🗨 **PERSPECTIVE**

Stewardship of global collective behavior

Joseph B. Bak-Coleman[a,b,1] 🄳, Mark Alfano[c,d] 🄳, Wolfram Barfuss[e,f] 🄳, Carl T. Bergstrom[g] 🄳,
Miguel A. Centeno[h], Iain D. Couzin[i,j,k] 🄳, Jonathan F. Donges[e,l], Mirta Galesic[m] 🄳, Andrew S. Gersick[n],
Jennifer Jacquet[o], Albert B. Kao[m] 🄳, Rachel E. Moran[a], Pawel Romanczuk[p], Daniel I. Rubenstein[n] 🄳,
Kaia J. Tombak[q], Jay J. Van Bavel[r,s], and Elke U. Weber[t,u] 🄳

Collective behavior provides a framework for understanding how the actions and properties of groups emerge from the way individuals generate and share information. In humans, information flows were initially shaped by natural selection yet are increasingly structured by emerging communication technologies. Our larger, more complex social networks now transfer high-fidelity information over vast distances at low cost. The digital age and the rise of social media have accelerated changes to our social systems, with poorly understood functional consequences. This gap in our knowledge represents a principal challenge to scientific progress, democracy, and actions to address global crises. We argue that the study of collective behavior must rise to a "crisis discipline" just as medicine, conservation, and climate science have, with a focus on providing actionable insight to policymakers and regulators for the stewardship of social systems.

collective behavior | computational social science | social media | complex systems

Collective behavior historically referred to instances in which groups of humans or animals exhibited coordinated action in the absence of an obvious leader (1–4): from billions of locusts, extending over hundreds of kilometers, devouring vegetation as they move onward; to schools of fish convulsing like some animate fluid while under attack from predators; to our own societies, characterized by cities, with buildings and streets full of color and sound, alive with activity. The characteristic feature of all of these systems is that social interactions among the individual organisms give rise to patterns and structure at higher levels of organization, from the formation of vast mobile groups to the emergence of societies with division of labor, social norms, opinions, and price dynamics.

Over the past few decades "collective behavior" has matured from a description of phenomena to a framework for understanding the mechanisms by which collective action emerges (3–7). It reveals how large-scale "higher-order" properties of the collectives feed back to influence individual behavior, which in turn can influence the behavior of the collective, and so on. Collective behavior therefore focuses on the study of individuals in the context of how they influence and are influenced by others, taking into account the causes and consequences of interindividual differences in physiology, motivation, experience, goals, and other properties.

The multiscale interactions and feedback that underlie collective behavior are hallmarks of "complex systems"—which include our brains, power grids, financial markets, and the natural world (8, 9). When perturbed, complex systems tend to exhibit finite

aCenter for an Informed Public, University of Washington, Seattle, WA 98195; beScience Institute, University of Washington, Seattle, WA 98195; cEthics & Philosophy of Technology, Delft University of Technology, 2628 CD Delft, The Netherlands; dInstitute of Philosophy, Australian Catholic University, Banyo Queensland 4014, Australia; eEarth System Analysis, Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, 14473 Potsdam, Germany; fTübingen AI Center, University of Tübingen, 72074 Tübingen, Germany; gDepartment of Biology, University of Washington, Seattle, WA 98195; hPrinceton School of Public and International Affairs, Princeton University, Princeton, NJ 08544; iDepartment of Collective Behaviour, Max Planck Institute of Animal Behavior, 78315 Radolfzell am Bodensee, Germany; jCentre for the Advanced Study of Collective Behaviour, University of Konstanz, 78464 Konstanz, Germany; kDepartment of Biology, University of Konstanz, 78464 Konstanz, Germany; lStockholm Resilience Centre, Stockholm University, 11419 Stockholm, Sweden; mSanta Fe Institute, Santa Fe, NM 87501; nDepartment of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; oDepartment of Environmental Studies, New York University, New York, NY 10012; pInstitute for Theoretical Biology, Department of Biology, Humboldt Universität zu Berlin, 10115 Berlin, Germany; qDepartment of Anthropology, Hunter College of the City University of New York, New York, NY 10065; rDepartment of Psychology, New York University, New York, NY 10003; sCenter for Neural Science, New York University, New York, NY 10003; tDepartment of Psychology, Princeton University, Princeton, NJ 08544; and uAndlinger Center for Energy and Environment, School of Engineering and Applied Science, Princeton University, Princeton, NJ 08544

resilience followed by catastrophic, sudden, and often irreversible changes in functionality (9, 10). Across a wide range of complex systems, research has highlighted how anthropogenic disturbance—technology, resource extraction, and population growth—is an increasing, if not dominant, source of systemic risk. Yet, scientific research on how complex systems are impacted by human technology and population growth has largely focused on the threats that these pose to the natural world (11–13). We have a far poorer understanding of the functional consequences of recent large-scale changes to human collective behavior and decision making. Our social adaptations evolved in the context of small hunter-gatherer groups solving local problems through vocalizations and gestures. Now we face complex global challenges from pandemics to climate change—and we communicate on dispersed networks connected by digital technologies such as smartphones and social media.

With increasingly strong links between ecological and sociological processes, averting catastrophe in the medium term (e.g., coronavirus) and the long term (e.g., climate change, food security) will require rapid and effective collective behavioral responses—yet it remains unknown whether human social dynamics will yield such responses (14–17). In addition to existential ecological and climatic threats, human social dynamics present other challenges to individual and collective wellbeing, such as vaccine refusal, election tampering, disease, violent extremism, famine, racism, and war.

Neither the evolutionary nor the technological changes to our social systems have come about with the express purpose of promoting global sustainability or quality of life. Recent and emerging technologies such as online social media are no exception—both the structure of our social networks and the patterns of information flow through them are directed by engineering decisions made to maximize profitability. These changes are drastic, opaque, effectively unregulated, and massive in scale.

The emergent functional consequences are unknown. We lack the scientific framework we would need to answer even the most basic questions that technology companies and their regulators face. For instance, will a given algorithm for recommending friends—or one for selecting news items to display—promote or hinder the spread of misinformation online? We do not have access to a theory-driven, empirically verified body of literature to inform a response to such a question. Lacking a developed framework, tech companies have fumbled their way through the ongoing coronavirus pandemic, unable to stem the "infodemic" of misinformation that impedes public acceptance of control measures such as masks and widespread testing (18).

In response, regulators and the public have doubled down on calls for reforming our social media ecosystem, with demands ranging from increased transparency and user controls to legal liability and public ownership. The basic debate is an ancient one: Are large-scale behavioral processes self-sustaining and self-correcting, or do they require active management and guidance to promote sustainable and equitable wellbeing (2, 19)? Historically, these questions have been addressed in philosophical or normative terms. Here, we build on our understanding of disturbed complex systems to argue that human social dynamics cannot be expected to yield solutions to global issues or to promote human wellbeing without evidence-based policy and ethical stewardship.

The situation parallels challenges faced in conservation biology and climate science, where insufficiently regulated industries optimize profits while undermining the stability of ecological and earth systems. Such behavior created a need for urgent evidence-based policy in the absence of a complete understanding of the systems' underlying dynamics (e.g., ecology and geosciences). These features led Michael Soulé to describe conservation biology as the "crisis discipline" counterpoint to ecology—an analogy to the relationship between medicine and comparative physiology (20). Crisis disciplines are distinct from other areas of urgent, evidenced-based research in their need to consider the degradation of an entire complex system—without a complete description of the system's dynamics. We feel that the study of human collective behavior must become the crisis discipline response to changes in our social dynamics.

Because human collective behavior is the result of processes that span temporal, geographical, and organizational scales, addressing the impact of emerging technology on global behavior will require a transdisciplinary approach and unprecedented collaboration between scientists across a wide range of academic disciplines. As our societies are increasingly instantiated in digital form, once-mathematical abstractions of social processes—networks are one prominent example—become very real parts of daily life (21–23). These changes present new challenges, as well as opportunities for measurement and intervention. Disciplines within and beyond the social sciences have access to techniques and ways of thinking that expand our ability to understand and respond to the effects of communication technology. We believe such a collaboration is urgently needed.

In what follows, we begin by framing human collective behavior as a complex adaptive system shaped by evolution, a system that much like our natural world has entered a heavily altered and likely unsustainable state (14, 24, 25). We highlight how communication technology has restructured human social networks, expanding, reorganizing, and coupling them to technological systems. Drawing on insight from complexity science and related fields, we discuss observed and potential consequences. Next, we describe how a transdisciplinary approach is required for actionable insight into the stewardship of social systems. Finally, we discuss some of the key ethical, scientific, and political challenges.

Communication Technology and Global Collective Behavior

Scholars have long sought to understand the mechanisms by which groups of individuals accomplish collective action (1, 2, 26). This phenomenon has been studied in a variety of disciplines, from anthropology, social psychology, sociology, political science, management, communication studies, economics, animal behavior, and sociobiology, to computer science, statistical physics, and the emerging domain of computational social science (27–35). These disciplines are largely differentiated by methods, scale of organization, and whether they study aspects of contemporary *Homo sapiens* society.

On an evolutionarily miniscule timescale, cultural and technological processes transformed our species' ecology (36). These changes that have transpired over this period have come about largely to solve issues at the scale of families, cities, and nations; only recently have cultural products begun to focus on solutions to worldwide problems and wellbeing. Our ability to detect and measure global challenges has coincided with an acceleration in the rate at which we are able to develop and adopt cheaply scalable communication technology.

Yet we lack the ability to predict how the technologies we adopt today will impact global patterns of beliefs and behavior tomorrow. Reliable prediction of social systems is among the

more elusive challenges in science (37). For instance, elections in countries such as the United States involve a discrete decision between two options and offer ample polling data—yet their outcomes remain difficult to predict (38). The key hurdle to predicting and managing emergent behavior is that social interactions and external feedback make it difficult, if not impossible, to reason about cross-scale dynamics through argument alone (i.e., these are complex adaptive systems) (25).

Scientists have confronted this type of problem before. The counterintuitive properties of emergent behavior frustrated early 20th century ethologists who reluctantly concluded that animal collectives such as flocking birds must employ telepathy to synchronize their harmonious short-term behaviors (1). To progress beyond these fanciful theories, researchers found ways to directly measure the collective dynamics of animal groups and developed approaches grounded in well-established sensory physiology and evolutionary theory (26, 39, 40). This body of literature has cataloged myriad ways in which collective functionality arises from natural selection shaping the behavioral rules that govern the actions and interactions of group members (41, 42). This research has highlighted that the remarkable capabilities of animal groups are not granted by supernatural forces but rather arise through the adaptation of collective behavior to ecological context (43, 44).

Collective animal behavior is one of many naturally occurring complex adaptive systems. Across the natural sciences, understanding and responding to the impact of human activity on complex systems are at the forefront of scientific inquiry. For example, in the last few decades it has become clear that population growth, technology, and overexploitation have had detrimental consequences on sustainability and ecosystem productivity (11, 13, 14). Earth scientists have responded by bridging disciplines and developing an applied approach aimed at providing regulators with information required for effective ecosystem stewardship. Brain science and medicine face similar challenges regarding how our psychological health and physical health are impacted by novel environmental conditions and substances. Evolutionary biology links conservation, medicine, epidemiology, and agriculture as they cope with impacts of rapidly changing selection landscapes (45).

By contrast, the long-term consequences of disturbance to human social dynamics remain unclear. For example, in the context of climate change there are strong arguments across disciplines suggesting that rapid behavioral change can bring about sustainability (16, 46, 47). At the same time, we cannot say whether a given communication technology will promote or prevent necessary changes from occurring. More generally, we lack the ability to foresee the externalities that communication technologies impose on aspects of human and ecosystem health and wellbeing. Below, we highlight four key ways in which recent changes to our social systems may have dramatically and unsustainably impacted social dynamics. Drawing on insights from a variety of academic disciplines, we describe how these changes are all but certain to have functional consequences at scale. Taken together, we argue that the changing functional properties of our global social network are unlikely to foster human wellbeing or ecological function and stability in the absence of evidence-based intervention.

Increased Scale of Human Social Networks. Perhaps the most obvious way in which human social networks differ from our ancestors and animal groups is in sheer scale. Our global social network of 7.8 billion people (3.6 billion of whom use social media) is distinct among macroscopic species. Among explanations for our large population size and geographic range is the agricultural revolution, in which humans domesticated crops and animals, paving the way for urbanization (but see refs. 48–50).

Connections between these groups formed states, nations, and the global social and economic network that now encompasses all but a few isolated groups (36, 51). Even language barriers are dissolving with global internet connectivity and effective machine translation. Cultural products, news, and information can spread far beyond their circumstance of origin. These remarkable changes to our social network size and structure and our institutions emerged over an extremely short time window of 12,000 y and well after the arrival of modern humans (48, 52).

The speed of recent changes to our society has largely precluded evolution by natural selection from altering our innate behavior and physiology in response. Hard-wired aspects of our individual and collective behavior are largely relics of earlier ecological and sociological contexts. Cultural evolution happens on a much faster timescale and has radically shaped collective human behavior (36, 51). This process has only accelerated, and our collective behavior now occurs in an environment that is defined by recent innovations in communication technology (e.g., social media, email, television) (53). While ideas for institutions and technology may be traced to individuals, their diffusion and shaping both arise from and alter collective, and historical, processes.

Expanding the scale of a collectively behaving system by eight orders of magnitude is certain to have functional consequences. Not only are societies at the scale of ours rare in the natural world; they also are often ecologically unstable where they do form (54). There are many possible challenges such large groups can face. Scarce resources, perhaps resulting from degraded commons or overpopulation, can cause intergroup or interindividual competition and war (55–57). Although there is evidence that shared commons can be sustainable, it is challenging to make them so—particularly at global scales (47).

Even if sufficient resources are available, changes in group size will have a host of functional consequences. Research in statistical physics and opinion dynamics demonstrates that group size can impact the tendency of collectives to settle on decisions (58, 59). Work from the collective intelligence literature suggests intermediate optimal group sizes in complex environments and highlights the difficulty of wise decision making in large groups (60, 61). Evolutionary mechanisms that encourage cooperation or coordination may be scale dependent, requiring institutions such as religion and governance to maintain these properties as group size increases (36, 62–64). Heterogeneous adoption of these institutions may further create conflict and erode cooperation (29, 65). In short, changes in scale alone have the potential to alter a group's ability to make accurate decisions, reach a clear majority, and cooperate.

Changes in Network Structure. The behavioral properties of a group arise not only from the number and properties of the individuals involved (i.e., the nodes of a social network) but also from the structure and temporal dynamics of the interactions between them (i.e., the edges). In other words, the same individuals arranged in a different network can exhibit different emergent behavior (66–68). Although offline networks from hunter-gatherers to urban dwellers bear structural similarities (69), the connectivity of technological social networks is starkly different (70).

Communication technologies allow people to interact more frequently and to do so with others from geographically distant

Bak-Coleman et al.
Stewardship of global collective behavior

PNAS | 3 of 10
https://doi.org/10.1073/pnas.2025764118

areas. Ties that span otherwise large network distances (i.e., long ties) can have profound consequences on the spread of disease and flow of information, including misinformation. For simple contagions, where a single interaction can lead to transmission (e.g., of disease), long ties can increase spreading (71, 72). Changes to simple contagions resulting from long ties online are among the easiest to model and reason about. As an example, online dating apps add long ties on sexual contact social networks—often by design, as they seek to connect strangers. This has the potential to increase disease burden even for those that do not use the services.

The spread of information and subsequent behavioral change often involves processes that go beyond simple contagion (73). While information may spread in a manner akin to disease, models must also account for how individuals integrate and adjust behavior, form opinions, and experience changes in emotion based on information from multiple sources (74, 75). Across disciplines, a host of interrelated models of information and behavior transmission have been developed, including complex contagion (computational social science), conformity (psychology, evolutionary anthropology), majority rule (political science, statistical physics), uses and gratifications (communication), and frequency-dependent learning (animal behavior) (28, 76–82). Virtually all of these models exhibit strong dependence on network structure. In many formulations, changes in network density, clustering, or the presence of influential individuals determine transmission dynamics. Such changes are unavoidable when groups adopt certain new communication technologies.

For most of our evolutionary past, individual *H. sapiens* may have maintained meaningful social contacts with, at most, hundreds of others and often far fewer (62, 69). Today, it is easy to connect and share information with thousands of other individuals on platforms such as Facebook, Instagram, and Twitter. More traditional forms of media such as TV, newspapers, and books allow individual authors and content creators to reach more people than were alive only a few thousand years ago. Highly connected individuals possess outsized influence, and it is unlikely that their centrality is solely related to being a producer of higher-quality information (83–86). Instead, their popularity may be a result of cumulative advantage or the tendency to evoke an emotional response (70, 87). Vested interests have taken advantage of new communication technology to spread misinformation, which partially explains why climate contrarians are overrepresented in nontraditional, digital media (88). In contexts where decisions depend upon accurate information about the world, these processes could undermine collective intelligence or promote dangerous behavior such as vaccine refusal (89, 90).

At a higher level of organization, our large population size combined with communication technology permits the development of novel network structures that were not possible historically. Macroscopic features of these structures, such as strong interconnectedness, long ties, and inequality of influence, drive many positive developments, such as transnational and transdisciplinary collaborations, rapid spread of scientific ideas, direct citizen engagement in science and politics, and overcoming isolation of individuals that do not fit in their local communities because of their beliefs and preferences (3, 30).

These structural features can also contribute to harmful phenomena: echo chambers and polarization, eroded trust in government, worldwide spread of local economic instabilities, global consequences of local electorate decisions, difficulty coordinating responses to pandemics, migrations driven by unreliable information about potential host countries, and others (70, 91, 92).

Novel large-scale structures can further impact the flow of information, altering the speed and accuracy with which information spreads (30, 93–96). Recent work suggests that network structural effects can lead to "information gerrymandering" that induces undemocratic outcomes whereby a majority of the electorate votes against the electorate's interest (97). These examples represent just a few of the many ways in which structure can impact collective functionality.

Information Fidelity and Correlation. While the structure and size of the global social network have changed, so too has the information that travels along its edges. Early human communication was largely biological (e.g., vocalizations, gestures, speech), relatively slow, and inherently noisy, allowing information to mutate and degrade as it moved throughout a network. Experimental and observational evidence suggests this natural decay allows influence from a given node to travel about three to four degrees of separation from the initiator (98, 99).

While noise, latency, and information decay are often viewed as unwanted in other areas of study, in collective systems they can serve several important functions. Noise can disrupt gridlock and promote cooperation (100), facilitate coherence (101), and improve detection of weak signals through phenomena akin to stochastic resonance (102). Evidence from fish schools revealed that noise and decay are important for preventing the spread of false alarms (39). Further, rapid information flows may overwhelm cognitive processes and yield less accurate decisions (103, 104). Through multiple iterations of high-fidelity transmission, communication technology allows information in tweets and articles to propagate beyond the three or four degrees of separation inherent to noisier forms of communication (83). Facsimiles of false information (e.g., misinformation and disinformation) can now spread across vast swaths of society without the risk of decay or fact checking along the way. Adding friction to this process has become one of the more promising approaches to reducing misinformation online (105).

Information is also increasingly cheap to produce and distribute. This eliminates barriers that may previously have functioned as filters on the type of information that is shared and alters the role of traditional informational gatekeepers such as journalists (106). On the one hand, this may make the sharing of information more egalitarian and promote the voices of historically disenfranchised groups; on the other hand, lowering such costs reduces incentives to produce high-quality and accurate information. This is exacerbated in contexts where trust, network structure, or other factors insulate public figures from fact checking and consequences of spreading false information (107–109). Anonymity online similarly permits the spread of low-quality information with minimal social cost and provides cover for bots brute-forcing a message onto a network (110).

As costs to inaccuracy decrease, individuals and institutions are better able to reap ideological and political benefits from outright lies (109). Portions of the society or networks repeatedly exposed to falsehood may normalize it or lack access to an information environment capable of sorting fact from fiction (107, 111, 112). The removal of filters that may have favored high-quality information, combined with rapid distribution of falsehood, may present one of the larger threats to human wellbeing when it comes to issues such as climate denial, vaccine refusal, treatment of minorities, and unfounded fears regarding the safety of genetically modified food.

Developments in media technology have reduced the granularity at which messages can be monetized in an information economy. Subscription-based models are receding as search engines, aggregator sites, social media platforms, and other innovations have created arenas of head-to-head competition among individual messages at the smallest scales of resolution. The unvarnished truth is no longer enough to prevail in the competition for attention (113). And that competition has become all of the more immediate as click-based advertising allows these microunits to be monetized directly and individually. New markets for pure misinformation emerge and thrive (114).

Innovations in the way we share information can have qualitative impacts as well—not only altering the rate, quantity, and fidelity of communication, but also fundamentally changing the types of information that can be stored in the first place (115). Changes to how information is stored and shared can alter and define power relationships. For example, the transition from oral to written history makes it possible to keep numerical records necessary for advanced commerce: debts can be recorded, taxes systematically extracted, and so forth. The advent of the printing press democratized not only who could own books, but also who could write them. The Internet has captured the long tail of human interests, allowing small groups of enthusiasts to find one another and document their passions in great detail. The advent of social media transferred the power to filter and screen content from professional editors to all of us, as we serve in an editorial capacity when we share information with our friends and thereby determine what they see (116). As technology develops, we will doubtless see other paradigm shifts. Being able to understand and predict the consequences of such shifts while, or even before, they are occurring must be a key focus of the study of human collective behavior.

Algorithmic Feedback. Inexpensive digital computing has reduced the cost of developing and implementing algorithms—mathematical recipes for manipulating information—and made them a pervasive aspect of our daily lives. Algorithms and artificial intelligence (AI) more specifically are used in many socially beneficial ways, from anticipating healthcare needs and making connections between potentially compatible individuals to regulating traffic and facilitating financial and policy decisions (117).

However, there is a growing concern regarding the impact of algorithmic decision making on individual and collective outcomes (118). For example, algorithms designed to filter, curate, and display the vast amount of information available online, combined with people's tendency to seek friendly social environments, may induce biases in perceived reality and contribute to societal polarization (119–122). Algorithms that aim to facilitate hiring, lending, healthcare, policing, and criminal justice may provide an illusion of objectivity while reinforcing human biases and creating feedback loops that further exacerbate injustice and inequality (123–125).

Algorithms designed to recommend information and products in line with supposed individual preferences can create runaway feedback wherein both the user's information preferences and subsequent exposure to content become more extreme over time (119, 126). Such path dependencies may have transformative effects, changing the preferences and values of the users themselves and leading to radicalization (127, 128). This may be reinforced by platforms recommending content based on the preferences of friends (129). Small fluctuations in initial popularity can drive differences in visibility and thus the "rich get richer" (130). For example, in a classic experiment, the popularity of all but the very best and worst songs was shown to be more related to stochastic early positive reception by other users than by their inherent quality (87).

Algorithms that recommend friends with similar beliefs introduce further complications. For example, highly followed Twitter users tend to receive many more new followers than less-followed users, in particular since Twitter began recommending users to follow in 2010 (131). This algorithmic change has increased the inequality in the number of followers between users—altering the overall network structure in ways that may exacerbate the spread of misinformation (83).

In sum, we are offloading our evolved information-foraging processes onto algorithms. But these algorithms are typically designed to maximize profitability, with often insufficient incentive to promote an informed, just, healthy, and sustainable society. Efforts to develop an appropriate scientific or ethical oversight and understanding are still in their infancy, and the black-box and proprietary nature of many algorithms slows down this progress (132). As a result, we have little insight into how the millions of seemingly minor algorithmic decisions that shape information flows every second might be altering our collective behavior.

Collective Behavior as a Crisis Discipline

Humanity faces global and existential threats including climate change, ecosystem degradation, and the prospect of nuclear war. We likewise face a number of other challenges that impact our wellbeing, including racism, disease, famine, and economic inequality. Our success at facing these challenges depends on our global social dynamics in a modern and technologically connected world. Given our evolved tendencies combined with the impact of technology and population growth, there is no reason to believe that human social dynamics will be sustainable or conducive to wellbeing if left unmanaged.

Online and offline forces that impact collective behavior and action are inextricable (133). Yet offline changes to how we share information may require years to percolate through the community, whereas changes in the digital world can be implemented and imposed in a matter of seconds. In this sense, online communication technology increases the urgency of stewardship while providing opportunities to enact evidence-based policies at scale. For these reasons, we expect that stewardship of social systems will require increased focus on digital technologies. However, we caution that online and offline dynamics cannot be disentangled and careful consideration of both will be necessary for identifying successful intervention strategies.

Given that the impacts of communication technology on patterns of behavior cross the lines that divide academic disciplines, a transdisciplinary synthesis and approach to managing our collective behavior are required. Between the complexity of our social systems, the specter of ongoing human suffering, and the urgency required to avert catastrophe, we must face these challenges in the absence of a complete model or full understanding (14, 134). In this way, the field of human collective behavior must join the ranks of other crisis disciplines such as medicine, conservation biology, and climate science (20).

Other crisis disciplines thrive on a close integration of observational, theoretical, and empirical approaches. Global climate models inform, and are informed by, experiments in the laboratory and the field. Mathematics describing disease dynamics suggest treatment paradigms in medicine, which can be tested and validated (135). Ecological models suggest strategies such as

Bak-Coleman et al.
Stewardship of global collective behavior

PNAS | 5 of 10
https://doi.org/10.1073/pnas.2025764118

establishing protected areas and using ecological cascades to manage deteriorating ecosystems (136). A similar approach can be adopted to study issues arising from communication technology.

For example, data-driven models of how information spreads may inform strategies to reduce political misinformation or anti-vaccine propaganda without requiring censorship. Similarly, modeling human interaction with recommendation algorithms may provide insight into best practices for detecting and deterring radicalization. Developing plausible mathematical theory will require integrating insight from scientists who rely on qualitative or mixed-methods approaches to study behavior online. Political communication research, in particular, has long described how alterations of networked communications technology appear to impact social movements, institutional politics, and political participation (133, 137).

A consolidated transdisciplinary approach to understanding and managing human collective behavior will be a monumental challenge, yet it is a necessary one. Given that algorithms and companies are already altering our global patterns of behavior for financial reasons, there is no safe hands-off approach. Below we chart a course for an applied, crisis-minded study of human collective behavior. We highlight some of the core challenges to doing so, issues requiring urgent attention, and necessary first steps.

Key Challenges and Future Directions

Stewardship will require incorporating our understanding of individual behavior with its emergent consequences at scale. Traditionally, fields such as psychology, social psychology, and behavioral economics have provided rich descriptions of individual behavior but have tended to study this behavior in experimental contexts with at most a few interacting individuals. By contrast, sociology, communication studies, science and technology studies, political science, and macroeconomics have measured or described patterns that occur at larger scales of organization using survey, ethnographic, and observational data, which can abstract away the underlying dynamics.

In the last few decades complexity science has begun to quantitatively link these scales of organization, generating a set of theoretical tools and frameworks for understanding how individual actions of interconnected agents give rise to social complexity (24, 42, 138). Incorporating a complex systems perspective is critical for understanding human behavior (24, 139, 140). Rigorous empirical tests of these models are still rare, which limits their usefulness for managing social dynamics.

Techniques adopted from the field of computational social science are well poised to bridge the gap between theory and measurement of collective behavioral processes (141, 142). Synchronous online experiments allow a detailed and controlled study of individuals interacting on social networks (143). These experiments can enable us to go beyond mathematically convenient but limited agent-based models and incorporate the richness and heterogeneity of individual behavior (97). Stewardship of collective processes will require an understanding of both individual motivations and their emergent consequences (144–146).

Moving from scientific to actionable insight will also require an understanding of law, public policy, systemic risk, and international relations. Our social systems are coupled to a variety of other tangible complex systems, including economies, supply chains of food and medicine, and utilities such as power grids. Proposed evidence-based policies will have to consider the risk that policy poses to the stability of other systems when communication technology interventions are applied at scale. At present, little such caution is exercised.

We should not expect to devise a single set of best practices that equitably addresses the totality of problems facing humanity. Often, solutions will instead focus on specific issues. Even in these cases, proposed solutions addressing a given issue in a given locality (i.e., vaccine misinformation in the United States) may have limited impact or even detrimental effects elsewhere. As with conservation biology and medicine, the stewardship of social systems inherently involves risk, trade-offs, and nonuniform benefits and costs (147). Scientific study of collective behavior may be able to provide a description of these features, yet questions of whether they should be adopted will often lie in the realms of the humanities and public policy.

If we hope to steward collective behavior, we need to find rapid ways to communicate research that avoid lengthy delays associated with peer review (148) to transmit basic research findings to those responsible for deploying interventions on timescales commensurate with evolving digital institutions (149). White papers aimed at regulators and journalists are common in climate science and recently played an important role in responding to electoral misinformation and communicating COVID-19–related research (107, 150). In lieu of peer review, multiinstitution and interdisciplinary collaboration provides a degree of error checking prior to publication. Subsequent to publication, rapid postpublication peer review on social media and other venues can substitute for slower formal mechanisms. For nontraditional methods of scientific communication to succeed, institutions and universities must find ways to incorporate these contributions into funding, hiring, and promotion decisions.

We suggest that there is an urgent need for an equivalent of the Hippocratic oath for anyone studying or intervening into collective behavior, whether from within academia or from within social media companies and other tech firms. Decisions that impact the structure of society should not be guided by voices of individual stakeholders but instead by values such as nonmaleficence, benevolence, autonomy, and justice. To the extent that values and needs vary across individuals and cultural contexts, decisions will require careful deliberation or context-specific solutions (151). Our approach must further consider the impact on those that lack access to communication technology, as interventions that improve digital life may lead to inequity offline. For instance, online vaccination or electoral registration programs risk relative disenfranchisement of groups that cannot take advantage of them. In the absence of a globally held normative framework for deciding what constitutes healthy societies or desirable sociotechnical interactions, it may be difficult to even agree on what ethical stewardship might entail. Developing ethical standards that consider the range of cultural perspectives, histories, and traditions impacted by communication technologies is no easy task.

Proposed interventions must consider direct ethical obligations toward individuals (e.g., freedom of speech, autonomy), nonhuman beings, and the environment, as well as more generic obligations toward society at large (e.g., limiting disease burden, establishing food security). The relevant sciences will help us to map out how various technical innovations and applications impact society as a whole, as well as distinct segments of society such as marginalized groups. Armed with this information, regulators and the public can make ethical and political choices about how—and whether—to proceed. These decisions should be as empirically informed as possible and must be rooted in needs, values, and concerns. As value priorities may differ across time

and cultural contexts, implementations that account for this variability must be considered.

As most communication technology is privately owned, the ability to study its impact, much less enact evidence-based policy, is constrained by the willingness of companies to cooperate. They may use insight from collective behavior to instead increase profits or simply refuse to act. For instance, there is evidence to suggest that a subset of users is engaged by misinformation, as well as emotionalized and moralized content (70, 83, 152, 153). From a company's perspective, this content retains users who provide economic value and its removal may not be economically favorable or even viable. This raises the possibility that some business models may be fundamentally incompatible with a healthy society (154). In such cases, identified interventions may not be in the interests of either the company or the users that prefer it. We anticipate these contexts to be particularly challenging and require ample evidence of harm to be presented to the public and regulators. Producing such evidence will be substantially more difficult if companies have a heavy hand in the production, funding, and communication of research (155, 156). Overall, profitable approaches promoting healthy online interaction—should they exist—will be easier to implement. Ongoing crises in digital spaces have generated substantial momentum and insight toward stewardship. Misinformation poses grave threats including the spread of conspiracy theories, rejection of recommended public health interventions, subversion of democratic processes, and even genocide (90, 107, 157, 158). In response, communication scholars have adopted decades-old theories of propaganda and mass communication to understand disinformation and media manipulation online (154, 159, 160). Social psychologists have developed "nudges" to encourage more discerning sharing of content online (105). More rapid responses to misinformation have come about through collaborations between social and computer scientists (107, 161).

Beyond misinformation, understanding the consequences of dark patterns—user interface design that guides people against their interests—and opaque algorithms is now a major topic of research. Owing to a near-complete lack of transparency from tech companies, description and measurement are critical first steps (162, 163). Despite the opacity, research has revealed how algorithms lead users to radical or age-inappropriate content (128, 164), exacerbate disparities in health (123), and increase bias in policing (124). Unfortunately, misinformation, algorithms, dark patterns, and other issues arise at a rate far greater than they can be adequately characterized, much less addressed.

The challenges that arise from new communication technologies will require identifying common classes of problems and associated solutions. This is the approach adopted in conservation biology, where crises observed across multiple contexts (e.g., ecosystem collapse, mismanaged commons) lead to an understanding of multiscaled dynamics yielding solutions that can be tailored to given sociological and ecological contexts (8, 47, 64). While this is a starting point, it is by no means a panacea.

Proposed solutions in conservation biology and other crisis disciplines, no matter how elegant, are often stymied by inability to convert workable solutions into large-scale behavioral change. Clever solutions aimed at social system stewardship will face similar challenges. In this regard, social media's influence provides a unique source of both risk and opportunity. Changes to a few lines of code can impact global behavioral processes. Such changes are ongoing, with or without scientific guidance. In the absence of evidence-informed policy recommendations, we should not expect the emergent consequences to be stabilizing or even beneficial. Collective behavior provides a framework for stewardship of social systems, not by supplanting other fields, but by stitching together disparate disciplines with a common goal.

Summary

Human collective dynamics are critical to the wellbeing of people and ecosystems in the present and will set the stage for how we face global challenges with impacts that will last centuries (14, 15, 64). There is no reason to suppose natural selection will have endowed us with dynamics that are intrinsically conducive to human wellbeing or sustainability. The same is true of communication technology, which has largely been developed to solve the needs of individuals or single organizations. Such technology, combined with human population growth, has created a global social network that is larger, denser, and able to transmit higher-fidelity information at greater speed. With the rise of the digital age, this social network is increasingly coupled to algorithms that create unprecedented feedback effects.

Insight from across academic disciplines demonstrates that past and present changes to our social networks will have functional consequences across scales of organization. Given that the impacts of communication technology will transcend disciplinary lines, the scientific response must do so as well. Unsafe adoption of technology has the potential to both threaten wellbeing in the present and have lasting consequences for sustainability. Mitigating risk to ourselves and posterity requires a consolidated, crisis-focused study of human collective behavior.

Such an approach can benefit from lessons learned in other fields, including climate change and conservation biology, which are likewise required to provide actionable insight without the benefit of a complete understanding of the underlying dynamics. Integrating theoretical, descriptive, and empirical approaches will be necessary to bridge the gap between individual and large-scale behavior. There is reason to be hopeful that well-designed systems can promote healthy collective action at scale, as has been demonstrated in numerous contexts including the development of open-sourced software, curating Wikipedia, and the production of crowd-sourced maps (165, 166). These examples not only provide proof that online collaboration can be productive, but also highlight means of measuring and defining success. Research in political communications has shown that while online movements and coordination are often prone to failure, when they succeed, the results can be dramatic (137). Quantifying benefits of online interaction, and limitations to harnessing these benefits, is a necessary step toward revealing the conditions that promote or undermine the value of communication technology.

A consolidated study of human collective behavior will be limited to providing mechanistic insight into the consequences of changes to our social system and potential solutions. The ethical issues raised by stewardship of social systems, like those associated with ecological systems, will require input from philosophy, public policy, and disciplines across the humanities (147). There is no viable hands-off approach. Inaction on the part of scientists and regulators will hand the reins of our collective behavior over to a small number of individuals at for-profit companies. Despite the scientific and ethical challenges, the risks of inaction both in the present and for future generations necessitate stewardship of collective behavior.

1. E. Selous, Thought transference (or what?) in birds. *Nature* **129**, 263 (1932).
2. Aristotle, *Politics* (Batoche Books, 1999).
3. M. Granovetter, The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
4. H. Blumer, Social problems as collective behavior. *Soc. Probl.* **18**, 298–306 (1971).
5. I. D. Couzin, J. Krause, Self-organization and collective behavior in vertebrates. *Adv. Stud. Behav.* **32**, 1–75 (2003).
6. T. Walker, D. Sesko, C. Wieman, Collective behavior of optically trapped neutral atoms. *Phys. Rev. Lett.* **64**, 408–411 (1990).
7. R. A. Bentley, M. J. O'Brien, Collective behaviour, uncertainty and environmental change. *Phil. Trans. R. Soc. A.* **373**, 20140461 (2015).
8. S. A. Levin, Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* **1**, 431–436 (1998).
9. R. M. May, S. A. Levin, G. Sugihara, Complex systems: Ecology for bankers. *Nature* **451**, 893–895 (2008).
10. M. Scheffer *et al.*, Anticipating critical transitions. *Science* **338**, 344–348 (2012).
11. P. J. Crutzen, W. Steffen, How long have we been in the Anthropocene era? *Clim. Change* **61**, 251–257 (2003).
12. W. Steffen, P. J. Crutzen, J. R. McNeill. The Anthropocene: Are humans now overwhelming the great forces of nature. *Ambio* **36**, 614–621 (2007).
13. A. D. Barnosky *et al.*, Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
14. W. Steffen *et al.*, Trajectories of the earth system in the Anthropocene. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8252–8259 (2018).
15. S. Carattini, S. Levin, A. Tavoni, Cooperation in the climate commons. *Rev. Environ. Econ. Pol.* **13**, 227–247 (2019).
16. I. M. Otto *et al.*, Social tipping dynamics for stabilizing Earth's climate by 2050. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2354–2365 (2020).
17. J. J.Van Bavel *et al.*, Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
18. J. Zarocostas, How to fight an infodemic. *Lancet* **395**, 676 (2020).
19. T. Hobbes, *Leviathan* (Penguin Books, Baltimore, MD, 1968).
20. M. E. Soulé, What is conservation biology? *Bioscience* **35**, 727–734 (1985).
21. D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
22. D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
23. A. L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
24. C. Schill *et al.*, A more dynamic understanding of human behaviour for the Anthropocene. *Nat. Sustain.* **2**, 1075–1082 (2019).
25. J. Holland, Complex adaptive systems. *A New Era in Computation* **121**, 17–30 (1992).
26. D. V. V. Radakov, *Schooling in the Ecology of Fish* (John Wiley & Sons, 1973).
27. D. J. Watts, A twenty-first century science. *Nature* **445**, 489 (2007).
28. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (University of Chicago Press, 1985).
29. C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
30. D. Centola, M. Macy, Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734 (2007).
31. M. De Condorcet, *Essay on the Application of Analysis to the Probability of Majority Decisions* (Imprimerie Royale, Paris, France, 1785).
32. L. Conradt, T. J. Roper, Group decision-making in animals. *Nature* **421**, 155 (2003).
33. R. E. Hertwig, U. E. Hoffrage, *Simple Heuristics in a Social World* (Oxford University Press, 2013).
34. W. Hoppitt, K. N. Laland, *Social Learning: An Introduction to Mechanisms, Methods, and Models* (Princeton University Press, 2013).
35. M. O. Jackson, *Social and Economic Networks* (Princeton University Press, 2010).
36. J. Henrich, *The Secret of Our Success* (Princeton University Press, Princeton, NJ, 2017).
37. J. M. Hofman, A. Sharma, D. J. Watts, Prediction and explanation in social systems. *Science* **355**, 486–488 (2017).
38. M. S. Lewis-Beck, M. Stegmaier, "Election forecasting, scientific approaches" in *Encyclopedia of Social Network Analysis and Mining*, R. Alhajj, J. Rokne, Eds. (Springer New York, New York, NY, 2016), pp. 1–8.
39. S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, I. D. Couzin, Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 201420068 (2015).
40. A. Strandburg-Peshkin, D. R. Farine, I. D. Couzin, M. C. Crofoot, Shared decision-making drives collective movement in wild baboons. *Science* **348**, 1358–1361 (2015).
41. D. J. T. Sumpter, The principles of collective animal behaviour. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **361**, 5–22 (2006).
42. I. Couzin, Collective minds. *Nature* **445**, 715 (2007).
43. D. Sumpter, *Collective Animal Behavior* (Princeton University Press, Princeton, NJ, ed. 1, 2010), vol. 1.
44. D. M. Gordon, The ecology of collective behavior in ants. *Annu. Rev. Entomol.* **64**, 35–50 (2019).
45. S. P. Carroll *et al.*, Applying evolutionary biology to address global challenges. *Science* **346**, 1245993 (2014).
46. T. K. Rudel, *Shocks, States, and Sustainability: The Origins of Radical Environmental Reforms - Oxford Scholarship* (Oxford University Press, Oxford, UK, 2019).
47. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 2015).
48. J. T. Stock, Are humans still evolving? Technological advances and unique biological characteristics allow us to adapt to environmental stress. Has this stopped genetic evolution? *EMBO Rep.* **9** (suppl. 1), 51–54 (2008).
49. P. Turchin, T. E. Currie, E. A. L. Turner, S. Gavrilets, War, space, and the evolution of Old World complex societies. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16384–16389 (2013).
50. H. J. Zahid, E. Robinson, R. L. Kelly, Agriculture, population growth, and statistical analysis of the radiocarbon record. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 931–935 (2016).
51. J. Henrich, *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous* (Farrar, Straus and Giroux, New York, NY, 2020), vol. 1.
52. K. Sterelny, From hominins to humans: How sapiens became behaviourally modern. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **366**, 809–822 (2011).
53. P. J. Boczkowski, The mutual shaping of technology and society in videotex newspapers: Beyond the diffusion and social shaping perspectives. *Inf. Soc.* **20**, 255–267 (2004).
54. M. W. Moffett, Supercolonies of billions in an invasive ant: What is a society? *Behav. Ecol.* **23**, 925–933 (2012).

8 of 10 | PNAS
https://doi.org/10.1073/pnas.2025764118

Bak-Coleman et al.
Stewardship of global collective behavior

55 J. L. Brown, Optimal group size in territorial animals. *J. Theor. Biol.* **95**, 793–810 (1982).

56 G. Hardin, The Tragedy of the Commons. *Science* **162**, 1243–1248 (1968).

57 C. R. Ember, M. Ember, Resource unpredictability, mistrust, and war. *J. Conflict Resolut.* **36**, 242–262 (1992).

58 S. Galam, Contrarian deterministic effects on opinion dynamics: "The hung elections scenario." *Phys. Stat. Mech. Appl.* **333**, 453–460 (2004).

59 S. Gekle, L. Peliti, S. Galam, Opinion dynamics in a three-choice system. *Euro. Phys. J. B* **45**, 569–575 (2005).

60 A. B. Kao, I. D. Couzin, Decision accuracy in complex environments is often maximized by small group sizes. *Proc. Biol. Sci.* **281**, 20133305 (2014).

61 M. Galesic, D. Barkoczi, K. Katsikopoulos, Smaller crowds outperform larger crowds and individuals in realistic task conditions. *Decision* **5**, 1–15 (2018).

62 R. I. M. Dunbar, Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **22**, 469–493 (1992).

63 A. R. Tilman, A. K. Dixit, S. A. Levin, Localized prosocial preferences, public goods, and common-pool resources. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5305–5310 (2019).

64 W. Barfuss, J. F. Donges, V. V. Vasconcelos, J. Kurths, S. A. Levin, Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12915–12922 (2020).

65 M. Casari, C. Tagliapietra, Group size in social-ecological systems. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2728–2733 (2018).

66 D. Lazer, A. Friedman, The network structure of exploration and exploitation. *Adm. Sci. Q.* **52**, 667–694 (2007).

67 T. N. Wisdom, X. Song, R. L. Goldstone, Social learning strategies in networked groups. *Cognit. Sci.* **37**, 1383–1425 (2013).

68 D. Barkoczi, M. Galesic, Social learning strategies modify the effect of network structure on group performance. *Nat. Commun.* **7**, 13109 (2016).

69 C. L. Apicella, F. W. Marlowe, J. H. Fowler, N. A. Christakis, Social networks and cooperation in hunter-gatherers. *Nature* **481**, 497–501 (2012).

70 W. J. Brady et al., Emotion shapes the diffusion of moralized content in social networks. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7313–7318 (2017).

71 J. Chan, A. Ghose, Internet's dirty secret: Assessing the impact of online intermediaries on HIV transmission. *MIS Quarterly*, **38**, 955–976 (2012).

72 J. J. Lehmiller, M. Ioerger, Social networking smartphone applications and sexual health outcomes among men who have sex with men. *PloS One* **9**, e86603 (2014).

73 W. J. Brady, M. J. Crockett, J. J. Van Bavel, The MAD Model of Moral Contagion: The role of motivation, attention and design in the spread of moralized content. *Perspect. Psychol. Sci.* **15**, 978–1010 (2020).

74 M. Kimura, K. Saito, "Tractable models for information diffusion in social networks" in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, J. Furnkranz, Ed. (Springer-Verlag, 2006), vol. 4213, pp. 259–271.

75 E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, "The role of social networks in information diffusion" in *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web* (ACM Press, New York, NY, 2012), pp. 519–528.

76 D. Centola, An experimental study of homophily in the adoption of health behavior. *Science* **334**, 1269–1273 (2011).

77 S. E. Asch, Opinions and social pressure. *Sci. Am.* **193**, 31–35 (1955).

78 J. G. Heinberg, Theories of majority rule. *Am. Polit. Sci. Rev.* **26**, 452–469 (1932).

79 K. N. Laland, Social learning strategies. *Anim. Learn. Behav.* **32**, 4–14 (2004).

80 P. L. Krapivsky, S. Redner, Dynamics of majority rule in two-state interacting spin systems. *Phys. Rev. Lett.* **90**, 238701 (2003).

81 M. W. Feldman, L. L. Cavalli-Sforzatt, Cultural and biological evolutionary processes: Gene-culture disequilibrium. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1604–1607 (1984).

82 T. E. Ruggiero, Uses and gratifications theory in the 21st century. *Mass Commun. Soc.* **3**, 3–37 (2000).

83 S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **359**, 1146–1151 (2018).

84 J. Becker, D. Brackbill, D. Centola, Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5070–E5076 (2017).

85 A. V. Banerjee, A. Chandrasekhar, E. Duflo, M. O. Jackson, Gossip: Identifying central individuals in a social network (2014). https://www.nber.org/papers/w20422. Accessed 10 June 2021.

86 C. O'Connor, J. O. Weatherall, "Modeling how false beliefs spread" in *The Routledge Handbook of Political Epistemology*, M. Hannon, J. de Ridder, Eds. (Routledge, 2021), pp. 203–213.

87 M. J. Salganik et al., Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).

88 A. M. Petersen, E. M. Vincent, A. L. R. Westerling, Discrepancy in scientific authority and media visibility of climate change scientists and contrarians. *Nat. Commun.* **10**, 3966 (2019).

89 A. Archer, A. Cawston, B. Matheson, M. Geuskens, Celebrity, democracy, and epistemic power. *Perspect. Polit.* **18**, 1–16 (2019).

90 K. Koltai, Vaccine information seeking and sharing: HOW private Facebook groups contributed to the anti-vaccine movement online. *AoIR Selected Papers of Internet Research* **10**, AoIR2020 (2020).

91 V. Narayanan et al., Polarization, partisanship and junk news consumption over social media in the US. arXiv [Preprint] (2018). https://arxiv.org/abs/1803.01845v1 (Accessed 10 June 2021).

92 S. Guriev, N. I. Melkinov, E. Zhuravskaya, Knowledge is power: Mobile internet, government confidence, and populism. *VOXEU CEPR* (2019). https://voxeu.org/article/mobile-internet-government-confidence-and-populism. Accessed 10 June 2021.

93 J. Pallavicini, B. Hallsson, K. Kappel, Polarization in groups of Bayesian agents. *Synthese* **198**, 1–55 (2018).

94 K. J. S. Zollman, Social network structure and the achievement of consensus. *Polit. Philos. Econ.* **11**, 26–44 (2012).

95 R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence: Models, analysis and simulation. *JASSS* **5**, 3/2 (2002).

96 G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents. *Adv. Complex Syst.* **03**, 87–98 (2000).

97 A. J. Stewart et al., Information gerrymandering and undemocratic decisions. *Nature* **573**, 117–121 (2019).

98 N. A. Christakis, J. H. Fowler, The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).

99 M. Moussaïd, S. M. Herzog, J. E. Kämmer, R. Hertwig, Reach and speed of judgment propagation in the laboratory. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4117–4122 (2017).

100 H. Shirado, N. A. Christakis, Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* **545**, 370–374 (2017).

101 C. A. Yates et al., Inherent noise can facilitate coherence in collective swarm motion. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5464–5469 (2009).

102 J. F. Lindner, B. K. Meadows, W. L. Ditto, M. E. Inchiosa, A. R. Bulsara, Array enhanced stochastic resonance and spatiotemporal synchronization. *Phys. Rev. Lett.* **75**, 3–6 (1995).

103 L. Chittka, P. Skorupski, N. E. Raine, Speed–accuracy tradeoffs in animal decision making. *Trends Ecol. Evol.* **24**, 400–407 (2009).

104 B. Bago, D. G. Rand, G. Pennycook, Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *J. Exp. Psychol. Gen.* **149**, 1608–1613 (2020).

105 G. Pennycook et al., Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595 (2021).

106 B. A. Williams, M. X. Delli Carpini, Monica and Bill all the time and everywhere. *Am. Behav. Sci.* **47**, 1208–1230 (2004).

107 Election Integrity Partnership, "The long fuse: Misinformation and the 2020 election" (Tech. Rep., Center for an Informed Public, Digital Forensic Research Lab, Graphika, & Stanford Internet Observatory, Stanford Digital Repository, Stanford, CA 2021).

108 W. L. Bennett, B. Pfetsch, Rethinking political communication in a time of disrupted public spheres. *J. Commun.* **68**, 243–253 (2018).

109 R. M. Entman, N. Usher, Framing in a fractured democracy: Impacts of digital technology on ideology, power and cascading network activation. *J. Commun.* **68**, 298–308 (2018).

110 T. Marlow, S. Miller, J. T. Roberts. Bots and online climate discourses: Twitter discourse on President Trump's announcement of U.S. withdrawal from the Paris Agreement. *Clim. Pol.*, 10.1080/14693062.2020.1870098 (2021).

111 D. M. J. Lazer et al., The science of fake news. *Science* **359**, 1094–1096 (2018).

112 G. Pennycook, D. G. Rand, *Research Note: Examining False Beliefs about Voter Fraud in the Wake of the 2020 Presidential Election* (Harvard Kennedy School Misinformation Review, 2021).

113 J. D. West, C. T. Bergstrom, Misinformation in and about science. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1912444117 (2021).

114 S. Subramanian, Inside the Macedonian fake-news complex. *Wired*, 15 February 2017. https://www.wired.com/2017/02/veles-macedonia-fake-news/. Accessed 16 June 2021.

Bak-Coleman et al.
Stewardship of global collective behavior

PNAS | 9 of 10
https://doi.org/10.1073/pnas.2025764118

115 E. Szathmáry, J. M. Smith, The major evolutionary transitions. *Nature* **374**, 227–232 (1995).

116 C. Bergstrom, J. West, *Calling Bullshit: The Art of Skepticism in a Data-Driven World* (Random House, New York, NY, ed. 1, 2020).

117 Z. R. Shi, C. Wang, F. Fang, Artificial intelligence for social good: A survey. arXiv [Preprint] (2020). https://arxiv.org/abs/2001.01818v1 (Accessed 10 June 2021).

118 I. Rahwan, Society-in-the-loop: Programming the algorithmic social contract. *Ethics Inf. Technol.* **20**, 5–14 (2018).

119 T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, J. A. Konstan, "Exploring the filter bubble" in *Proceedings of the 23rd International Conference on World Wide Web - WWW '14* (ACM Press, New York, NY, 2014), pp. 677–686.

120 E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).

121 E. Bozdag, Bias in algorithmic filtering and personalization. *Ethics Inf. Technol.* **15**, 209–227 (2013).

122 B. Toff, R. K. Nielsen, "I just google it": Folk theories of distributed discovery. *J. Commun.* **68**, 636–657 (2018).

123 Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).

124 K. Lum, W. Isaac, To predict and serve? *Significance* **13**, 14–19 (2016).

125 C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Random House, New York, NY, ed. 1, 2016), vol. 1.

126 J. A. Evans, Electronic publication and the narrowing of science and scholarship. *Science* **321**, 395–399 (2008).

127 M. Alfano, J. A. Carter, M. Cheong, Technological seduction and self-radicalization. *J. Am. Philos. Assoc.* **4**, 298–322 (2018).

128 M. H. Ribeiro, R. Ottoni, R. West, V. A. F. Almeida, W. M. W. Meira, "Auditing radicalization pathways on YouTube" in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, Inc., New York, NY, 2020), vol. 11, pp. 131–141.

129 S. Deng, L. Huang, G. Xu, X. Wu, Z. Wu, On deep learning for trust-aware recommendations in social networks. *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 1164–1177 (2017).

130 R. K. Merton, The Matthew effect in science. *Science* **159**, 55–63 (1968).

131 J. Su, A. Sharma, S. Goel, "The effect of recommendations on network structure" in *Proceedings of the 25th International Conference on World Wide Web - WWW '16* (ACM Press, New York, NY, 2016). pp. 1157–1167.

132 J. Stoyanovich, J. J. Van Bavel, T. V. West, The imperative of interpretable machines. *Nat. Mach. Intel.* **2**, 197–199 (2020).

133 B. Bimber, A. J. Flanagin, C. Stohl, *Collective Action in Organizations: Interaction and Engagement in an Era of Technological Change* (Cambridge University Press, 2012).

134 J. Rockström *et al.*, A roadmap for rapid decarbonization. *Science* **355**, 1269–1271 (2017).

135 R. Birger, R. Kouyos, J. Dushoff, B. Grenfell, Modeling the effect of HIV coinfection on clearance and sustained virologic response during treatment for hepatitis C virus. *Epidemics* **12**, 1–10 (2015).

136 D. Fortin *et al.*, Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. *Ecology* **86**, 1320–1330 (2005).

137 H. Margetts, P. John, S. Hale, T. Yasseri, *Political Turbulence* (Princeton University Press, Princeton, NJ, ed. 1, 2016), vol. 1.

138 N. Goldenfeld, L. P. Kadanoff, Simple lessons from complexity. *Science* **284**, 87–89 (1999).

139 J. H. Miller, S. E. Page, *Complex Adaptive Systems* (Princeton University Press, Princeton, NJ, 2007).

140 J. Duffy, J. M. Epstein, R. Axtell, Growing artificial societies: Social science from the bottom up. *South. Econ. J.* **64**, 791 (1998).

141 D. J. Watts, "Computational social science: Exciting progress and future directions" in *Frontiers of Engineering* (National Academies Press, Washington, DC, 2013).

142 D. M. J. Lazer *et al.*, Computational social science: Obstacles and opportunities. *Science* **369**, 1060–1062 (2020).

143 N. Paton, A. Almaatouq, *Empirica: Open-Source, Real-Time, Synchronous, Virtual Lab Framework* (Zenodo, 2018).

144 X. Chen, S. C. J. Sin, Y. L. Theng, C. S. Lee, "Why do social media users share misinformation?" in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (Institute of Electrical and Electronics Engineers Inc., New York, NY, 2015), **vol. 2015**, pp. 111–114.

145 E. A. Rosa, O. Renn, A. M. McCright, *The Risk Society Revisited: Social Theory and Risk Governance on JSTOR* (Temple University Press, Philadelphia, PA, ed. 1, 2014).

146 M. Schlüter *et al.*, A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecol. Econ.* **131**, 21–35 (2017).

147 F. S. Chapin *et al.*, Ecosystem stewardship: Sustainability strategies for a rapidly changing planet. *Trends Ecol. Evol.* **25**, 241–249 (2010).

148 D. S. Himmelstein, K. Powell, Analysis for "the history of publishing delays" blog post v1.0 (2016). https://zenodo.org/record/45516#.YLld9japHlw. Accessed 1 February 2021.

149 H. Else, How a torrent of COVID science changed research publishing - in seven charts. *Nature* **588**, 553 (2020).

150 National Academies, Societal experts action network. https://www.nationalacademies.org/our-work/societal-experts-action-network#sl-three-columns-d2bc460d-5bb3-41ce-991f-80e4dd0bdae8. Accessed 1 February 2021.

151 S. H. Schwartz, An overview of the Schwartz theory of basic values. *Online Read. Psychol. Culture* **2**, 1–20 (2012).

152 G. Pennycook, D. G. Rand, Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *J. Pers.*, in press.

153 J. J. Van Bavel, A. Pereira, The partisan brain: An identity-based model of political belief. *Trends Cognit. Sci.* **22**, 213–224 (2018).

154 Y. Benkler, R. Farris, H. Roberts, *Network Propaganda* (Oxford University Press, 2018), vol. 1.

155 C. O'Connor, J. O. Weatherall, Scientific polarization. *Eur. J. Phylos. Sci.* **8**, 855–875 (2018).

156 N. Oreskes, E. M. Conway, Defeating the merchants of doubt. *Nature* **465**, 686–687 (2010).

157 J. Whitten-Woodring, M. S. Kleinberg, A. Thawnghmung, M. T. Thitsar, Poison if you don't know how to use it: Facebook, democracy, and human rights in Myanmar. *Int. J. Press Politics* **25**, 407–425 (2020).

158 N. Velásquez *et al.*, Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. *In Review* (2020). https://www.researchsquare.com/article/rs-110371/v1. Accessed 10 June 2021.

159 B. Silverstein, Toward a science of propaganda. *Polit. Psychol.* **8**, 49 (1987).

160 J. Donovan, B. Friedberg, "Source hacking media manipulation in practice executive summary" (Tech. Rep., Data & Society, 2019).

161 J. Kaiser *et al.*, Mail-in voter fraud: Anatomy of a disinformation campaign. Berkman Klein Center (2020). https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020. Accessed 10 June 2021.

162 S. Baluja *et al.*, "Video suggestion and discovery for you tube: Taking random walks through the view graph" in *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08* (ACM Press, New York, NY, 2008), pp. 895–904.

163 A. Mathur *et al.*, "Dark patterns at scale: Findings from a crawl of 11K shopping websites." in *Proceedings of the ACM on Human-Computer Interaction* (ACM, 2019), vol. 3.

164 K. Papadamou *et al.*, "Disturbed Youtube for kids: Characterizing and detecting inappropriate videos targeting young children" in *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020* (AAAI Press, Palo Alto, CA, 2020), pp. 522–533.

165 A. Kittur, R. E. Kraut, "Harnessing the wisdom of crowds in Wikipedia: Quality through coordination" in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* (ACM Press, New York, NY, 2008), pp. 37–46.

166 C. Prandi, P. Salomoni, S. Mirri, "Mpass: Integrating people sensing and crowdsourcing to map urban accessibility" in *2014 IEEE 11th Consumer Communications and Networking Conference, CCNC 2014* (IEEE Computer Society, 2014), pp. 591–595.

10 of 10 | **PNAS**
https://doi.org/10.1073/pnas.2025764118

Bak-Coleman et al.
Stewardship of global collective behavior