# 20

# Reflections on Collaboration

## A Roadmap for Future Research

Paul F. M. J. Verschure

This project has examined the nature and dynamics of collaboration, asking what it is, how it is realized, and what its underlying processes are. The tentative answers to these questions form a broad and partially contradictory panoply of positions expressing various levels of optimism about humans' capability to collaborate. To enable more definitive answers, we need to advance our understanding and practice of collaboration. Here, I outline such an endeavor, starting with reflections on the high-level questions of what collaboration is and how we can know its underlying principles. Thereafter, I sketch out some of the profound questions that need to be addressed with respect to the human substrate of collaboration and its organization and a collection of outstanding questions that emerged from this project. Humanity's global challenges require global collaboration, and I hope we will live up to our collective responsibility to tackle these issues, starting with placing our understanding and practice of collaboration on a solid transdisciplinary scientific foundation.

The concept of collaboration has transformed from its pragmatic Latin origins of *collaborare,* compounding *com-* (together) and *laborare* (to work) into an ambiguous and vague form that encompasses attributes such as aspiration, valued relationships, treason, moral judgment, and injustice (see Chapters 3 and 4; Digeser 2022). We began with a working definition that viewed collaboration as *cooperation between agents toward mutually constructed goals* but found that such a tight link to goals is unnecessary. Goals do not necessarily shape collaboration, nor do they need to be mutually constructed. Humans can collaborate for the sake of collaboration; goals might not be the same for each individual in a collective given sub-goals, and goals can change due to the collaboration itself. How, then, should collaboration be distinguished from cooperation?

Our discussions brought the original notion of a singular external goal as the main driver of a monolithic collective process into question and pointed, instead, to the goal-oriented nature of collaboration as a post hoc interpretation of the emergent structure of the dynamics of a collaborating collective but not as its exclusive cause. Moreover, given that collaboration is, by necessity,

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

future oriented, a more inclusive definition would be a *collective action driven by a co-constructed imagined future embedded in a shared worldview* (Chapters 4, 5, and 17). This definition places cooperation and collaboration on a continuum of the information required from proximal external physical cues, driving cooperation to distal cognitive constructs defining collaboration. Consider the chemo-mechanical signal exchange that defines ant cooperation versus a community of fate that can make humans collaborate to strive toward abstract objectives such as justice, freedom, or various malignant visions. Viewing collaboration in this manner resolves the ambiguity between these constructs and shows they can co-exist within a collective of agents, creating sub-structures and synergies. The notion of the cooperation-collaboration continuum opens new avenues of research and for shaping human collaboration in more heterogeneous and diverse dynamic structures.

Scientific progress depends on identifying effective paradigms to define the practice of research. Elinor Ostrom's work on managing common pool resources is an outstanding example of this in economics. It shows that humans can effectively and universally manage commons without needing external top-down forces of the state or the market once organized properly through rules, norms, and strategies or "institutions." As with all models, complexity needs to be compressed to achieve epistemic parsimony. Ostrom's model is effective because it addresses the specific case of small-scale groups managing physical resources to avoid a tragedy of the commons under an assumption of collaboration (Ostrom 2015). This authoritative model might give the impression that goals, operationalized as the management of physical resources, are central to human collaboration, notwithstanding the speculative broadening toward intellectual property and software. Yet, the model's scalability in shaping human collaboration to address global challenges is not obvious despite its suggestion of polycentric governance (Chapter 13). It also stands in a tense relation to global challenges by providing a rationale for special interest and neoliberal exploitation. Hence, being decoupled from a moral reference, the model will not automatically be sufficient to avoid a global tragedy of the commons. Ironically, it can directly contribute to it by providing an argument for the core value of neoliberalism: deregulation. Despite their utility, insisting on economic models would propagate a perspective on human collaboration organized around generic operational notions such as property, rational choice, incentives, and their derived constructs into domains of knowledge and experience where they might not fit. This Forum had to navigate this implicit form of what has been dubbed economics imperialism. The Ostrom principle has been useful yet we must recognize and place the multidimensional phenomenon of collaboration at the center of our efforts if we are to understand and mobilize it to address common global challenges.

Despite the contested nature of the meaning of collaboration, we converged on one point: its future-oriented directedness or intentionality. This, however, creates a fundamental challenge for a scientific treatment of this phenomenon

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

as it falls outside the epistemological approach taken by modern third-person-oriented Western science. The intentionality of behavior, or its *telos*, was removed from the study of mind and behavior during the early twentieth century as the legacy of the behaviorist revolution and its empty organism model. Indeed, by linking behavior to externally defined factors, such as the commons and economic games, economics is reviving this behaviorist model. Despite the undeniable methodological contributions of the operationalization offered by behaviorism, its incompleteness in understanding adaptive behavior has been well-documented since the 1930s (Verschure 2016a). Denying the richness and complexity of a phenomenon as a methodological convenience comes at a high cost, which must be corrected in the study of collaboration. As a result, the science of collaboration is as contested as its definition, trapped in the schism of J. P. Snow's two cultures, uncoupled from all higher-order aspects of human action and experience and Aristotle's final cause (Snow 1959). Understanding and shaping collaboration is an invitation to address this fundamental challenge in building a transdisciplinary science of the human condition.

The reorientation from pragmatic goals to an imagined future has several implications that deserve further attention. When an imagined future anchors our understanding of collaboration, the scope of potential collaborations is only limited by the possible futures a collective can imagine. This implies that collaboration exists both in the mental and physical worlds, first person and third person perspectives merging along the collaboration information continuum. What unique properties of collaborating agents would allow them to balance these two perspectives?

From the imagined future perspective, and following Vygotsky's constructivist view of cognitive development and education, collaboration is a dynamic scaffolded process in which a proximal zone of attainment is predicated on existing collective cognitive structures and capabilities (Vygotsky 1978). This raises the question of which properties in a collaborating collective will facilitate or hinder the exploration of the proximal zone of potential collaborative opportunities. Exploration is, in essence, adversarial because, by necessity, it is embedded in a dialectic relation to an obstinate reality that it seeks to change. Moreover, usually, the adversary is another collaborative collective. This raises important questions about the principles governing the internal dynamics of the collaboration versus those governing the external competition and their interlinking. By defining collaboration as explorative and adversarial, it is intrinsically dynamic, responsive, and transient. If collaboration succeeds, it will transform from a deliberate, open, future-oriented process into an exploitative process based on habits and rules, thus shrinking the role of deliberative exploration and effectively moving from the collaboration end of the information continuum to the cooperation one. The resulting (more conservative complexity-shunning bureaucratic) process intrinsically conflicts with the disruptive nature of collaboration, which can potentially hamper further collaboration within the same domain. This raises the important question of how the fruits of

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

collaboration can be consolidated without limiting or even extinguishing the creative power of collaboration through a smothering bureaucracy.

Collaboration begins to take form when individual agents generate actions and communicate and interpret conventions and norms mediated and modulated by their embodied minds. This implies that the atomic unit of collaboration is the human agent. As a result, collaboration is fragile because it rests on a web of motivations, beliefs, and emotions of the collaborating agents, freeloaders, parasites, and defectors and is maintained or collapses through their actions, inactions, and interactions. The collective transcends the capabilities of the individuals forming it through complementarities, synergies, and various feedback loops. Hence, to understand the limitations and potentiality of collaboration, one must understand its atomic element: humans. Building on the outcomes of the Forum, my goal in this chapter is to highlight how we can position the study of humans in the context of collaboration to open up new avenues of research that are mission-critical if we are to address global collaboration in ways that will benefit all of humanity. In this effort, I will consider human nature or the evolutionary and biological roots, the capabilities for entertaining scenarios on the future, consciousness, and free will.

A theory of collaboration must make assumptions about human nature as the product of evolution. Where do we place humans on the Hobbes versus Rousseau continuum: Does human nature define a life that is "solitary, poor, nasty, brutish, and short," requiring strong governance, or are we "noble savages" who are inherently good and peaceful yet corrupted by society's development and institutions? Insights from this volume see these perspectives reflected in a "cage fight" mentality of all against all, expressed by the former president of the CERN board (S. de Jong podcast[1]), the destructive collaboration shaped by the Nazi's in engineering the Holocaust (Chapter 4), or cancer as a collaboration model (R. Axelrod podcast). Alternatively, others put love, "good heart," and flourishing at the center of collaboration (D. Narayan and J. Manzolli podcasts; see also Chapter 13). The brute versus noble savage contrast, however, is circumspect as a constructed master narrative that served the political development of Western society rather than an established ground truth on the nature of *Homo sapiens* (Graeber and Wengrow 2021). Hobbes's beast and Rousseau's angel are archetypical extremes on a continuum that defines human nature and its diversity in human populations. We must understand this diversity and how it impacts collaboration rather than cartoonish extremes.

The ability to collaborate sets humans apart from other great apes (Chapter 2) and may be seen as one of the key transitions in evolution that enabled the capabilities of joint/collective intentionality, theory of mind (ToM), norm-driven adaptation, moral agency, and social selection against cheaters, from which a notion of "group-mindedness" emerged (Tomasello 2019). The ability

[1] Podcasts are available at https://esforum.de/forums/ESF32_Collaboration.html?opm=1_3.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

to understand and align with the mental states of others is crucial for effective collaboration, and this requires advanced forms of mentalizing or virtualization (Verschure 2016b). This mental innovation allowed hominids to form socially shared agencies, creating norm and convention-driven feedback systems facilitating complex social coordination. This hypothesis requires further unpacking and raises questions on the evolutionary transitions that boosted coordination and cooperation into collaboration, the mental innovations this entailed, and the underlying somatic and neuronal modifications. As an example, let us consider ToM. Asserting that ToM is a critical evolutionary innovation assumes capabilities to apply it in a collaborative context. In other words, collaborating agents must answer the questions: Who and where are the other agents, and what is their ToM? One perspective holds that since evolution has pre-equipped vertebrates with robust and fast reactive control systems, it has already injected intention into the world as a prior (Chapters 5 and 17; Verschure 2016b). Accordingly, agents can display social interaction and collaboration by virtue of ontological commitments based on strong priors of "the other," including their intentionality and potential for competition and collaboration. Hence, ToM is exercised instinctively, triggered by those states of the world that require interpretation and explanation based on an intentionality prior (Lallee and Verschure 2015). When combined with the intentionality prior, ToM can be applied to any complex phenomenon, providing an "intentional stance" as a shortcut to comprehending environmental complexity in intentional terms (Dennett 1989). However, discussions reflected in this volume show that this cognitive tool goes beyond such pragmatics. Deploying ToM to an environment through the intentionality prior creates a proactive social attention mechanism. This will boost the potential for collaboration and scale it up from direct proximal interactions in the physical world of physical objects, a commons and mother earth to other agents and abstract entities such as unions, companies, religious groups, or nation-states. This generalization is facilitated because any entity can now be conceived as an intentional agent with its goals with whom joint intentionality could be formed, irrespective of the factual presence of agency. The *telos* of agents, whether a physical person or not, are part of the virtualizations and projections of the beholder, serving as alignment promoting information compression rather than being explicitly realized in the multiscale dynamics of collaboration as it unfolds in the physical world. Mentalizing allows humans to collaborate beyond the physicality of the commons and creates new affordances. Indeed, other agents can be instrumentalized and seen as virtual tools serving individual or collective goals, which can be expressed in the allocation of roles and responsibilities in collaborative collectives. Moreover, the intentionality prior can be generalized to time itself, creating an illusion of purpose in the ongoing dynamics of reality and instilling a necessary optimism in the progress and directedness of collaborative efforts.

Explaining collaboration from the evolutionary innovation of "group-mindedness" requires understanding the constituent processes to avoid circularity.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

This invites us to look at consciousness, volition, mental time travel, mentalizing, intrinsic motivation, norm acquisition, and social learning and cognition, to name a few, from the perspective of collaboration. Each of these processes is not well understood in its own right, let alone integrated into a comprehensive theory of the embodied mind. In addition, consciousness and volition are considered controversial because they do not seem to fit the standard physics-grounded model of reductionist Western science (Sapolsky 2023; Wegner 2003; see also Chapters 5, 13, and 17). To start, we can consider that the underlying processes can be grouped together under the capability of virtualization to express the decoupling from the real world and the future-orientedness of the associated internal processes and representations. Virtualization comprises an evolutionary transition that I speculate occurred shortly after the Cambrian explosion 500 M years ago, creating the conditions for social interaction, cooperation, and collaboration to emerge (see Chapters 5 and 17; Verschure 2016b). As an example of a more specific research agenda, we can consider one well-researched aspect of virtualization: the ability to entertain scenarios of possible futures or mental time travel (Suddendorf and Corballis 1997). In rodents, this has been observed in so-called forward sweeps in the memory system of the hippocampus at decision points in a maze (Johnson and Redish 2007). In these sweeps, populations of neurons that encode specific locations, or place cells, respond in the order in which they would be activated if the animal traverses the associated location. However, the animal is at rest at the decision point when these sweeps occur and effectively explores future trajectories. Collections of place cells form allocentric maps (i.e., world-centric) that support the spatial cognition underlying goal-oriented navigation (Buzsáki and Moser 2013) and instantiate a cognitive map proposed by Edward Tolman in the 1930s. This concept is supported by an overwhelming amount of empirical evidence in the neuroscience and psychology of episodic memory, with broad ramifications in language, affect, and cognition (Moscovitch et al. 2016). Mental time travel has been directly observed in these maps, and more recently, the representation of social maps within the same memory system has been reported (Tavares et al. 2015). This would suggest that prosociality and collaboration are realized upon the substrate of spatial cognition. If this hypothesis is correct, human collaborators enter any social situation with an abstract map informing their social navigation of the collaboration modulated by their objectives and traits, imagining possible futures, expectations, and interactions, and creating conditions for their individual roles to emerge. This implies that the intention of a collaborative process is an amalgamation of all the individual goals the participating agents bring to bear in their actions and expectations. In this view, a collaboration would have an emergent direction, and only the participating agents would have goals. The innovations of the social and collaborative brain underlying these adaptations have barely been explored and provide a new perspective on core brain mechanisms of memory and cognition. Recently, it was discovered that volition has a decisive global modulatory role on these

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

memory dynamics in the human brain (Pacheco et al. 2021). In the context of collaboration and the social cognitive map hypothesis, this suggests a mechanistic coupling between individual agency and prosociality.

Collaboration finds coherence in the narratives the participants share, whether ideological, mythological, religious, or natural (E. Slingerland and M. Levi podcasts; see also Chapters 5, 13, and 17). Mental time travel allows the mind to order events and links them to meaning, motivations, emotions, and goals, creating ontological frames for these stories shaping experience and the potential for collaboration. Understanding mental time travel is necessary for scrutinizing how these narratives are constructed and maintained. Narratives, in turn, allow scaling the complexity of collaborations by providing cognitive offloading through symbol systems and the construction of common ontologies, limiting and directing the space of possible futures and forming the proximal zone of viable collaboration trajectories. This raises the question of how the cognitive mechanisms and ontological priors of event structures and narratives contribute to effective collaboration and what their origins are in mixing nature and nurture.

The proximal zone of collaboration is not static; it is reshaped by the dynamics present in the collaboration itself. On the one hand, there is a feedforward influence through the initial conditions of the collaboration and the dynamics of agents and their physical and social environments. On the other, an emergent effective environment created by the collaboration itself biases the perceptual, emotional, and cognitive structures of the agents involved, including virtualizations, imagined futures, and narratives. Hence, collaborations unfold in a continuous feedback loop between the task space, agents, their virtualizations, and the resistance created by opposing forces, which can be seen as a form of niche construction (Chapters 5 and 12). Deciphering these feedback loops will allow us to identify the architecture of a collaboration (Chapter 14). Understanding this architecture and its different topologies (e.g., implicit or explicit, flat or hierarchical, voluntary or enforced) will allow us to better understand the success, failure, and dynamics of collaborative processes in various contexts. As with all complex control systems, the question is: What constraints must a collaboration satisfy to be deconstrained in its task space? For instance, leadership, agent features, joint intention, "group-mindedness," norms, communication, and incentives might be considered necessary for successful collaboration, yet what are the potential unforeseen consequences when these constraints are not met or absent? For example, think of the devastating effects of failing and malignant leadership (Chapter 19), the abuse of trust (Chapter 11), or the distortions introduced by disinformation and misleading narratives, all which contribute to failing or pathological collaborations. Especially in the design of future collaborative artificial and hybrid systems, understanding these architectural constraints will be of the essence (Chapter 5).

Collaboration is guided by conventions and norms, implying that collaborating agents have a sense of self, desires, beliefs, and the imagination of

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

future scenarios. Based on various observations gathered during the Forum, collaborating agents can evaluate and predict concrete and abstract outcomes, reason, act autonomously, and learn. None of these features would appear controversial, yet they form the constituent features of an agent-centric pragmatic model of free will (Murphy et al. 2009; Strawson 2010; Wolf 2013). In this agent-centric model of responsible autonomy, volition emerges from complex information-based interactions that cannot be reduced to physical causation. Instead, as collaboration, it can be seen to result from the recurrently coupled levels of organization of biological systems and their multiscale "amplification logic," realized in the processes underlying biological evolution, including genome, organism, and their physical, social, and cultural environments (Deacon 2011). The freedom that comes with the responsible autonomy of collaborating agents is bounded and can be captured as the freedom to relocate, disobey, and form new social bonds (Graeber and Wengrow 2021). This implication of our analysis of collaboration defines a controversial research agenda because it assumes that free will is compatible with our understanding of the deterministic natural world, which is rejected by the eliminative hard determinisms of the so-called incompatibalists. Experts expect no breakthroughs in this age-old confrontation (Strawson 2010). Yet, our analysis shows that collaboration is intrinsically coupled to our concept of free will. Hence, understanding and shaping collaboration implies that we need to take a position on the status of free will both as a natural phenomenon and a right. Reducing it away on methological grounds will not suffice.

Psychological traits vary in populations of humans, creating a tension between the role diversity and similarity play in collaboration. Diversity can be a potential driver of innovation and adaptability, while similarity could enhance efficiency and communication. Yet, the question is whether there is a proper balance between these two and whether certain combinations of traits we find in individual agents make them good or bad collaborators. For instance, in Yip et al.'s view, attachment theory shows a link between personality characteristics and collaboration effectivity (Yip et al. 2018). An innate attachment behavioral system motivates people to seek support from others in times of need following an anxious, avoidant, or secure style (Bowlby 1969). Early childhood experiences modulate attachment styles and directly impact how people operate in teams and develop and maintain trust. For instance, secure attachment is linked to autonomy and creativity, while team members with an avoidant attachment style tend to resist leadership. In addition, it has been shown that changes in the employment relationship can initiate attachment-seeking behaviors among employees toward their organization (Albert et al. 2015). Another example can be found in our understanding of stress, which directly impacts the balance of deliberate and reactive/habitual control (Sapolsky 2017). Stress triggers the release of glucocorticoids, which significantly alter the brain's control architecture, shifting it from deliberate, frontal lobe-dependent processes to more emotional, habitual, and instinctive responses. This

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

modulation will directly impact the cognitive systems underlying ToM, mental time travel, and rational choice and can have runaway detrimental effects on collectives (Vodovotz et al. 2024). The control systems of individual agents are not static. Rather, they continuously reconfigure to satisfy multiple and ever-changing constraint boundaries from fast reactive systems in emergencies to slow deliberation facing complex challenges (Kahneman 2012; Verschure 2016b). Unraveling these effects and understanding their impact on collaboration requires that all core processes of the mind (i.e., motivation, personality, emotion, perception, cognition, language, consciousness, agency, and volition) are investigated comprehensively and integrated into one standard model. The last attempt to achieve such a model stranded in the work of Clark Hull 70 years ago (Hull 1952). Moving toward a standard model of the human embodied mind in all its complexity and variability is a formidable foundational transdisciplinary challenge requiring large-scale scientific collaboration. Building such a model will directly inform and calibrate our attempts to realize synthetic collaborators (Chapter 5).

The discussions at the Forum have shown that collectives and organizations frequently make the assumption that humans can collaborate "out of the box." This raises the question whether it is reasonable to expect that a random individual human will automatically be able to act effectively in a collaborative context, or whether this requires preparation and training in our rapidly technocentric, diverse, and global yet fragmented world? Panksepp (2004) proposed that evolution introduced collective play (or collaboration, for the sake of collaboration) as one of the seven core emotional systems of the vertebrate brain. This would imply that evolution prepared our brains for collaboration in simple contexts and provided the necessary mechanisms for bootstrapping our prosociality toward complex technocratic societies through cultural and technological development. Another example of innate mechanisms that prepare the mind for interaction, communication, and collaboration is the so-called interaction engine proposed by Levinson (2006), which putatively allows humans to explore the social and cultural world. The question is whether these evolutionary predispositions are enough. If we want humans to effectively collaborate in complex environments that span the globe and beyond and involve thousands of active physical and social elements, we cannot only rely on biological priors rendered by evolution; we have to develop a pedagogy of collaboration that matches the complexity of the collaborative challenges humanity is facing (Chapter 13). Today, such programs do not exist at scale, although the military has dedicated the most efforts to this critical aspect (L. Sciulli and R. Popovic podcasts).

Collaborating collectives define an identity for their members, forming communities of fate (M. Levi podcast). In the adversarial nature of the collaboration, identity can be sharpened through schismogenesis, i.e., the definition of identity through contrasts proposed by Bateson (1958). Schismogenesis fundamentally influences collaboration dynamics because the distinction between

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

competing collectives needs to be maintained by reaffirming identity, which can potentially become incrementally polarized in an identity arms race. This can lead to entrainment of the collaboration by the drive for identity instead of targeting its initial objectives or a collaboration dynamics that serves identity creation. In addition, schismogenesis can lead to challenges to the coherence of the collaborating collective as multiple references for identity building emerge, for instance, linked to various roles and relations. A close link between collaboration and identity is demonstrated by recent experiments that show that humans limit their shared attention and empathy to in-group members (Hein et al. 2010; Vanman 2016). This defines a potential limit to the collaboration potential of a collective to those with whom one shares identity. Hence, identity is another constraint that deconstrains and is a key variable to understand and manage in collaboration. Schismogenesis and the psychology and prosociality of identity create a collaboration paradox because, whereas a shared identity is fundamental to building and maintaining collaboration, it also creates conditions for its collapse by driving escalating contrast with competing collectives and internal fragmentation due to an individual drive for identity. The complex relationship between identity and collaboration must be respected and further investigated, and it again illustrates that collaboration carries in itself the mechanisms for its demise.

We marvel at the unimaginable challenges humans can overcome through collaboration. Yet, as numerous examples reveal—from warfare and systematic, state-sponsored genocide during World War II to industrial strategies that optimize profit at great cost to humanity—there is no automatic link between collaboration, human flourishing, and morality. Since collaboration is adversarial, this intrinsic dialectic easily translates into moral dichotomies of good and evil, collaboration versus collusion. Within the collaborating group, there exists a notion of a common vision and shared ontology with its associated virtues, yet there will, in most cases, be another collective that aims to negate that vision and disrupt its collaborative process, setting its own standards of reality and morality. Given the malleability of the human mind, this should not surprise us. We can think of the battle between the fossil fuel industry and grassroots ecologists to address climate change or between global powers to obtain dominance in world affairs. In this context, we can revisit the paradox created by the connotation of collaboration that emerged after the First and Second World Wars, where it became synonymous with treason to the nation-state (R. Van der Laarse podcast; Digeser 2022): treason because an individual or group willingly switched sides in the adversarial interaction between two collectives and their associated national identities. This implies that the collaborator exchanged the norms and values of the initial in-group with that of the opposing group, thus personifying schismogenesis. It must be emphasized that the term collaborator took on this connotation especially in countries where only a small fraction of the population was involved in active resistance against the Nazis while the majority took on a bystander role. One can speculate that outrage and

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

prejudice expressed by this majority resulted from the cognitive dissonance of violating the basic tenet of collaboration, attachment, and identity, now spanning the divide of adversarial relations compounded by collective guilt. If we want to call this collaboration, we must acknowledge it as a special case and see it as a warning when attempts are made to merge collaborating collectives or individuals are asked to switch sides. This example illustrates that studies of collaboration should consider that this complex phenomenon occurs in various qualities, from aspirational and serving the common good to destructive and evil as well as all possible variations. Going forward, our approach must become less monolithic and more inclusive so that it holds whether one speaks of the construction of pre-Columbian earth mounds by North American cultures, warfare, genocide, the mafia, a company, public institutions, research consortia, or democracy. Lumping all of these together under one concept and a single model will prevent us from understanding its nature and dynamics and limit our ability to master collaboration in the service of improving the human condition.

Humans are entering a new age of enhanced collaboration with technology in the form of machines and algorithms (Chapter 5). Artificial intelligence (AI) has already irreversibly entered society on a large scale. We will witness accelerating hybrid human-machine collaboration, creating opportunities in communication, decision making, and productivity versus risks such as bias, privacy, dependence, job displacement, and transparency. If we look at the mid- to long-term impacts, new opportunities, and risks appear. AI tools can monitor the quality of collaborative processes and their participants and provide feedback on improving and sustaining them. It can do so by monitoring all the critical parameters of the human collaborative process, some outlined above, and providing properly timed and packaged nudges tuned to the roles of the participants. It can also provide high-level representations of collaboration dynamics and predict its future trajectories, decision points, and consequences, creating new levels of monitoring and understanding collaboration. Yet, as the Forum's analysis shows, our humanity is intricately linked to our ability to collaborate, and it is not the case that these technologies are directly engineered or potentially self-evolved to constructively match the critical parameters of human collaboration. This development will require a focused and regulated path. Given the exclusively commercial interests driving the current AI revolution, one can expect a bias toward short-term financial success that will exploit the foundations on which human collaboration rests, masked by the misplaced optimism and hype of "technoreligion." A negative scenario foresees contemporary surveillance capitalism (Zuboff 2018), where humans are exploited for their data, transforming into a new version of exploitative capitalism where humans perform labor in the illusion of collaborating while being controlled by AI and its human masters. For instance, as discussed, the ability of humans to collaborate is built on constraints that deconstrain our prosociality. We project intention into the world, are prosocial by design, and seek affirmation and

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

identity while being norm-sensitive and reason-responsive. Current AI algorithms are not and probably will never match these capabilities in a similar way. Through this mismatch, they can break and co-opt human collaboration by violating or hacking these constraints. Indeed, it is likely that new ways of manipulating human collaboration will be found and exploited beyond the destructive narratives of disinformation. A simple hack of prosociality is to provide human users with demographics-compatible positive social feedback, an effective trick of virtual companion sites. Currently, this targets single users, but generalization to groups will follow sooner rather than later, creating adverse collaborative dynamics, especially when Large Language Models can generate increasingly more psychologically plausible narratives. Alternatively, one can drag users down so-called radicalization pathways to get them to respond to clickbait, reinforcing micro-identities and fractioning collective identities, as is on full display during significant world events. Once packaged in AI-generated multi-modal anthropomorphic form, these cues will become more powerful with convincing social salience (Inderbitzin et al. 2013). Lastly, we will see the emergence of pure synthetic forms of collaboration where algorithms and machines will develop their own collaboration architectures operating at complexity and speed that humans cannot comprehend or follow. The future will see artificial collaboration pathways created and natural ones blocked in the service of human-controlled commercial and political forces and autonomous collaborating artificial systems. The most significant risks humanity faces are, as usual, the unintended consequences of these developments based on a naïve techno optimism. Mitigating this risk calls for action to build regulatory frameworks based on a true understanding of collaboration.

A second form of human-machine collaboration at the horizon will be human augmentation. In this case, humans are not necessarily in competition with technology but rather interface with it to augment their existing capabilities, including "group-mindedness," empathy, and other collaboration skills. This could imply direct neurotechnological interfaces to the processes underlying our prosociality and its constituent processes. Enhancing sensitivity to social cues, the capacity of the map of social space, and broadening communication channels could strengthen collaboration potential. New forms of perception and cognition serving collaboration could be constructed through collective sensing technologies, such as artificially synchronizing brains to promote collaboration (Chapter 10). This step, although still in the long-term future, opens new domains of manipulation and exploitation, especially because they will be directly and subliminally linked to the user's experience, making detection practically impossible.

The field of collaborative AI is investigating how fully synthetic collaboration can be structured (Chapter 5). The most promising way to achieve this is to emulate the architecture of the human body–brain–mind nexus in a drive toward neuroAI. This step will advance humanity's technology agenda, potentially

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

realizing entirely artificial societies dedicated to mission-critical tasks and, in doing so, validating the principles underlying human collaboration.

We started this project four years ago in the hope of advancing our understanding of human collaboration and, in that way, contributing to improving it in the face of the global challenges that humanity must overcome to properly manage spaceship Earth (Fuller 1969). We did not achieve a comprehensive checklist so that each collaboration can be efficiently structered and realized. Rather, we found a universe of challenges below the surface of our initial definition. Our initial direct questions of what collaboration is, how we collaborate, what the underlying processes are, and when it breaks down made us discover that it is the most complex phenomenon in the universe. Answering these questions led us to the heart of what it is to be human. Hence, to address the pressing global challenges that can spell the end of the Anthropocene as we know it requires us to find out who we are as a collaborating species.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

From “The Nature and Dynamics of Collaboration,”
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144