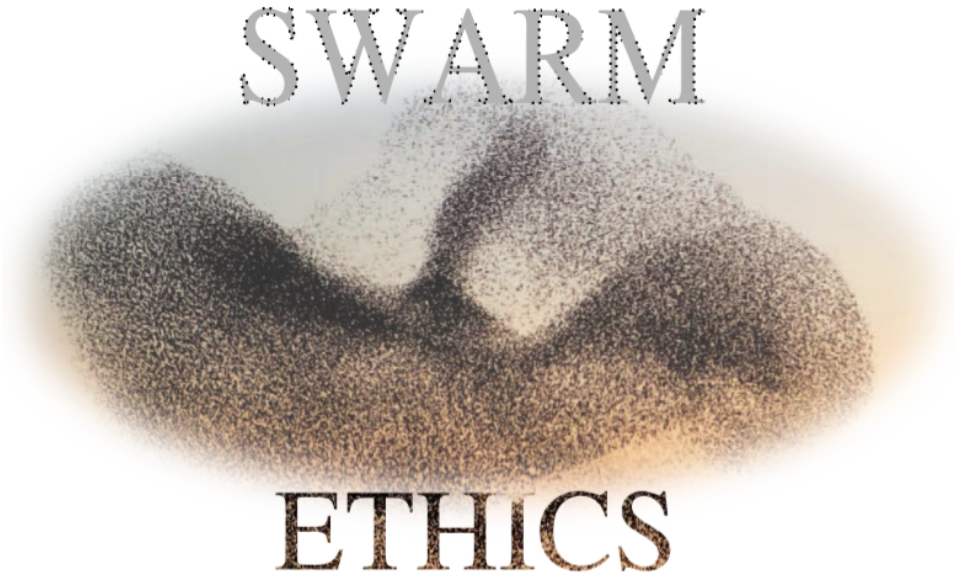
Swarm Ethics: A New Collective and Decentralized Ethics in the Digital Age

Katja Rausch[1] and Daniele Proverbio[2]

[1]Founder, House of Ethics.
[2]University of Luxembourg, Luxembourg Centre for Systems Biomedicine (LU)

Submission to the conference "CEPE 2023: Human-AI Interaction and the Future: Values, Responsibilities, and Freedoms" and the Vivian Weil Best Paper Award, Illinois Institute of Technology, Chicago.



In face of nowadays vibrant, fluid and ever-changing societies, old ethical paradigms are challenged. No more a singleton, the individual is embedded in dynamical communities, in turn subjected to open environments and stimuli. We thus introduce a new concept fit for the digital age and the "individual-within-open-societies" : Swarm Ethics.

Swarm Ethics as a new multi-disciplinary phenomenon runs at the cross-roads of behaviorism, complex systems and modern digital technologies. Thirty years ago, Gerardo Beni defined Swarm Intelligence as a "'swarm' of agents (biological or artificial) which, without central control, collectively (and only collectively) carry out (unknowingly, and in a somewhat-random way) tasks normally requiring some form of 'intelligence'" (1989). Building on it, Swarm Ethics equally identifies multitudes of intelligent agents that manage to perform coordinated, self-organized ethical activities by sharing information and acting upon dynamical environments. Swarm Ethics is about "emerging group ethics in open systems", fueled and scaled by digital technologies that speed up all processes.

The novel approach transcends traditional Western ethics and its cognition-based evolution of morality. In normative ethics, be it deontological, consequential or virtue ethics, ethical decision-making aims at a judgmental outcome, often leading to classical ethical dilemmas. With Swarm Ethics, we finally welcome an alternative to the classical two-track ethical dilemma, and open up for the multi-layered "ethical imperative" by Heinz von Foerster.

Our proposed concept circumvents the predominantly used and over-used approach to morality and ethics as being an individual evolutionary cognitive process (Laurence Kohlberg, "The Development of Modes of Thinking and Choices in Years 10 to 16" (1958)). Swarm Ethics, on the contrary, is "emerging" group ethics at large scale. The self-emergence process of ethical shaping within a group is driven by purpose to build trust and navigate within societies.

As paramount difference with traditional ethics and its pre-shaped mental moral constructs and norms, self-organization in Swarm Ethics is not encoded. It is rather an emergent property of the system as a whole, following the self-organizing principle of Swarm Intelligence. Ethical virtues and values are not "hard-coded" in anybody's cognition, but they are shaped and finally self-emerge by means of information sharing and coordination with other agents, eventually being recognised and labeled by beholders. The initially random shaping of (ethical) patterns as a characteristic of intelligence is fundamental for purpose-driven Swarm Ethics to be actionable and scalable.

A second significant driver lies in its horizontal, decentralized model. It is not based on a top-down (control-driven) hierarchy, but builds its own intelligence by following simple rules and quickly reacting to external stimuli as well as internal processes. In Swarm Ethics, agents are both autonomous and collective entities. The evolution of the swarm does not require spectacularly complex and nuanced rules to achieve collective consensus over shared principles. Instead, simple and horizontal behavioral patterns are often sufficient to elicit complex emergent phenomena, if individuals share similar "cost functions" (that is, similar purposes). A paradigmatic example is Schelling's model of segregation (T. Schelling, "Dynamic models of segregation" (1971)). Put together many individuals, belonging to two groups. Each individual follows a simple rule: "If there are more than x% of neighbors that belong to the other group, move". Even when the percentage is very small, the resulting effect is segregation between the two groups. Albeit simplistic, it shows that societal division may emerge from cascading simple (and apparently inoffensive) behaviors.

The third major difference to traditional ethics consists in the fact that Swarm Ethics is manyfold enhanced by digital technology. It is an agile, fast-paced form of emerging group ethics. The social ethical swarm is grouped, amplified and scaled by digital social media, which accelerate information speed, enlarge their diffusion potentials and provide almost-instantaneous feedback to all agents.

We have identified six main characteristics for Swarm Ethics : 1. no top-down model but a horizontal grouping of autonomous individuals; 2. based on perception-action and imitation behavior; 3. based on simple rules/principles : imitation, care and purpose-driven; 4. depen-

dent on social media to scale, amplify and synchronize; 5. digital communication fuels rapid information sharing and massive swarm building; 6. high potential in social communities, business and political environments.

These characteristics shape Swarm Ethics into a new robust ethical phenomenon that thrives on three major catalysts.

As first catalyst, Swarm Ethics is perception-action based as opposed to the traditional cognition-based approach to ethics. Past ethical ideas were born when nature was thought to be linear, mechanical and individual-based. Modern information theories (Claude Shannon, "The Mathematical Theory of Communication" (1948); Norbert Wiener, "Cybernetics: Or Control and Communication in the Animal and the Machine" (1948)) have shown that single entities interact, and feedback energy and information in ever-changing networks. Swarm Ethics fully acknowledges humans as complex systems within complex systems: dynamic and responsive agents in open, mutant systems. These open systems have no leader nor hierarchy (as opposed to the traditional theological Kantian system in deontological ethics, for example) but a synchronic, empathetic movement created out of disorder.

The second catalysts are mutuality (imitation) and empathy (care). In Swarm Ethics the Other Next is important as the agent is no singleton but embedded in a group. Turning to behavioral anthropology is of prime importance to capture the pace and agility that characterizes Swarm Ethics. Our proposed concept builds also on the works in philosophical anthropology by the French social scientist René Girard and his Mimetic Theory ("Anorexie et désir mimétique" (2008)) as well as on modern findings in adaptive behavior by Dutch ethologist Frans de Waal ("The Age of Empathy" (2009)).

For the first two catalysts to happen, Swarm Ethics must go beyond cybernetics and early system theories. Instead, it addresses the open systems with exchanges of information, external stimuli, and dynamical interactions. The "system" becomes an "open system" (Ludwig von Bertalanffy, "The Theory of Open Systems in Physics and Biology" (1950)), subject to fluctuations. We say it is "out of equilibrium": not a monolith with adamant convictions and without connection to the external world, but a subject to information fluxes which may not mutually counter-balance.

All these catalysts, combined, light the ethical fire and start to combust, scale and thrive through interconnected digital technologies, thus allowing rapid scalability of emerging group ethics to happen transnationally and inter-culturally. In the 1960s, the media theorists Marshall McLuhan and John Culkin already insisted on the systemic link between behavior, technology and society. Fuirther influenced by modern Habermasian "media-medium" ethics (Jürgen Habermas, "The Theory of Communicative Action: Reason and the Rationalization of Society" (1984)), combined to complex systems and modern social technologies, Swarm Ethics would be part of applied ethics, spurring agile and fast decision-making in the digital age.

The novel concept will lead to thrilling research opportunities in applied ethics, digital, social and business ethics. It can be a game-changer in digital ethics with the rise of collaborative and companion robots, artificial caretakers, chatbots or any technological agent in social digital interaction or web3 metaverses.

New philosophical questions arise: How do we understand individuals per se and within societies, as open agents and mutating systems? Can we develop conceptual and actuable frameworks that can cope with dynamical changes within a community, society, company or nation? Even meta-ethical questions spur: Which actions and purposes act as catalysts for agents to integrate, modify and propose ethical principles? Is it primarily perception-action and satisfaction of "rational" cost-functions (purposes) like trust building? Or reactive imitation? Or else, empathy and other "irrational" impulses? How much does each catalyst concur in shaping the overall behaviors?

Digital polarization and political/ideological groupings can be further analyzed in the light of Swarm Ethics. As unexpected "emerging group ethics", Swarm Ethics has been observed on many instances within online communities, businesses, political events at national or international level. In such observed cases, new purpose-driven group patterns have been emerging

fast and rapidly building robustness in order to become actionable. Many societal questions can thus be addressed with different perspectives. What is the influence of mutual information exchanges, and how does their speed influence the construction of individual beliefs? How is Swarm Ethics different from "nudging" or simple "influence"? Or take a more machiavellian turn with "How to influence/enable ethics towards desired directions?". Where "desired" should again be an emergent consensus, therefore involving multiple actors and factions.

Within this novel paradigm, ethics is fully recognised as "difficult" but, at the same time, it regains its prestige: if everything influences ethics, ethics influences everything. Will Swarm Ethics be a new lever capable of bridging complex humanness and velocious technology, and consequently help us, once again, to survive adversarial, volatile and uncertain environments as a human species?