THE JOURNAL OF PHILOSOPHY
VOLUME CXVIII, NO. 10, OCTOBER 2021

DIVERSE POPULATIONS ARE CONFLATED WITH HETEROGENEOUS COLLECTIVES*

The concept of difference has a long and important tradition in philosophical research,[1] and noticing differences is fundamental for building our most basic categories,[2] social systems,[3] models,[4] and their causal explanations.[5] Given this context, our goal

* We are deeply grateful for the insightful advice of the anonymous reviewers—including regarding this article's title—and the generous help from our colleagues and students: Anat Kolumbus and Yael Silver for their ideas and data; Eli Pitkovski, David Enoch, Yochai Ataria, and Tamar Hager for crucial advice; and the participants of the Tel Hai Seminar, "Involved Philosophy" Workshops, and PSA 2016 conference.

[1] Space limitations and the content of our question prevent engagement with this important history, which includes Leibniz's Principle of Identity of Indiscernibles ("there are never two things in nature which are exactly alike and in which it is impossible to find a difference" (Gottfried Wilhelm Leibniz, *Discourse on Metaphysics and the Monadology*, trans. George R. Montgomery (Amherst, NY: Prometheus Books, [1686] 1992), p. 608)) and subsequent discussions by Kant, Strawson, Hacking, and more recently by Charles B. Cross ("Causal Independence, the Identity of Indiscernibles, and the Essentiality of Origins," this JOURNAL, CVI, 5 (May 2009): 277–91).

[2] "Basic categories are those which carry the most information. . .and are, thus, the most differentiated from one another." See Eleanor Rosch et al., "Basic Objects in Natural Categories," *Cognitive Psychology*, VIII, 3 (July 1976): 382–439, at p. 382.

[3] Philip Kitcher clarified why foregrounding differences is one major feature that separates democratic from totalitarian regimes and science from other knowledge-production practices. See Philip Kitcher, *Science in a Democratic Society* (Amherst, NY: Prometheus Books, 2011).

[4] Within the large body of literature on models, we use Michael Weisberg's definition of models as incomplete, idealized, and abstract representations, specified by descriptions such as "words, pictures, equations, diagrams or computer programs and are accompanied by legends" (Michael Weisberg, *Simulation and Similarity: Using Models to Understand the World* (Oxford: Oxford University Press 2013), p. 45) and interpretations, which "set up relations of denotation between the model and the real-world target, and. . . give criteria for evaluating the goodness of fit between a model and a target" (*ibid.*, p. 39). A model's description constitutes its causal explanation and the interpretation to its prediction.

[5] A model's casual explanation is a "story about why that phenomenon occurred. . . . A causal factor makes a difference to a phenomenon just in case its removal from a

0022-362X/21/1810/525–548

is to invite a discussion on a new set of "big" questions—not answer them—by noticing an important but heretofore overlooked distinction in the meaning and practice of measuring difference.[6]

We explicate and formalize two different meanings of 'difference'—'diversity' and 'heterogeneity'—and argue that 'diversity' can describe a population well enough but does not describe a collective well, whereas 'heterogeneity' better describes the latter and therefore ought to describe it. A population is any subset within a larger group[7] with no prerequisite of any joint process or structure, whereas exactly such joint aspects characterize a collective group.[8] We argue that such different kinds of groups are better described by different measurements of difference, and that ignoring these distinctions sometimes leads to a surprising and disturbing conflict between diversity and heterogeneity.

In a nutshell, the statement "a zoo is diverse whereas an ecosystem[9] is heterogeneous" encapsulates our distinction. The same number of species, populations, and individual animals could exist in the same spatial and temporal proximity in either a zoo or an ecosystem. But a zoo is an aggregated collection of non-interacting populations, whereas populations within an ecosystem interact within a collective.

---

causal model prevents the model from entailing the phenomenon's occurrence." This definition of causality, attributed by Weisberg to Michael Strevens, helps clarify the model framework. See *ibid.*, p. 101.

[6] We thank an anonymous reviewer for this framing.

[7] According to the Oxford English Dictionary (*OED Online* (Oxford: Oxford University Press, 2021), https://www.oed.com/Entry/147922, accessed March 24, 2021), the extended use of 'population' is "A group of people, esp. regarded as a class or subset within a larger group." For this article, the relevant technical uses of the term are from statistics and biology. For statistics a group is "a (real or hypothetical) totality of objects or individuals under consideration, of which the statistical attributes may be estimated by the study of a sample or samples drawn from it." For biology it is "a group of animals, plants, or humans, within which breeding occurs," so the concept of population requires no mention of inner or outer structure.

[8] Using the term 'collective' presumes a non-aggregative entity, holding some joint processes and a common structure. James Griesemer, "Landscapes of Developmental Collectivity," in Snait B. Gissis, Ehud Lamm, and Ayelet Shavit, eds., *Landscapes of Collectivity in the Life Sciences* (Cambridge, MA: The MIT Press, 2017), pp. 25–48, at p. 32, captures this point for humans and animals alike: "In a collective, the items collected for landscape representation seem somehow to *belong* together." When discussing the trait of group collectivity, Griesemer elaborates that "collectivity seems to be the concept we fall back on when our sense of individuality falters or when we try to comprehend processes by which heterogeneous individuals may gather (or be gathered) together to form a group that is somehow 'more' than an aggregate" (*ibid.*, p. 31).

[9] "An ecosystem consists of a biological community, its physical and chemical environment, and the dynamic interactions that link them" (A. K. Salomon, "Ecosystems," in S. E. Jørgensen and B. D. Fath, eds., *Encyclopedia of Ecology*, vol. 2 (Boston: Elsevier, 2008), pp. 1155–65).

The two types of difference are neither mathematically interchangeable nor empirically correlated with one another (see APPENDIX), and describing an ecosystem's populations or species (that is, its diversity) without tracking the structure of interactions between them (for example, hierarchical relationships in an organization or who-eats-whom relationships in a natural food web) will lead inevitably to failures in managing it. As a general rule, sufficing with measures of 'diversity' for models of collective phenomena (for example, epistemic, social, cultural, or biological groups) is a value-laden choice with nontrivial epistemic, moral, and environmental results. In particular, focusing on the 'diversity' of human communities can be self-defeating for those who truly care about the importance of diversity.[10] Similarly, if an ecosystem is rare or threatened, such a failure to distinguish diversity from heterogeneity may be costly.

A full analysis of the concept of difference is beyond the scope of a single article. We focus here only on 'difference' in the context of models that measure group differences for model-based policies, social and environmental alike. We argue that distinguishing 'diversity' from 'heterogeneity' clarifies other core concepts in philosophy of science—'group', 'robustness', and 'objectivity'—and that noticing 'heterogeneity' will improve model accuracy and validity. Furthermore, we argue that 'diversity' and 'heterogeneity' not only are nonequivalent, but also are sometimes conflated. To illustrate this point, we briefly look into the socio-demography of philosophy programs during the past few decades and suggest that a routine, model-based policy that focuses on academic diversity alone will plausibly reduce academic heterogeneity and thus deepen epistemic injustice while maintaining social injustice intact.

Our main argument exposes a heretofore unnoticed gap between two measures of 'difference'. We begin with an example—the usual model-based policy for reducing social inequality in the university—that reveals a surprising trade-off between diversity and heterogeneity measurements (section I). This case demonstrates the relevance and importance of explicating a yet unnoticed gap in the meaning of 'difference', both in common parlance (section II) and in mathematical formulations (section III). We note that current measures of difference, which may adequately fit diverse populations, inaptly apply to heterogeneous collectives. To fill this lacuna, we use network metrics to introduce a new "index of heterogeneity" (section IV). To

[10] We thank an anonymous reviewer for this elegant way of summarizing the moral crux of our paper.

conclude, since ‘difference’ plays a fundamental role in shaping our lives and thoughts, and since foregrounding only diversity measures in eco-social collectives is empirically linked to increasing their inner-divergence and henceforth social inequality, then explicating the distinction between ‘diversity’ and ‘heterogeneity’ and using relevant measures of difference for the appropriate contexts may be a first step toward minding—in the sense of noticing and correcting—an important gap in a very basic concept (section v).

I. AN EXAMPLE: A TRADE-OFF BETWEEN DIVERSITY AND HETEROGENEITY IN THE UNIVERSITY

One typically cares more about identifying differences regarding members of groups experiencing increased pressures or constraints so as to resist such constraints and ameliorate unjust pressures. Therefore, modeling the academic success of ethnic, racial, or gender minorities in the university not only is an epistemic and empirical question, but also is a moral one. One’s research perspective directs attention to certain phenomena and models that explain these phenomena by tracking and describing certain casual chains.[11] We argue that the default research perspective for modeling diversity in the university harms the very issue it aims to promote.

Such data-models routinely count the number and abundance of individuals within different groups—represented by self-identified people in each group—but ignore the connective *interactions* within and between these groups, their structure, and level of engagement. In so doing, the default interpretation for academic inequality neglects a major “difference-maker” cause in its model description. Because such models neglect a major factor in their causal explanations, they will produce inaccurate or biased results.

Diversity is monitored in universities because it is easy to measure, track, and model discrete, constant, and context-independent parameters. Hence, variables such as race, ethnicity, and gender are measured and related to one’s grade-point average (GPA), whereas interactive and context-dependent variables such as sense of belonging rarely if ever are measured or monitored at large scales or used for model-based policy. Even when interactive variables such as “campus climate” are finally measured, surveys do not track one’s intra- and inter-group dynamics within a joint collective but rather one’s sense

[11] James R. Griesemer, “Reproduction and the Scaffolded Development of Hybrids,” in Linnda R. Caporael, James R. Griesemer, and William C. Wimsatt, eds., *Developing Scaffolds in Evolution, Culture, and Cognition* (Cambridge, MA: The MIT Press, 2014), pp. 23–55.

of ease and safety in the university's unspecified public sphere. Since one tends to feel safer in one's local—typically homogenous—group, and since minorities constantly meet hegemonic "others," then such climate measures score higher when minorities are exposed to a "safe zone," where only like-members are invited.[12]

Such a safe refuge clearly increases one's sense of belonging to their own group, and in that sense is vital to one's well-being, but it also increases one's sense of divergence from other campus groups and reduces one's sense of belonging to the academy as a whole, followed by academic alienation and its substantial costly affects. Since universities are practical and achievement-oriented institutions, they will be rationally inclined to invest more in those changes they can more easily model, measure, and for which they can quantify improvement. As a result, measuring population diversity becomes institutionalized in academic protocols, and safe zones spread on campuses for nearly any kind of minority group.

When annual censuses of diversity yield disappointing results or miss stated targets, university leaders declare their deep commitment to affirmative action and recruiting and enrolling 'minorities'. Heads are counted identically across different populations and power structures, yet students, staff, faculty, or trustees are rarely asked about their ties with members of minority groups or about their sense of belonging to a larger whole. No management recommendations are suggested to improve *interactions* between 'majority' and 'minority' groups. Inter-group dynamics, if they occur, are not the university's responsibility but left to individual initiatives in individual classrooms, fraternities, or small-scale projects.

We assuredly do not suggest abandoning affirmative action, which has resulted in demonstrable benefits inside and outside universities. Yet only tracking and analyzing measures of diversity without regard for measures of heterogeneity—for example, sense of belonging, sense of community, inter-group collaboration over homework, teaching, grants, or papers—foregrounds the differences among the various populations on campus and backgrounds their daily academic interactions within a common collective. That is, since tracking 'diversity' on campus focuses only on tracking differences without noting their collective context—whether it holds some common ground or not—this praxis of measurement increases political divergence.[13]

[12] See experimental results by J. Katz et al., "Effect of Exposure to a Safe Zone Symbol on Perceptions of Campus Climate for Sexual Minority Students,"*Psychology of Sexual Orientation and Gender Diversity*, III, 3 (September 2016): 367–73.

[13] For details on tracking, see James R. Griesemer, "Tracking Organic Processes: Representations and Research Styles in Classical Embryology and Genetics," in Manfred D.

Critical social awareness for differences is clearly of value, and a first important step toward increasing equality in our educational system,[14] yet focusing on difference and diversity has been shown to be associated with increased alienation from other groups and from one's collective (which, in the case of minority groups, is dominated by other groups). Among minority groups, increased alienation and a reduced sense of belonging to the collective (for example, the university) has been correlated significantly with lower levels of health, well-being, GPA, and college drop-out rates.[15] The lack of concern for dimensions of heterogeneity, such as 'sense of belonging,' may partly explain why most minority groups are still woefully underrepresented among philosophy faculties even though the percentage of minority undergraduates studying philosophy has increased and substantial funds have been dedicated to targeted scholarships for decades.[16]

Some may read the data in the footnotes yet still maintain that the path toward academic social justice requires only continuing to promote diversity as much as possible in order to make progress toward a more heterogeneous and just society in the future. At least in philosophy programs, it is important to note that increasing the diversity on campus has *not* functioned as a stepping stone toward increased heterogeneity and academic success of minority groups. In fact, decades of emphasis on simple measures of population diversity may have led (inadvertently) to a positive feedback between reducing campus heterogeneity and increasing its divergence. The latter reduced the sense of belonging to the university among minorities and increased their alienation, along with its disturbing significant impact on other aspects of their lives.

Laubichler and Jane Maienschein, eds., *From Embryology to Evo-Devo: A History of Developmental Evolution* (Cambridge, MA: The MIT Press, 2007), pp. 375-433.

[14] See the first educational stage toward conscientization and cultural revolution, as put forward by Paulo Freire, "Cultural Action and Conscientization," in Michael W. Apple and Wayne Au, eds., *Critical Education,* vol. 2 (New York: Routledge, 2014), pp. 5–26.

[15] Gregory M. Walton and Geoffrey L. Cohen, "A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students," *Science,* cccxxxi, 6023 (March 2011): 1447–51. See many more studies on 'sense of belonging' as a reliable indicator and a robust casual factor of minority students' well-being and achievement: Terrell L. Strayhorn, *College Students' Sense of Belonging: A Key to Educational Success for All Students* (New York: Routledge, 2019).

[16] One can follow the steady demographic graphs of college graduates in philosophy from 1995 up to the present or read about the representation of female authors in philosophy journals from 2004 to 2015, where "in all years and for all journals, the percentage of female authors was extremely low, in the range of 14–16%" in Isaac Wilhelm, Sherri Lynn Conklin, and Nicole Hassoun, "New Data on the Representation of Women in Philosophy Journals: 2004–2015," *Philosophical Studies,* clxxv, 6 (June 2018): 1441–64.

This scenario not only is a *prima facie* problem of social injustice but also represents an epistemic injustice, described by Fricker as "a wrong done to someone specifically in their capacity as a knower."[17] Both individual agents and epistemic collectives have different knowledge and research perspectives. Their lack of inclusion in various models often leads to mismatched interpretations[18] and erroneous predictions.[19] More generally, because knowledge is typically unequally distributed between all agents and all groups—some of it is known to all, some to certain groups, and some only to specific individuals[20]—when high levels of diversity lead to divergence and alienation, reduced interaction on campus keeps pieces of information latent. Diversity indices cannot capture this phenomenon, whereas heterogeneity indices specifically identify and track it. A focus on heterogeneity at the expense of diversity, at least in this important case and perhaps in other similar cases, is likely to increase shared knowledge, improve scientific models targeting academic inequality, and perhaps help change the grim picture minorities face in philosophy departments.

II. DISTINGUISHING DIVERSITY FROM HETEROGENEITY

The aforementioned example motivated this section's deeper analysis of the diversity-heterogeneity distinction. We start by explicating the conceptual difference, then examine how it helps to clarify the meaning of 'group', 'objectivity', and 'robustness', each of which is a core concept in philosophy of science with a parallel rich philosophical literature. The aim of identifying a distinction between diversity and heterogeneity is to clarify these concepts and their implications,

[17] Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (New York: Oxford University Press, 2007), p. 1.

[18] One example is a broader and more politized interpretation of 'scientific pluralism' due to repeated meetings with Allawian women. See Helen E. Longino, "Interaction: A Case for Ontological Pluralism," *Interdisciplinary Science Reviews*, XLV, 3 (October 2020): 432–45.).

[19] For example, Henrich and colleagues revealed that "95% of psychological samples come from countries with only 12% of the world's population," a sub-population they named WEIRD (Western, Educated, Industrialized, Rich, and Democratic). We thank Anat Kolumbus for this input. See more in Joseph Henrich, Steven J. Heine, and Ara Norenzayan, "The Weirdest People in the World?," *Behavioral and Brain Sciences*, XXXIII, 2–3 (June 2010): 61–83, and a wider overview in Joseph Henrich, *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous* (New York: Farrar, Straus and Giroux, 2020).

[20] See an epistemic argument in Miriam Solomon, "Norms of Epistemic Diversity," *Episteme*, III, 1–2 (June 2006): 23–36; and a sociological analysis in Elihu M. Gerson, "Integration of Specialties: An Institutional and Organizational View," *Studies in History and Philosophy of Biological and Biomedical Sciences*, XL, 4A (December 2013): 515–24.

not to reflect a universally unanimous linguistic intuition, although our distinction does agree with common usage.[21]

II.1. Diversity and Heterogeneity in Common Usage. Attributing heterogeneity to something (for example, a classroom or a multicellular organism) implies that this something is a complex coordination of interactions or structures among the different individuals or groups that jointly compose it. In contrast, attributing diversity to something (for example, paintings hanging on a wall or the people in a queue) does not imply interactions or structures or joint coordination. Thus, 'heterogeneous' only applies to a *collective*, whereas 'diverse' applies to aggregated populations or a *collection*. One could note a diverse collection of paintings on a wall, but not a diverse collection of organs in one's body.

The key distinction is that a wall is described and identified independently of the paintings hanging on it, whereas describing or identifying an organism already assumes coordinated interaction among its different organs. That is, 'diversity' refers only to the distinctions between entities, ignoring any higher-level interrelation between entities even if such ties exist, whereas 'heterogeneity' refers to entities that interact and integrate—to some degree—their differences within a larger, complex whole. Thus, the concept and measure of diversity *presupposes divergence*, not integration nor a neutral sense of difference.

II.2. Heterogeneity Is a Special Case of Diversity. 'Heterogeneous' applies only to collectives and their comprising entities that together meet three necessary conditions: (1) *difference* among the entities of which an entity is composed, (2) *interaction* among these different entities, and (3) *integration* of the different interacting entities into a complex collective structure. Diverse entities need only fulfill the first condition. Heterogeneity is thus a special case of diversity: a diverse entity *may* be heterogeneous, but a heterogeneous entity *must* also be diverse. Examining the structure and relations of the different entities within and between an examined entity is thus crucial for modeling that entity as a heterogeneous collective. Abstracting away from such empirical background presumes the targeted entity is, or can be modeled well enough as, a diverse collection.

[21] We deeply thank Anat Kolumbus for her idea for, and work of, searching in online linguistic databases, each containing over 500 million words. The frequencies of 'collective', 'whole', 'integration', and 'interaction' co-occurred significantly more with 'heterogeneous' than with 'diverse' (improved prediction by, respectively, 24, 8, 11, and 11%; chi-square tests for non-random frequencies), thus supporting our hypothesis. Corpora searched: Mark Davies, "The Corpus of Contemporary American English" (2008), https://www.english-corpora.org/coca/, accessed October 15, 2021; and Mark Davies, "The Wikipedia Corpus" (2015), https://www.english-corpora.org/wiki/, accessed October 15, 2021.

An entity's amount of diversity affects its amount of heterogeneity, but the effect need not be additive or strictly monotonically increasing.[22] In a group heterogeneity model, a minimum level of group diversity is needed for group heterogeneity to occur at all. Yet for certain group traits or goals, or within certain socioecological contexts in which the *collective* character of the group is crucial, adding diversity beyond a certain amount rarely increases, and may actually reduce, its heterogeneity. This reduction occurs because diversity disregards the causal factors on which a collective depends (that is, diversity assumes difference). Because heterogeneity depends on such factors (that is, heterogeneity assumes a collective), then for activities that can be performed only by collectives, a substantial increase in diversity may magnify within-group differences and decrease a group's integration and effective group-level organized performance. Whether or when an increase in diversity changes (reduces) the heterogeneous, collective nature of a group is an empirical question, not a conceptual one. Regardless, we argue that there is a distinction between diversity and heterogeneity and that at least in some cases, for example, in philosophy departments, there is an inverse relationship between the two that can lead to a diversity-heterogeneity conflict or trade-off. The general domain of such case studies is neither rare nor trivial.

This domain was first described by Warren Weaver as "organized complexity." Weaver portrayed past and future research directions while also popularizing Claude Shannon's measure of 'entropy' in the transmission of information.[23] Shannon's entropy, synonymized as 'diversity', has propagated through many fields, including ecology, theories of networks, systems, fuzzy sets, and economics.[24] These fields,

[22] We thank an anonymous reader for this important insight, which is formalized below (see section III) and illustrated graphically with empirical data from ecological food webs in the APPENDIX.

[23] We thank an anonymous reviewer for this reference and for putting our work in this context. According to Weaver, the behavior of the gene, brain, or society "are all problems which involve dealing simultaneously with a *sizable number of factors which are interrelated into an organic whole*. They are all, in the language here proposed, problems of *organized complexity*." Warren Weaver, "Science and Complexity," *American Scientist*, XXXVI, 4 (October 1948): 536–44, at p. 541, italics in original.

[24] Shannon quantified the uncertainty ('entropy') of a string of letters or numbers as $H' = -\sum_i p_i \ln p_i$, where p_i is the proportion of identical letters in the string. The more letters or numbers there were (that is, the more 'diverse' the series) and the more equal their relative proportions, the more difficult it would be to predict the next one. See Claude E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, XXVII, 3 and 4 (July and October 1948): 379–423 and 623–56; and Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949). Norbert Wiener independently studied information theory in his work on anti-aircraft defenses, which led to the development

and many others, use systems theory to define ties more clearly between individuals and groups, and to improve multi-cultural or multidisciplinary communication.[25] In all these fields the 'group' in question refers to a population or collective and may face the diversity-heterogeneity conflict.

II.3. Diversity and Heterogeneity Are Both Predicates of a Group, but Diverse Groups and Heterogeneous Groups Are Very Different. Diversity and heterogeneity are both predicates of the same type of entity—a group. This similarity can obscure an understanding of a conflict or trade-off within or between groups. Recent literature has clarified the conceptualization of biological[26] and human groups,[27] but our focus here is not on groups *per se* but on differences within and between them. In

---

of cybernetics and the initial formalization of systems theory. See Flo Conway and Jim Siegelman, *Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics* (New York: Basic Books, 2005).

Shannon and Weaver noted that as applied to communications, their measure of entropy quantified not what was said, but what could be said. This probabilistic notion underlies fuzzy set theory (see Lotfi A. Zadeh, "Fuzzy Sets," *Information and Control*, VIII, 3 (June 1965): 338–53), and Shannon's H' computes the diversity (entropy) of a fuzzy set from the uncertainty ('fuzziness') of its members (see Liu Xuecheng, "Entropy, Distance Measure and Similarity Measure of Fuzzy Sets and Their Relations," *Fuzzy Sets and Systems*, LII, 3 (December 1992): 305–18). This same index, H', has been widely adopted (and often misused) as a measure of species diversity in ecosystems. See, for example, Ian F. Spellerberg and Peter J. Fedor, "A Tribute to Claude Shannon (1916–2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the 'Shannon–Wiener' Index," *Global Ecology and Biogeography*, XII, 3 (May 2003): 177–79. An equivalent formulation was derived by Weitzman, who showed that H' satisfies a basic dynamic programming equation that yields an optimal classification scheme to justify the economic preservation of biological, historical, and cultural diversity; see Martin L. Weitzman, "On Diversity," *The Quarterly Journal of Economics*, CVII, 2 (May 1992): 363–405.

[25] "Knowledge of individual group behavior must be improved. Communication must be improved between people of different languages and cultures, as well as between all the varied interests which use the same language, but often with such dangerously differing connotations" (Weaver, "Science and Complexity," *op. cit.*, p. 544). Both Weaver and Zadeh were interested in similar problems and sought similar solutions. See Rudolf Seising, "Warren Weaver's 'Science and Complexity' Revisited," *Studies in Fuzziness and Soft Computing*, CCLXXIII (November 2012): 58–87.

[26] See Gissis, Lamm, and Shavit, eds., *Landscapes of Collectivity in the Life Sciences*, *op. cit.*

[27] See Margaret P. Gilbert, *Joint Commitment: How We Make the Social World* (New York: Oxford University Press, 2013); Michael, E. Bratman, *Shared Agency: A Planning Theory of Acting Together* (New York: Oxford University Press, 2014); Philip Petit, "How to Tell if a Group Is an Agent," in Jennifer Lackey, ed., *Essays in Collective Epistemology* (Oxford: Oxford University Press, 2014), pp. 97–121; Deborah P. Tollefson, *Groups as Agents* (Cambridge, UK: Polity Press, 2015); Michael S. Brady and Miranda Fricker, eds., *The Epistemic Life of Groups: Essays in the Epistemology of Collectives* (Oxford: Oxford University Press, 2016); and Katherine Hawley, "Social Mereology," *Journal of the American Philosophical Association*, III, 4 (Winter 2017): 395–411.

this context, Gabriel Uzquiano's "identification of groups with variable plural embodiment"[28] is helpful. It formally accounts for differences we identify between collectives and aggregated collections or populations while recognizing that both are no more than individuals coincidently co-existing in space and time and that any group could be part of another group. For Uzquiano, "[t]his account illuminates the difference between highly structured groups like committees and less cohesive groups like queues. The main difference is to be located in the principles of generation corresponding to each."[29] As shown hereafter, heterogeneous and diverse groups rely on different casual processes to maintain their different character.

'Epistemic collectives' nicely demonstrate this point. According to Margaret Gilbert, an epistemic collective is a group holding "a collective belief that *p* without all or most—or indeed any—members of the population in question believing that *p*."[30] This concept makes sense of statements like "the university believes that *p*" and the practice of a basketball team, meeting only once, randomly assembled and jointly deciding the team's strategy against its rival. Heterogeneous epistemic groups apply collective intentionality[31] via complex interactions, typically explicit deliberations accompanied by subtle social mechanisms, which, *ceteris paribus*, diverse epistemic groups need not have.

For example, an epistemic population holding a diversity of views is justified in conducting a simple voting process over *p* without deliberation or other forms of active participation for the sake of mutual understanding or agreement. It can rationally adopt its majority vote (as in pure "adversary democracy"[32]), average result (as in the "wisdom of crowds"[33]) or another measure regarding *p*,[34] whereas any such result

[28] Gabriel Uzquiano, "Groups: Toward a Theory of Plural Embodiment," this JOURNAL, CXV, 8 (August 2018): 432–52, at p. 451.

[29] *Ibid.*

[30] Margaret Gilbert, "Collective Epistemology," *Episteme*, I, 2 (October 2004): 95–107, at p. 98.

[31] As mentioned, the extensive literature on this matter is outside the scope of this article. See also Marija Jankovic and Kirk Ludwig, eds., *The Routledge Handbook of Collective Intentionality* (New York: Routledge, 2017).

[32] We thank an anonymous reviewer for pointing to adversary and unitary democracy here. See Jane Mansbridge, "Beyond Adversary Democracy," in Derek Barker and David W. Brown, eds., *Higher Education Exchange* (Dayton, OH: The Kettering Foundation, 2017), pp. 6–13.

[33] James M. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (New York: Doubleday, 2004).

[34] For example, if an investigator studies a certain group *x*, and *p* states that *x* is racist, then even if only 30% percent of *x*'s members would answer affirmatively to the question "do you believe in white supremacy?," the investigator would be justified to state that *p* is true. We thank Eli Pitkovski for this comment.

alone cannot suffice for justifying p in a heterogeneous epistemic collective, which requires deliberation in search of areas of agreement within multifaced spaces of differences.[35]

Both groups aim for a true decision about p, yet only the latter also aims for a *joint process* of decision making, though not necessarily a consensus verdict. In contrast, the former is focused only on that the final verdict reflects *individual* preference as much as possible. In addition, a group decision over p could be based on the conclusion alone (p) or also on its premises (q_1, q_2), and different procedures for group decision—with or without first deciding the premises—could deduce conflicting group conclusions (p or $\sim p$).[36] Populations, holding diverse views not necessarily connected or deliberated, fit well with a simple majority vote on a single proposition p (or multiple disconnected single conclusions $p_1, p_2, \ldots p_n$), thereby shrinking the public sphere in accord with a minimal liberal account. Heterogeneous groups, which necessarily try to interconnect their differences, are inclined to deliberate, for example, to add q_1 and q_2 into the deliberation over p, thus increasing the public sphere in accord with a comprehensive liberal account.[37] That is, the structure of differences within a groups could be relevant to its decision structure and the extent of its public sphere.

II.4. Diversity and Heterogeneity Can Affect Objectivity, Robustness, and Empirical Accuracy. Different knowledge production processes may be considered less objective, robust, or empirically accurate. We discuss here the meaning of these fundamental concepts in philosophy of science in the framework of the diversity-heterogeneity distinction.

Helen Longino argued that objectivity increases if different contextual values are conserved as necessary elements of a pluralistic scientific community.[38] Contextual values are the personal, social, and cultural values of groups or individuals about "what ought to be" in science.[39] We agree with Longino that confirmation is done within a

[35] "To maintain its legitimacy, a democracy must have both a unitary and an adversary face. . . becoming neither and absorbing neither, but holding them together." See Mansbridge, "Beyond Adversary Democracy," *op. cit.*, pp. 8–9.

[36] This is known as the "discursive dilemma," and its relevance for aggregative/non-aggregative collective judgment is reviewed by Christian List, "The Discursive Dilemma and Public Reason," *Ethics*, CXVI, 2 (January 2006): 362–402. We thank David Enoch for this reference and its relevance.

[37] *Ibid.*

[38] See Helen E. Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry* (Princeton, NJ: Princeton University Press, 1990); and Helen E. Longino, *The Fate of Knowledge* (Princeton, NJ: Princeton University Press, 2002).

[39] Longino, *Science as Social Knowledge*, *op. cit.*, pp. 4 and 83–86; cf. Shannon and Weaver's notion of entropy ('diversity') as what could be [said]. See footnote 24.

context; that scientific models and practices are laden with contextual values; that a pluralistic deliberation of different values adds relevant doubt and contexts to improve scientific confirmation; and that when this pluralistic process of examination reaches a joint result, it is more objectively confirmed. Longino's 'confirmation' requires a heterogeneous scientific community and a democratic process of decision making, combining unitary procedures seeking agreement via personal deliberation with adversarial ones foregrounding interpersonal conflict. Such confirmation requires scientists to actively welcome and value disagreements within and outside their community.[40]

Because this process of pluralistic deliberation cannot occur without coordinated interaction of agreement and—more importantly—disagreement within a larger group, seeking to improve scientific objectivity first requires one to recognize the epistemic, social, and political dimensions of the discussion. Merely recognizing the existence of diverse values and interests while following science's implicit 'majority vote' at each given time does not suffice for objective confirmation. Although such recognition promotes openness to differences, it is not pluralistic in the sense promoted by Longino but rather promotes conformism with the hegemonic or majority point of view. On the other hand, a heterogeneous scientific collective, combining adversary and unitary democratic procedures and reaching joint results from a pluralistic perspective, is not only better confirmed, but also more just.

Different paths leading to the same result characterize another core scientific concept: robustness. Jonah Schupbach described 'robustness analysis' as "exploring the differences in past means of detection that have had no (or a negligible) effect on the result—and also by exploring variations that have made a difference to the result."[41] A robust model is effectively true, because, as famously stated by Richard Levins, "our truth is the intersection of independent lies."[42] William Wimsatt clarified that full independence cannot be obtained,[43] and Schupbach abandoned independence all together. Instead, he described robustness analysis as "explanatorily discriminating bits of evidence, which successively eliminate more and more of *H*'s competitors [that is, competing hypotheses]."[44]

[40] Longino, "Interaction: A Case for Ontological Pluralism," *op. cit.*

[41] Jonah N. Schupbach, "Robustness Analysis as Explanatory Reasoning," *The British Journal of Philosophy of Science*, LXIX, 1 (March 2018): 275–300, at p. 277.

[42] Richard Levins, "The Strategy of Model Building in Population Biology," *American Scientist*, LIV, 4 (December 1966): 421–31, at p. 423.

[43] William C. Wimsatt, *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (Cambridge, MA: Harvard University Press, 2007).

[44] Schupbach, "Robustness Analysis as Explanatory Reasoning," *op. cit.*, p. 288.

Adding the diversity-heterogeneity distinction can refine Schupbach's analysis, according to which "…it is really not so relevant whether means of detection are strongly diverse or sufficiently heterogeneous in some absolute sense, independent of considered hypotheses. What matters for RA [robustness analysis] diversity is that the means (which may actually be quite similar in most respects) are different in just the sense required to rule out the target hypothesis's salient competitors."[45] To empirically test and subsequently reject or support a hypothesis, each competing hypothesis needs to incorporate at least some—but not all—major structural elements of the model's description (interacting, casual factors) and deduce from that descriptive difference a different interpretation (empirical hypothesis and prediction) about the target in the world.[46]

Differences among model interpretations could be heterogeneous or diverse. If heterogeneous, then describing the interacting casual factors that produced a different prediction is necessary. To compare the fit of the prediction to the existing data, it is necessary to compare the fit of underlying casual factors. Since a comparison needs common ground, then at least some casual factors must be shared among the different models. The more descriptive similarity there is between competing models, the stronger they compete and the clearer the verdict of the analysis becomes. As more heterogeneous predictions of other models are outcompeted by the prediction of one model, the latter can be said to be better supported via robustness analysis.

Diverse predictions are less constrained. Predictions from different models could differ from one another not because one model's casual structure more accurately fits the phenomenon, but because of internal—and often unknown or inappropriate—differences among model descriptions ("causes") that are invisible to users.[47] Therefore, even if a set of diverse predictions were outcompeted by those of a particular model, the conclusion may not be robust because the predictions were inappropriately supported by specious causal chains.

[45] *Ibid.*, pp. 288–89.

[46] For details, see the analysis of similarity in Weisberg, *Simulation and Similarity*, *op. cit.*, chapter 13.

[47] We thank an anonymous reviewer for this point, and for noting that if we do not understand how a model makes a prediction, it could be using "causal" variables such as race or gender that might be associated with actual causes for contextual reasons (for example, African-American men may have higher rates of mortality from an infectious disease not because of their gender or race but because they are more likely to suffer discrimination by Caucasian healthcare professionals). This problem can be offset at least in part by mathematical descriptions of models that aim to make all assumptions explicit.

Given Schupbach's meaning of 'difference'—". . . in just the sense required to rule out the target hypothesis's salient competitors"[48]—then it would matter if a diverse or heterogeneous hypothesis were eliminated because one reasonably expects competing hypotheses to be more robust if they are heterogeneous rather than merely diverse.

If one accepts our view of 'robustness' and 'objectivity', and since it is generally accepted that a more robust and objective confirmation leads to truer scientific results,[49] one can expect heterogeneous epistemic groups, at least in some contexts, to succeed better than homogenous or diverse groups. Indeed, Scott Page empirically showed that heterogeneous groups *are* significantly more efficient at finding correct answers to complex puzzles than equal-sized homogenous groups of elite cognitive achievers or diverse, larger crowds.[50] The advantage of heterogeneous groups disappeared when they became too large and thus their integrative interactions less efficient, or when the puzzle was too simple for deliberation to help find an answer (for example, an ox's weight or a submarine's location).[51] Given the rate of knowledge accumulation and the fact that in these hyperspecialized times[52] it is impossible for anyone to master a discipline or even subdiscipline,[53] small heterogeneous groups may be expected to become an even more visible and successful route for scientific progress in tackling Weaver's "organized complex" ("fuzzy") problems. In sum, noticing the diversity-heterogeneity gap can help one use the specific advantages of each type of difference, yet also necessitates identifying different parameters for each type of difference. A formal account of what these parameters are and how to track them will be explicated in the next two sections.

[48] Schupbach, "Robustness Analysis as Explanatory Reasoning," *op. cit.*, p. 289.

[49] Mieke Boon, "Understanding Scientific Practices: The Role of Robustness Notions," in Léna Soler et al., eds., *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (New York: Springer, 2012), pp. 289–316.

[50] Scott E. Page, *Diversity and Complexity* (Princeton, NJ: Princeton University Press, 2011).

[51] On the loss of information due to deliberation, see Miriam Solomon, "*Groupthink* versus *The Wisdom of Crowds*: The Social Epistemology of Deliberation and Dissent," *The Southern Journal of Philosophy*, XLIV, S1 (Spring 2006): 28–42.

[52] Millgram convincingly argues that philosophers are especially relevant for these hyperspecialized times, given their professional training in logic and dialogue. See Elijah Millgram, *The Great Endarkenment: Philosophy for an Age of Hyperspecialization* (Oxford: Oxford University Press, 2015).

[53] We thank an anonymous reviewer for foregrounding the link between heterogeneity, Weaver's work, and interdisciplinarity.

III. FORMAL DEFINITIONS OF DISSIMILARITY

III.1. Fundamental Statistical Estimators Are Predicates of Single Populations. Formal ways of measuring dissimilarity more clearly illustrate the conflation between diversity and heterogeneity. Briefly, measurable properties ('variables') of a group of individual entities (a 'population' of cells, organisms, and so on) rarely are perfectly similar. Rather, they will take on a range of values $\boldsymbol{y} = \{y_1, y_2, y_3, \ldots y_n\}$, where the value of the variable measured for the i^{th} individual is denoted y_i. The average (mean or expected) value $\overline{y}$ of the distribution of the measured variables equals the sum of all the individual measurements divided by the number of individuals, n: $\overline{y} = \sum_{i=1}^{i=n} \frac{y_i}{n}$. The variance is the sum of the squared differences between each individual measurement and the mean: $\sigma^2 = \sum_{i=1}^{i=n} (y_i - \overline{y})^2$.[54] The standard error of the mean ($\frac{\sqrt{\sigma}^2}{n}$) provides an intuitive estimate of the variability of a set of measurements.[55]

III.2. Measures of Diversity Are Predicates of One or More Populations. 'Variance' and 'mean' are predicates of single populations (groups of measurements), whereas heterogeneity and diversity are not one-dimensional predicates. The former are composite properties of a 'sample'—a group of measurements taken from more than one population[56]—whereas the latter are derived from a collection of datasets that describe a wide range of different, sometimes incommensurate, variables from one or more sampled populations. Statisticians rarely,

[54] Ronald A. Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver and Boyd, 1925).

[55] Under reasonable assumptions, ≈63% of the measurements fall within ±1 standard error of the mean, and ≈95% fall within ±2 standard errors of the mean. See Aaron M. Ellison and Brian Dennis, "Paths to Statistical Fluency for Ecologists," *Frontiers in Ecology and the Environment*, VIII, 7 (September 2010): 362–70, for a comprehensive discussion of the assumptions behind these estimates and calculations of associated confidence intervals.

[56] For example, the classical analysis of variance (ANOVA) can be used to determine if two or more populations differ in their average measured traits. ANOVA assumes that the variances of the populations being compared are equal; this is referred to as "homogeneity of variance" or "homoskedasticity." In contrast, if variances are unequal ('heterogeneous' or 'heteroskedastic'), mathematical transformations of the data must be done to ensure that variances are homogeneous prior to comparing populations using ANOVA. See Ronald A. Fisher, "The Correlation between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society of Edinburgh*, LII, 2 (1918): 399–433; and Nicholas J. Gotelli and Aaron M. Ellison, *A Primer of Ecological Statistics*, 2nd ed. (Sunderland, UK: Sinauer Associates, 2012), for additional details on the difficulties raised by heterogeneity of variances in classical statistics. Note that the usage of heterogeneity in reference to among-group variances describes a problem to overcome so that valid statistical comparisons among different populations can be made. This use of heterogeneity is different from that used elsewhere in this article.

if ever, refer to or use measures of diversity, but biologists and social scientists regularly use many of them.

III.3. Measuring Diversity.

All diversity metrics in the biological or social realm are derived from two fundamental quantities. The first is the number of different objects (individuals, species, ethnicities, professions, and so on) in an assemblage (S), and the second is the relative abundance p_i of each of the i^{th} objects ($p_i = \frac{n_i}{\sum_{i=1}^{S} n_i}$).

These two quantities are combined in what are called 'Hill numbers':[57]

$$(1)\quad {}^{q}D = \left(\sum_{i=1}^{S} p_i^{q} \right)^{1/(1-q)}$$

The value of q determines how sensitive the measure of diversity is to the distribution of relative abundances.

When $q = 0$, there is no contribution of relative abundance to the measure of diversity, so diversity is simply the number of different objects: ${}^{q}D \equiv {}^{0}D = S$.

When $q = 1$, common and rare objects are weighted equally and in proportion to their relative abundances:[58]

$$(2)\quad {}^{1}D = \exp\left(- \sum_{i=1}^{S} p_i \ln p_i \right)$$

When $q = 2$, equation (1) is equivalent to the Simpson Diversity Index[59] (used by ecologists studying biological diversity), and the Herfindahl-Hirschman Index[60] (used by economists), which were developed concurrently and independently, and employ the same mathematical formulation. Both indices weight abundant (common) objects more than rare ones:

$$(3)\quad {}^{2}D = 1 \Big/ \sum_{i=1}^{S} p_i^{2}$$

[57] Mark O. Hill, "Diversity and Evenness: A Unifying Notation and Its Consequences," *Ecology*, LIV, 2 (March 1973): 427–32.

[58] Note that equation (1) is undefined for $q = 1$ because the denominator $(1 - q)$ of its exponent is undefined. However, the limit as $q \to 1$ converges to the value given by equation (2). The quantity being exponentiated in Equation 2 is equivalent to Shannon's H' discussed in footnote 24.

[59] Edward H. Simpson, "Measurement of Diversity," *Nature*, CLXIII, 4148 (April 1949): 688.

[60] This index is combined from presentations in two manuscripts: Albert O. Hirschman, *National Power and the Structure of Foreign Trade* (Berkeley: University of California Press, 1945); and Orris C. Herfindahl, "Concentration in the US Steel Industry," PhD diss. (Columbia University, 1950).

Equations 1–3 quantify the number of objects present in a sample of interest and the abundance of each object, but *not* their interactions—for example, competition, mutualism, symbiosis—nor their location or function within a system. Both indices are widely used and applied in biodiversity and economics, and their similarity attests to the many implicit ties that society weaves between social and ecological realms.

Unlike variance, which is interpretable on its own, diversity has little meaning unless it is associated with a sample or a population. Ecologists 'partition' diversity into three distinct components: alpha, beta, and gamma diversity. Alpha diversity is the simplest diversity of a single local population, and gamma diversity is the diversity of the regional area in which multiple local populations occur. Thus, alpha diversity of any local population is a subset of gamma diversity, and by definition, the union of all alpha diversities within a region is its gamma diversity $\alpha \subseteq \gamma; \cup_i \alpha_i = \gamma$. Beta diversity measures the dissimilarity among populations and is computed as the quotient of alpha and gamma diversity: $\frac{\alpha}{\gamma} = \beta$.[61]

Many researchers in the social sciences use 'diversity' as a catchall term not attached to any particular measured process. In contrast, Page suggests three kinds of diversity: (1) *variation*, or diversity within a type, referring to quantitative differences in a specific variable; (2) *diversity of types*, referring to qualitative differences between types; and (3) *diversity of composition*, or the way types are arranged.[62] Page's *variation* is directly analogous to the ecologist's alpha and gamma diversity, whereas his *diversity of types* and *diversity of composition* are analogous to different ecological conceptualizations of beta diversity. Neither has much to do with heterogeneity.

Finally, in philosophy, 'diversity' and 'heterogeneity' are used interchangeably. In political and moral philosophy these terms are used in critical discussions of inclusion and epistemic injustice; for example, when noticing that philosophy programs are especially homogenous in terms of their 'alpha diversity' of self-identified ethnicity, race, or gender within a department.[63] In the next section we elaborate on

[61] For additional details on partitioning diversity, see Aaron M. Ellison, "Partitioning Diversity," *Ecology*, XCI, 7 (July 2010): 1962–63.

[62] Page, *Diversity and Complexity*, *op. cit.*

[63] Currently 3.1% of the doctoral students registered in the American Philosophical Association self-identify as African American, compared to an average of 8.8% in all doctoral programs. Women occupy 50.8% of the American population, 53.3% of all doctoral programs, and only 25.6% of the doctorates in the American Philosophical Association. See Carolyn D. Jennings et al., "The Diversity and Inclusivity Survey: Final Report" (Washington, DC: APA Grants, 2019).

what distinguishes the concepts of diversity and heterogeneity, and their effect on other core concepts in philosophy of science.

IV. A MEASURE OF HETEROGENEITY

In a nutshell, zoos are diverse whereas ecosystems are heterogeneous. Like a zookeeper enumerating her different animals, measuring diversity in a department or ecosystem means counting individuals (defined by group memberships) present at a certain place and time. But simply being present does not imply that one will interact with another, and it is those interactions and their relative strengths that distinguish a viable heterogeneous collective from a diverse collection of different things. Weaver distinguished between the two and called upon new mathematical tools for tackling the former, while Wiener independently[64] articulated the mathematical foundations of what we now call systems theory. Here, we use recent applications of systems theory developed for analyzing networks of interacting entities (for example, species in a food web, individuals in a social network) to suggest a range of measures of heterogeneity—from the simple to the complex—that could be used to determine the extent to which a group is heterogeneous or diverse.

IV.1. Definition of a Network. A 'network' is a set of connected things. More formally, a network consists of a set of interacting *nodes* (individuals, species, buildings sharing infrastructure, and so on) that are connected by *edges* (links, paths) that define the relationships between each pair of nodes. Canonical examples of networks include food webs, in which the nodes represent species and the edges define who-eats-whom relationships, and social networks, in which the nodes represent individual people or social groups and the edges define how they are connected to one another (for example, 'employee' and 'boss' in an organization or 'followers' on Twitter). Nodes and edges may be weighted by, respectively, their abundance and strength or direction of their interactions (for example, strong positive interactions, weak negative interactions, and so on), or they may be unweighted.

IV.2. Network Connectance: A Basic Metric of Heterogeneity. Network metrics characterize its structure at levels ranging from individual nodes or edges through sub-networks (groups of two or more nodes plus any

[64] See footnote 24 and Conway and Siegelman, *Dark Hero of the Information Age*, *op. cit.*

edges between them) to the entire network.[65] For assessing system-wide heterogeneity, whole-network and sub-network metrics are likely to be more informative than individual-level metrics.

As with metrics of diversity (section III.3), metrics of heterogeneity are legion but all are derived from two basic quantities: the number of nodes, N (which may be the number of species in an assemblage, ethnicities in the college, people in a sample, and so on) and the number of edges (interactions) between the nodes, L. If every node is connected to every other node, there would be $L = N(N-1)/2$ total edges. But in most networks, not every node is connected to all others. For example, in a basic food web, herbivores eat plants but not predators. We therefore define the 'connectance' C of a network as the number of edges that do occur relative to those that could:

$$(4) \quad C = \frac{L}{[N\,(N-1)/2]}$$

C is the most basic measure of heterogeneity. A group of species or individuals, no matter how diverse, will have $C = 0$ if none of them interact with one another.[66] As interactions increase, so will C. A fully connected network will have $C = 1$. Increasing connectivity among heterogeneous nodes is necessary to change a diverse collection into a collective, yet collectivity requires more than connectivity. Analogous with our simplest measure of diversity (${}^0D = S$), which has no information on relative abundance, C as a measure of heterogeneity includes no information on strength or direction of interactions.

IV.3. Ascendency Includes Interaction Strength. It is mathematically straightforward to include interaction strength in measures of network connectivity.[67] We need two additional quantities: T_{ij} and X_j. The first, T_{ij}, measures the transfer of energy, materials, information, and so on from node (species, individual) i to node j. T_{ij} is in measured (raw data) units (for example, calories/day). The second, X_j, is the rate of any additional, external inputs to node j (that is, inputs not coming from another node in the network). The proportion that each node i contributes (outputs) to the total input of node j is then

[65] For an extended discussion of network metrics in ecology, see Matthew K. Lau et al., "Ecological Network Metrics: Opportunities for Synthesis," *Ecosphere*, VIII, 8 (August 2017): e01900. For a more general introduction, see Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations* (New York: Springer, 2005).

[66] See, for example, Beverly Daniel Tatum, *"Why Are All the Black Kids Sitting Together in the Cafeteria?": And Other Conversations about Race* (New York: Basic Books, 1997).

[67] These and other metrics of network structure can be computed using the '`enaR`' and '`econullnetr`' packages of the `R` software system, https://r-project.org/, accessed February 15, 2020.

computed as $g_{ij} = T_{ij}/T_{.j}X_j$, where the subscripted 'dot' indicates the summation over the missing index $(T_{.j} = \sum_{i=1}^{N} T_{ij})$. g_{ij} ranges from 0 to 1.

A more complete metric of network heterogeneity that includes both the number of connections and their strengths is network *ascendency*,[68] A:

$$(5)\quad A = \sum_{i,j} T_{ij} \ln\left(\frac{T_{ij}T_{..}}{T_{i.}T_{.j}}\right)$$

The ascendency of a network is a combination of its overall organization and the amount of material, energy, or information flowing through it. In ecological food webs, it tends to increase initially with the number of different species but then levels out (see APPENDIX).[69] To our knowledge, ascendency and other network metrics that account for interaction strength have not yet been applied to human social networks.

IV.4. The Potential Importance of Individuals and Sub-networks in Network Function. Most recently, ecologists and computational biologists have been linking measures of species diversity (especially ${}^2D$, which emphasizes dominant species in an assemblage; section III.3) and the structure of tightly interacting sub-networks (for example, densely connected nodes called "rich clubs," or fully connected rich clubs called "cliques") to improve our understanding of network stability and resiliency.[70] Similar methods have been applied to analysis of the internet and other social networks,[71] but not to social situations where diversity not only is described but also is a focus of change because of historical and ongoing epistemic, social, and environmental injustices

[68] Noticing the interwoven ties between science and society, upon coining the term *ascendency*, Ulanowicz deliberately used a non-standard spelling (with a second "e") to distinguish it from ascendancy (with an "a") and to avoid the latter's negative connotations and emotive weight associated with dominance. See Robert E. Ulanowicz, "Quantitative Methods for Ecological Network Analysis," *Computational Biology and Chemistry*, XXVIII, 5–6 (December 2004): 321–39.

[69] For a detailed discussion of ascendency and other network metrics that take advantage of interaction strength, see Robert E. Ulanowicz, *Ecology, the Ascendent Perspective* (New York: Columbia University Press, 1997); and Robert E. Ulanowicz, Robert D. Holt, and Michael Barfield, "Limits on Ecosystem Trophic Complexity: Insights from Ecological Network Analysis," *Ecology Letters*, XVII, 2 (February 2014): 127–36.

[70] See Zhanshan (Sam) Ma and Aaron M. Ellison, "Dominance Network Analysis Provides a New Framework for Studying the Diversity–Stability Relationship," *Ecological Monographs*, LXXXIX, 2 (May 2019): e01358.

[71] For examples, see Shi Zhou and Raúl J. Mondragón, "The Rich-Club Phenomenon in the Internet Topology," *IEEE Communications Letters*, VIII, 3 (March 2004): 180–82; and Mark E. J. Newman and Aaron Clauset, "Structure and Inference in Annotated Networks," *Nature Communications*, VII (June 2016): 11863.

(see section V). We encourage others to use these different measures of heterogeneity rather than suffice with "headcounts" (alpha diversity) as a uniform measure of diversity.

We emphasize that a collection of nodes (individuals, social groups, species) with randomly placed edges (connections) between the nodes are unlikely to lead to heterogeneous networks. Rather, emergent properties of complex systems, including stability and resiliency, arise only from particular sets of connections with unequal strengths. Network topology can describe the connections, but ascendency (Equation 5) and other measures associate the topology with strengths of interactions and can predict emergent properties.[72]

V. CONCLUSION

Diversity is not heterogeneity, and a continued focus on the former does not necessarily increase the latter. Disputing readers could argue that because heterogeneity is a special case of diversity (section II.2), a trade-off or conceptual tension between them is unnecessary. Practically speaking, some may argue that instead of dwelling on complex semantic distinctions, we should ardently promote diversity whenever possible in order to make gradual progress toward a more heterogeneous and just society. We believe we have shown both lines of thought to be misleading. Our suggested concept of heterogeneity is justified not only because of its mathematical difference from measures of diversity, but also—and perhaps more importantly—because of the moral, epistemic, and empirical costs of the common practice to advocate for diversity as a sufficiently measurable goal for attaining all types of difference in all types of groups. Demography of a university is one such example (section I); describing and managing ecosystems is another (see APPENDIX). More generally, the problem is that simply measuring 'difference' is inherently ambiguous, since the same word adheres to two different concepts—diversity and heterogeneity—applied to different types of groups—populations and collectives—committed to different model descriptions—ignoring structure and engagement as causal factors or targeting them—and deducing different policy recommendations. Whereas both concepts are needed for addressing the problem of modeling group differences, one cannot practically do both at the same spatiotemporal scale. In the case of modeling collectives or modeling an aggregative population that

[72] See Ulanowicz, *Ecology, the Ascendent Perspective*, *op. cit.*; and Ma and Ellison, "Dominance Network Analysis Provides a New Framework for Studying the Diversity–Stability Relationship," *op. cit.*

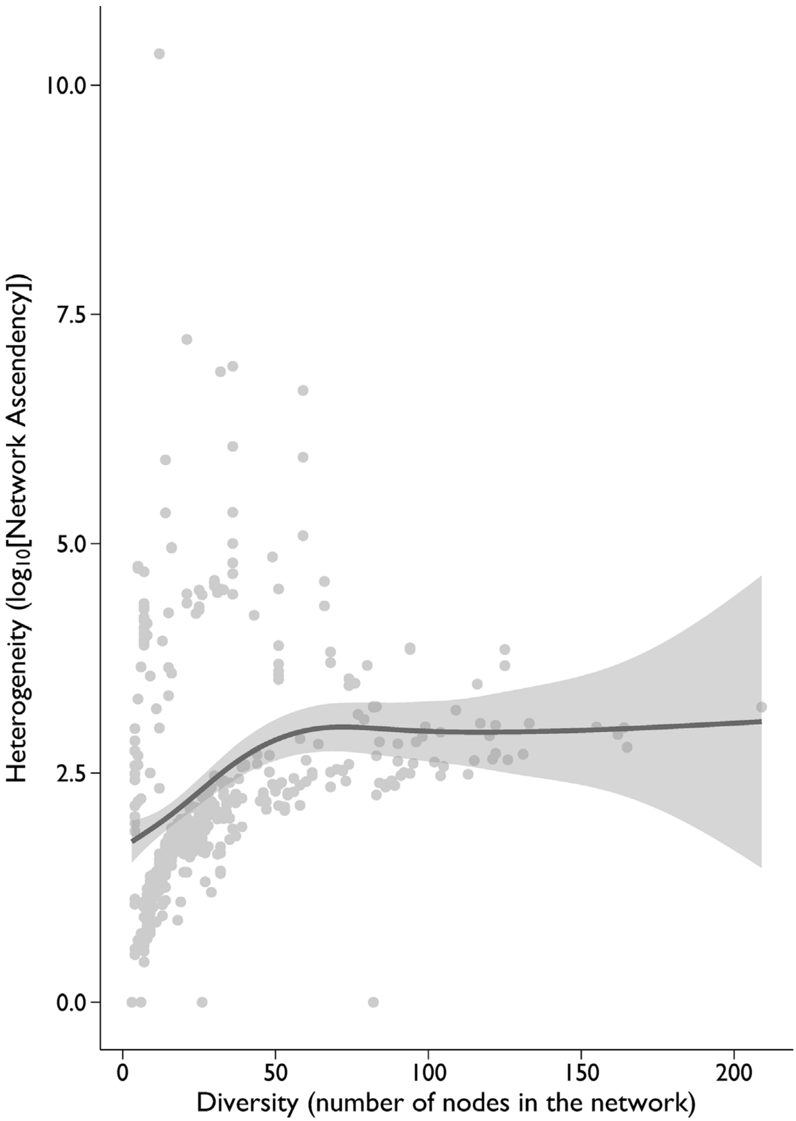
one hopes will become a collective, we advise to first measure heterogeneity.

Furthermore, clarifying the concept of heterogeneity has helped to clarify other basic scientific concepts such as 'group', 'robustness', and 'objectivity'. Formalizing this concept opened a path for its interdisciplinary usages in many other fields for addressing a range of additional questions regarding Weaver's domain of complex phenomena. Finally, implementing heterogeneity can more rapidly advance certain social goals, including integrating different groups, identities, perspectives, and sources of information. Finally, because heterogeneity also facilitates social interactions among groups that carry epistemic and moral implications, paying attention to it may help to reduce epistemic injustice. Conversely, diversity alone often leads to divergence, is insufficient for resisting social injustice, and misses epistemic opportunities that result from integrative working interactions. For conceptual, practical, epistemic, and moral reasons, we suggest learning from the difference between heterogeneity and diversity, working with both in different contexts, and foregrounding the relevant type of difference in the appropriate context.

APPENDIX

See the figure on the following page. The relationship between the diversity of species in 430 unique ecological networks (food webs) and heterogeneity of those same networks increases initially and then levels off. Diversity and heterogeneity of each food web is represented by a single gray dot. The thick black line is the best-fit general additive model to all the data, and the gray shading is the estimated 95% confidence interval on the fitted model.[73]

[73] Data compiled by Aaron M. Ellison and Nicholas J. Gotelli. Data and code to reproduce this figure are available online at https://doi.org/10.6073/pasta/5dc19f7d3cc4db6fb86a2984fe126f4a, accessed January 15, 2021.



AYELET SHAVIT

Tel Hai College and Technion

AARON M. ELLISON

Harvard University and Sound Solutions for Sustainable Science