherlands Organisation for Scientific Research (Veni, 451.10.032)

## References

- Anand, S., & Sen, A. (1994). *Human Development Index: methodology and measurement*. Human Development Occasional Paper No. 2. Oxford University Press on behalf of the United Nations Development Program. Available online at <http://hdr.undp.org/sites/default/files/oc12.pdf>
- Anderson, K. G. (2010). Life expectancy and the timing of life history events in developing countries. *Human Nature*, 21(2), 103–123. doi:10.1007/s12110-010-9087-z.
- Apostolou, M. (2010). Sexual selection under parental choice in agropastoral societies. *Evolution and Human Behavior*, 31(1), 39–47. doi:10.1016/j.evolhumbehav.2009.06.010.
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150–166. doi:10.1177/1088868306294907.
- Ashton, M. C., & Lee, K. (2009). The HEXACO–60: a short measure of the major dimensions of personality. *Journal of Personality Assessment*, 91(4), 340–345. doi:10.1080/00223890902935878.
- Barber, N. (2000). The sex ratio as a predictor of cross-national variation in violent crime. *Cross Cultural Research*, 34(3), 264–282. doi:10.1177/106939710003400304.
- Barber, N. (2003). Paternal investment prospects and cross-national differences in single parenthood. *Cross Cultural Research*, 37(2), 163–177. doi:10.1177/1069397103037002001.
- Barthes, J., Godelle, B., & Raymond, M. (2013). Human social stratification and hypergyny: toward an understanding of male homosexual preference. *Evolution and Human Behavior*, 34(3), 155–163. doi:10.1016/j.evolhumbehav.2013.01.001.
- Borgerhoff Mulder, M. (2001). Using phylogenetically based comparative methods in anthropology: more questions than answers. *Evolutionary Anthropology Issues News and Reviews*, 10(3), 99–111. doi:10.1002/evan.1020.
- Borgerhoff Mulder, M., Nunn, C. L., & Towner, M. C. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology Issues News and Reviews*, 15(2), 52–64. doi:10.1002/evan.20088.
- Borgerhoff Mulder, M., Bowles, S., Hertz, T., Bell, A., Beise, J., Clark, G., & Hooper, P. L. (2009). Intergenerational wealth transmission and the dynamics of inequality in small-scale societies. *Science*, 326(5953), 682–688. doi:10.1126/science.1178336.

- Bosker, R. J., Kremers, E. J. J., & Lugthart, E. (1990). School and instruction effects on mathematics achievement. *School Effectiveness and School Improvement*, 1(4), 233–248. doi:10.1080/0924345900010401.
- Boyce, M. S., & McDonald, L. L. (1999). Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, 14(7), 268–272. doi:10.1016/S0169-5347(99)01593-1.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Brand, A., & Bradley, M. T. (2012). More voodoo correlations: when average-based measures inflate correlations. *Journal of General Psychology*, 139(4), 260–272. doi:10.1080/00221309.2012.703711.
- Brand, A., Bradley, M. T., Best, L. A., & Stoica, G. (2010). Multiple trials may yield exaggerated effect size estimates. *Journal of General Psychology*, 138(1), 1–11. doi:10.1080/00221309.2010.520360.
- Brashares, J. S., Arcese, P., Sam, M. K., Coppolillo, P. B., Sinclair, A. R. E., & Balmford, A. (2004). Bushmeat hunting, wildlife declines, and fish supply in West Africa. *Science*, 306(5699), 1180–1183. doi:10.1126/science.1102425.
- Brown, J. L., Morales, V., & Summers, K. (2009). Home range size and location in relation to reproductive resources in poison frogs (Dendrobatidae): a Monte Carlo approach using GIS data. *Animal Behaviour*, 77(2), 547–554. doi:10.1016/j.anbehav.2008.10.002.
- Burger, J., Kirchner, M., Bramanti, B., Haak, W., & Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences*, 104(10), 3736–3741. doi:10.1073/pnas.0607187104.
- Byrne, B. M. (2009). *Structural equation modeling with AMOS: Basic Concepts, Applications, and Programming*. London, UK: Routledge.
- Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure. *Journal of Cross-Cultural Psychology*, 30(5), 555–574. doi:10.1177/0022022199030005001.
- Cagliani, R., Guerini, F. R., Fumagalli, M., Riva, S., Agliardi, C., Galimberti, D., & Fenoglio, C. (2012). A trans-specific polymorphism in ZC3HAV1 is maintained by long-standing balancing selection and may confer susceptibility to multiple sclerosis. *Molecular Biology and Evolution*, 29(6), 1599–1613. doi:10.1093/molbev/mss002.
- Cashdan, E., & Steele, M. (2013). Pathogen prevalence, group bias, and collectivism in the standard cross-cultural sample. *Human Nature*, 24, 59–75. doi:10.1007/s12110-012-9159-3.
- Cavusgil, S. T., & Das, A. (1997). Methodological issues in empirical cross-cultural research: a survey of the management literature and a framework. *Management International Review*, 37(1), 71–96.
- Chang, K.-T. (2003). *Introduction to geographic information systems*. New York: McGraw-Hill.
- Comaroff, J. L. (1980). *The meaning of marriage payments*. London: Academic.
- Connolly, P. (2006). Summary statistics, educational achievement gaps and the ecological fallacy. *Oxford Review of Education*, 32(02), 235–252.
- Costa, P. T., & MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional manual*. Odessa: Psychological Assessment Resources.
- Currie, T. E., & Mace, R. (2012). Analyses do not support the parasite-stress theory of human sociality. *Behavioral and Brain Sciences*, 35(2), 83–85. doi:10.1017/S0140525X11000963.
- Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., & Mace, R. (2010). Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature*, 467(7317), 801–804. doi:10.1038/nature09461.
- Dahl, E., Ivar Elstad, J., Hofoss, D., & Martin-Mollard, M. (2006). For whom is income inequality most harmful? A multi-level analysis of income inequality and mortality in Norway. *Social Science and Medicine*, 63(10), 2562–2574. doi:10.1016/j.socscimed.2006.06.002.
- Dama, M. S. (2011). Sex ratio at birth and mortality rates are negatively related in humans. *PLoS One*, 6(8), e23792. doi:10.1371/journal.pone.0023792.
- Dama, M. S. (2012). Parasite stress predicts offspring sex ratio. *PLoS One*, 7(9), e46169. doi:10.1371/journal.pone.0046169.
- Dama, M. S. (2013). Cognitive ability correlates positively with son birth and predicts cross-cultural variation of the offspring sex ratio. *Naturwissenschaften*, 100(6), 559–569. doi:10.1007/s00114-013-1052-3.
- DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L. M., & Little, A. C. (2010). The health of a nation predicts their mate preferences: cross-cultural variation in women's preferences for masculinized male faces. *Proceedings of the Royal Society B: Biological Sciences*, 277(1692), 2405–2410. doi:10.1098/rspb.2009.2184.
- DeBruine, L. M., Jones, B. C., Little, A. C., Crawford, J. R., & Welling, L. L. M. (2011). Further evidence for regional variation in women's masculinity preferences. *Proceedings of the Royal Society B: Biological Sciences*, 278(1707), 813–814. doi:10.1098/rspb.2010.2200.

- DEFRA. (2011). Family food datasets. Retrieved June 6, 2012, from <http://www.defra.gov.uk/statistics/foodfarm/food/familyfood/datasets/>.
- Dow, M. M. (2007). Galton's Problem as multiple network autocorrelation effects cultural trait transmission and ecological constraint. *Cross Cultural Research*, 41(4), 336–363. doi:10.1177/1069397107305452.
- Dow, M. M., & Eff, E. A. (2008). Global, regional, and local network autocorrelation in the standard cross-cultural sample. *Cross Cultural Research*, 42(2), 148–171. doi:10.1177/1069397107311186.
- Duncan, L. A., Schaller, M., & Park, J. H. (2009). Perceived vulnerability to disease: development and validation of a 15-item self-report instrument. *Personality and Individual Differences*, 47(6), 541–546. doi:10.1016/j.paid.2009.05.001.
- Dutton, D. G. (1994). Patriarchy and wife assault: the ecological fallacy. *Violence and Victims*, 9(2), 167–182.
- Eff, E. A. (2004). Does Mr. Galton still have a problem? Autocorrelation in the standard cross-cultural sample. *World Cultures*, 15(2), 153–170.
- Eggan, F. (1954). Social anthropology and the method of controlled comparison. *American Anthropologist*, 56(5), 743–763. doi:10.1525/aa.1954.56.5.02a00020.
- Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), 3801–3808. doi:10.1098/rspb.2010.0973.
- Eppig, C., Fincher, C. L., & Thornhill, R. (2011). Parasite prevalence and the distribution of intelligence among the states of the USA. *Intelligence*, 39(3), 155–160. doi:10.1016/j.intell.2011.02.008.
- Fincher, C. L., & Thornhill, R. (2012). Parasite-stress promotes in-group assortative sociality: the cases of strong family ties and heightened religiosity. *Behavioral and Brain Sciences*, 35(2), 65–79. doi:10.1017/S0140525X11000021.
- Fincher, C. L., Thornhill, R., Murray, D. R., & Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biological Sciences*, 275(1640), 1279–1285. doi:10.1098/rspb.2008.0094.
- Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, 43(4), 557–572. doi:10.2307/2094779.
- Frankenhuis, W. E., & Panchanathan, K. (2011). Balancing sampling and specialization: an adaptationist model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, 278(1724), 3558–3565. doi:10.1098/rspb.2011.0055.
- Frankenhuis, W. E., Panchanathan, K., & Clark Barrett, H. (2013). Bridging developmental systems theory and evolutionary psychology using dynamic optimization. *Developmental Science*, 16(4), 584–98. doi:10.1111/desc.12053.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Gelman, A., Shor, B., Bafumi, J., & Park, D. (2007). Rich state, poor state, red state, blue state: what's the matter with Connecticut? *Quarterly Journal of Political Science*, 2(4), 345–367. doi:10.1561/100.00006026.
- Golding, J., Pembrey, M., & Jones, R. (2001). ALSPAC—the Avon Longitudinal Study of Parents and Children. *Paediatric and Perinatal Epidemiology*, 15(1), 74–87. doi:10.1046/j.1365-3016.2001.00325.x.
- Gravelle, H. (1998). How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *British Medical Journal*, 316(7128), 382–385. doi:10.1136/bmj.316.7128.382.
- Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: a life history theory approach. *Journal of Personality and Social Psychology*, 100(6), 1015–1026. doi:10.1037/a0022403.
- Hackman, J., & Hruschka, D. (2013). Fast life histories, not pathogens, account for state-level variation in homicide, child maltreatment, and family ties in the U.S. *Evolution and Human Behavior*, 34(2), 118–124. doi:10.1016/j.evolhumbehav.2012.11.002.
- Hammel, E. A. (1980). The comparative method in anthropological perspective. *Comparative Studies in Society and History*, 22(2), 145–155. doi:10.2307/178402.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61(1), 139–151. doi:10.1006/anbe.2000.1518.
- Harvey, P. H., Read, A. F., & Nee, S. (1995). Why ecologists need to be phylogenetically challenged. *Journal of Ecology*, 83(3), 535–536.
- Hawkes, K. (1994). On life-history evolution. *Current Anthropology*, 35(1), 39–41. doi:10.2307/2744136.
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903–918. doi:10.1037/0022-3514.82.6.903.



- Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? *Psychological Science*, *19*(4), 309–313. doi:10.1111/j.1467-9280.2008.02085.x.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, *33*(2–3), 61–83. doi:10.1017/S0140525X0999152X.
- Herkenrath, M. (2002). Quantitative cross-national analysis as a research tool in the sociology of developing countries: a critical examination. *Current Sociology*, *50*(4), 517–530. doi:10.1177/0011392102050004003.
- Hill, K. R., Walker, R. S., Božičević, M., Eder, J., Headland, T., Hewlett, B., & Wood, B. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, *331*(6022), 1286–1289. doi:10.1126/science.1199071.
- Holý, L. (1987). *Comparative anthropology*. (L. Holý, Ed.). Oxford: Blackwell.
- Howell, D. C. (2010). *Fundamental statistics for the behavioral sciences*. Belmont: Wadsworth.
- Hox, J. J. (2010). *Multilevel analysis: techniques and applications*. London: Taylor & Francis.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, *16*(2), 131–152. doi:10.1177/0022002185016002001.
- Huisman, M., Oldehinkel, A. J., De Winter, A., Minderaa, R. B., De Bildt, A., Huizink, A. C., & Ormel, J. (2008). Cohort profile: the Dutch “TRacking Adolescents’ Individual Lives Survey” (TRAILS). *International Journal of Epidemiology*, *37*(6), 1227–1235. doi:10.1093/ije/dym273.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: history, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: theory and research* (2nd ed., Vol. 2, pp. 102–138). New York: Guilford Press.
- Kanazawa, S. (2006). IQ and the wealth of states. *Intelligence*, *34*(6), 593–600. doi:10.1016/j.intell.2006.04.003.
- Kievit, R. A., Frankenhuys, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson’s Paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*, 513. doi:10.3389/fpsyg.2013.00513.
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2003). Enhancing the validity and cross-cultural comparability of measurement in survey research. *The American Political Science Review*, *97*(4), 567–583. doi:10.1017/S000305540400108X.
- Kruger, D. J., & Nesse, R. M. (2006). An evolutionary life-history framework for understanding sex differences in human mortality rates. *Human Nature*, *17*(1), 74–97. doi:10.1007/s12110-006-1021-z.
- Kruger, D. J., & Schlemmer, E. (2009). When men are scarce, good men are even harder to find: life history, the sex ratio, and the proportion of men married. *Journal of Social Evolutionary and Cultural Psychology*, *3*(2), 93–104.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. doi:10.1207/s15327906mbr3902\_8.
- Little, A. C., Apicella, C. L., & Marlowe, F. W. (2007). Preferences for symmetry in human faces in two cultures: data from the UK and the Hadza, an isolated group of hunter-gatherers. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1629), 3113–3117. doi:10.1098/rspb.2007.0895.
- Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: when predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, *2*(2), 363. doi:10.1037/1076-8971.2.2.363.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201–226. doi:10.1146/annurev.psych.51.1.201.
- Mace, R., & Pagel, M. (1994). The comparative method in anthropology. *Current Anthropology*, *35*(5), 549–564.
- Mace, R., Jordan, F., & Holden, C. (2003). Testing evolutionary hypotheses about human biological adaptation using cross-cultural comparison. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, *136*(1), 85–94. doi:10.1016/S1095-6433(03)00019-9.
- Marlowe, F. W. (2003). The mating system of foragers in the standard cross-cultural sample. *Cross Cultural Research*, *37*(3), 282–306. doi:10.1177/1069397103254008.
- Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., & Tracer, D. (2008). More “altruistic” punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1634), 587–592. doi:10.1098/rspb.2007.1517.
- Marlowe, F. W., Berbesque, J. C., Barrett, C., Bolyanatz, A., Gurven, M., & Tracer, D. (2011). The “spiteful” origins of human cooperation. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1715), 2159–2164. doi:10.1098/rspb.2010.2342.
- McCormick, C. M. (2013). Watch where and how you stick pins when playing with voodoo correlations. *Journal of General Psychology*, *140*(1), 82–86.
- McCulloch, C. E., & Neuhaus, J. M. (2001). *Generalized linear mixed models*. New York: Wiley Online Library.

- Menzel, H. (1950). Comment on Robinson's "Ecological correlations and the behavior of individuals." *American Sociological Review*, 15, 674.
- Mgawe, P., Borgerhoff Mulder, M., Caro, T., Martin, A., & Kiffner, C. (2012). Factors affecting bushmeat consumption in the Katavi-Rukwa ecosystem of Tanzania. *Tropical Conservation Science*, 5(4), 446–462.
- Moore, J. H. (1994). Putting anthropology back together again: the ethnogenetic critique of cladistic theory. *American Anthropologist*, 96(4), 925–948. doi:10.2307/682452.
- Moore, F. R., Coetzee, V., Contreras-Garduño, J., Debruine, L. M., Kleisner, K., Krams, I., ... Suzuki, T. N. (2013). Cross-cultural variation in women's preferences for cues to sex- and stress-hormones in the male face. *Biology Letters*, 9(3). doi:10.1098/rsbl.2013.0050.
- Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. *Ethnology*, 8(4), 329–369. doi:10.2307/3772907.
- Naroll, R. (1965). Galton's Problem: the logic of cross-cultural analysis. *Social Research*, 32(4), 428–451.
- Nasif, E. G., Al-Daeaj, H., Ebrahimi, B., & Thibodeaux, M. S. (1991). Methodological problems in cross-cultural research: an updated review. *Management International Review*, 31(1), 79–91.
- National Child Development Study. (n.d.). National Child Development Study. Retrieved June 06, 2012, from <http://www.esds.ac.uk/longitudinal/access/ncds/133004.asp>.
- Nee, S., Read, A. F., Greenwood, J. J. D., & Harvey, P. H. (1991). The relationship between abundance and body size in British birds. *Nature*, 351(6324), 312–313. doi:10.1038/351312a0.
- Nee, S., Read, A. F., & Harvey, P. H. (1996). Why phylogenies are necessary for comparative analysis. In E. P. Martins (Ed.), *Phylogenies and the comparative method in animal behavior* (pp. 399–411). Oxford: Oxford University Press.
- Nettle, D. (2009). Ecological influences on human behavioural diversity: a review of recent findings. *Trends in Ecology & Evolution*, 24(11), 618–624. doi:10.1016/j.tree.2009.05.013.
- Piantadosi, S., Byar, D. P., & Green, S. B. (1988). The ecological fallacy. *American Journal of Epidemiology*, 127(5), 893–904.
- Plewis, I., Ketende, S., & Millennium Cohort Study. (2006). *Millennium Cohort Study*. Technical Report on Responses. Retrieved from <http://www.esds.ac.uk/longitudinal/access/mcs/133359.asp>.
- Poortinga, Y. H. (1989). Equivalence of cross-cultural data: an overview of basic issues. *International Journal of Psychology*, 24(1), 737–756. doi:10.1080/00207598908247842.
- Powell, A., Shennan, S., & Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science*, 324(5932), 1298–1301. doi:10.1126/science.1170165.
- Quinlan, R., & Quinlan, M. (2008). Human lactation, pair-bonds, and alloparents. *Human Nature*, 19(1), 87–102. doi:10.1007/s12110-007-9026-9.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London: Sage.
- Raven, J. (2000). Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. doi:10.1006/cogp.1999.0735.
- Rehkopf, D. H., & Buka, S. L. (2006). The association between suicide and the socio-economic characteristics of geographical areas: a systematic review. *Psychological Medicine*, 36(02), 145–157.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357.
- Ross, M. H., & Homer, E. (1976). Galton's Problem in cross-national research. *World Politics*, 29(1), 1–28. doi:10.2307/2010045.
- Rushton, S. P., Lurz, P. W. W., Fuller, R., & Garson, P. J. (1997). Modelling the distribution of the red and grey squirrel at the landscape scale: a combined GIS and population dynamics approach. *Journal of Applied Ecology*, 34(5), 1137–1154.
- Schaller, M., & Murray, D. R. (2008). Pathogens, personality, and culture: disease prevalence predicts worldwide variability in sociosexuality, extraversion, and openness to experience. *Journal of Personality and Social Psychology*, 95(1), 212–221. doi:10.1037/0022-3514.95.1.212.
- Scheiner, S. M., Cox, S. B., Willig, M., Mittelbach, G. G., Osenberg, C., & Kaspari, M. (2000). Species richness, species-area curves and Simpson's Paradox. *Evolutionary Ecology Research*, 2(6), 791–802. doi:10.1152/ajpendo.00279.2010.
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, 94(1), 168–182. doi:10.1037/0022-3514.94.1.168.
- Schwartz, S. (1994). The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *American Journal of Public Health*, 84(5), 819–824.
- Sheppard, L. (2003). Insights on bias and information in group-level studies. *Biostatistics*, 4(2), 265–278. doi:10.1093/biostatistics/4.2.265.

- Shrira, I., Wisman, A., & Webster, G. D. (2013). Guns, germs, and stealing: exploring the link between infectious disease and crime. *Evolutionary Psychology*, *11*(1), 270–287.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B (Methodological)*, *13*(2), 238–241. doi:10.1038/168063d0.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Stephen, I. D., Scott, I. M. L., Coetzee, V., Pound, N., Perrett, D. I., & Penton-Voak, I. S. (2012). Cross-cultural effects of color, but not morphological masculinity, on perceived attractiveness of men's faces. *Evolution and Human Behavior*, *33*(4), 260–267. doi:10.1016/j.evolhumbehav.2011.10.003.
- Subramanian, S. V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting Robinson: the perils of individualistic and ecologic fallacy. *International Journal of Epidemiology*, *38*(2), 342–360. doi:10.1093/ije/dyn359.
- National Longitudinal Survey of Youth. (2007). National Longitudinal Survey of Youth, 1997. Inter-university Consortium for Political and Social Research (ICPSR). doi:10.3886/ICPSR03959.v2.
- Te Grotenhuis, M., Eisinga, R., & Subramanian, S. V. (2011). Robinson's ecological correlations and the behavior of individuals: methodological corrections. *International Journal of Epidemiology*, *40*(4), 1123–1125. doi:10.1093/ije/dyn357.
- Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovova, L., Ahn, C., Ahn, H., & Angleitner, A. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science*, *310*(5745), 96–100. doi:10.1126/science.1117199.
- Thomas, F., Daoust, S. P., Elguero, E., & Raymond, M. (2013). Do mothers from rich and well-nourished countries bear more sons? *Journal of Evolutionary Medicine*, *1*, 1–11. doi:10.4303/jem/235568.
- Thorndike, E. L. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *American Journal of Psychology*, *52*(1), 122–124.
- Thornhill, R., & Fincher, C. L. (2011). Parasite stress promotes homicide and child maltreatment. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *366*(1583), 3466–3477. doi:10.1098/rstb.2011.0052.
- Tovée, M. J., Swami, V., Furnham, A., & Mangalparsad, R. (2006). Changing perceptions of attractiveness as observers are exposed to a different culture. *Evolution and Human Behavior*, *27*(6), 443–456. doi:10.1016/j.evolhumbehav.2006.05.004.
- Towner, M. C., Grote, M. N., Venti, J., & Borgerhoff Mulder, M. (2012). Cultural macroevolution on neighbor graphs. *Human Nature*, *23*(3), 283–305.
- Turchin, P., Currie, T. E., Turner, E. A. L., & Gavrilets, S. (2013). War, space, and the evolution of Old World complex societies. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16384–16389. doi:10.1073/pnas.1308825110.
- Tybur, J. M., & De Vries, R. E. (2013). Disgust sensitivity and the HEXACO model of personality. *Personality and Individual Differences*, *55*(6), 660–665. doi:10.1016/j.paid.2013.05.008.
- Tybur, J. M., Bryan, A. D., Lieberman, D., Caldwell Hooper, A. E., & Merriman, L. A. (2011). Sex differences and sex similarities in disgust sensitivity. *Personality and Individual Differences*, *51*(3), 343–348. doi:10.1016/j.paid.2011.04.003.
- United Nations Development Programme (UNDP). (2011). Human Development Index (HDI) data. <http://hdr.undp.org/en/statistics/hdi>. Accessed November 2011.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park: Sage.
- Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, *31*(1), 33–51. doi:10.1177/0022022200031001004.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *54*(2), 119–135. doi:10.1016/j.erap.2003.12.004.
- Vleeschbarometer. (2011). Thuisverbruik (Home Consumption). Retrieved June 06, 2012, from <http://www.vlam.be/marketinformatiedocument/files/Thuisverbruik2010.pdf>.
- Von Korff, M., Koepsell, T., Curry, S., & Diehr, P. (1992). Multi-level analysis in epidemiologic research on health behaviors and outcomes. *American Journal of Epidemiology*, *135*(10), 1077–1082.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*(3), 274–290. doi:10.1111/j.1745-6924.2009.01125.x.
- Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data*. New York: Wiley.
- Wasserman, E. A. (1993). Comparative cognition: beginning the second century of the study of animal intelligence. *Psychological Bulletin*, *113*(2), 211–228. doi:10.1037/0033-2909.113.2.211.

- Wechsler, D. (1981). *WAIS-R manual: wechsler adult intelligence scale-revised*. San Antonio: Psychological Corporation.
- Wells, J. C. K. (2012). Sexual dimorphism in body composition across human populations: associations with climate and proxies for short- and long-term energy supply. *American Journal of Human Biology*, 24(4), 411–419. doi:10.1002/ajhb.22223.
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.
- White, D. R., Betzig, L., Mulder, M. B., Chick, G., Hartung, J., Irons, W., & Spencer, P. (1988). Rethinking polygyny: co-wives, codes, and cultural systems [and comments and reply]. *Current Anthropology*, 29(4), 529–572.
- Wilkie, D. S., Starkey, M., Abernethy, K., Effa, E. N., Telfer, P., & Godoy, R. (2005). Role of prices and wealth in consumer demand for bushmeat in Gabon, Central Africa. *Conservation Biology*, 19(1), 268–274. doi:10.1111/j.1523-1739.2005.00372.x.
- Wilkin, T. A., Garant, D., Gosler, A. G., & Sheldon, B. C. (2006). Density effects on life-history traits in a wild population of the great tit *Parus major*: analyses of long-term data with GIS techniques. *Journal of Animal Ecology*, 75(2), 604–615. doi:10.1111/j.1365-2656.2006.01078.x.
- Winkleby, M., Cubbin, C., & Ahn, D. (2006). Effect of cross-level interaction between individual and neighborhood socioeconomic status on adult mortality rates. *American Journal of Public Health*, 96(12), 2145–2153. doi:10.2105/AJPH.2004.060970.
- Wolff, H., Chong, H., & Auffhammer, M. (2011). Classification, detection and consequences of data error: evidence from the human development index. *The Economic Journal*, 121(553), 843–847. doi:10.1111/j.1468-0297.2010.02408.x.
- Yip, P. S. F., & Liu, K. Y. (2006). The ecological fallacy and the gender ratio of suicide in China. *British Journal of Psychiatry*, 189(5), 465–466.
- York, R., & Gossard, M. H. (2004). Cross-national meat and fish consumption: exploring the effects of modernization and ecological context. *Ecological Economics*, 48(3), 293–302. doi:10.1016/j.ecolecon.2003.10.009.

**Thomas Pollet** is an assistant professor in the Department of Social and Organizational Psychology at VU University Amsterdam. He obtained his PhD in psychology from Newcastle University (UK) in 2008. He has previously worked at the University of Groningen. His research covers a broad range of topics in the field of evolution and human behavior such as kinship, individual differences, and mating. He currently studies aspects of the evolution of stature.

**Joshua Tybur** is an assistant professor in the Department of Social and Organizational Psychology at VU University in Amsterdam. He received his PhD in evolutionary psychology from the University of New Mexico in 2009. His research is currently focused on understanding some of the puzzles posed by disgust, and how the solutions to these puzzles can inform our understanding of food preferences and aversions, mating psychology, and moral judgment.

**Willem Frankenhuys** is an assistant professor in developmental psychology at Radboud University Nijmegen in the Netherlands. He obtained his PhD in 2013 in biological anthropology at the University of California, Los Angeles. His empirical work examines cognitive development in harsh environments. His current studies test the hypothesis that people who live in stressful conditions developmentally tailor, or specialize, their mental abilities for solving problems that are ecologically relevant to them (e.g., detecting and memorizing threats). His theoretical work examines how natural selection shapes developmental mechanisms and how these mechanisms produce phenotypic variability, depending on ecology and internal state. In addition, he uses statistical decision theory and stochastic dynamic programming to better understand tradeoffs during development.

**Ian Rickard** is a lecturer in the Department of Anthropology at the University of Durham, UK. He has a broad range of professional interests, but his largest research focus is at the intersection of evolutionary biology, ecology and health. He is particularly interested in the question of how evolutionary life-history theory can help us understand, predict and remedy health disparities within and between contemporary human populations.