



A System-Theoretical Account to Kant's Interactional Ethics*

Peter beim Graben, Kiersten Xiangqi Zhu†

Abstract

A healthy open society depends on individuals who are capable of autonomous judgment while remaining responsive to the perspectives and claims of others. This creates a tension, as its members must remain autonomous, while sustaining the conditions that make moral autonomy possible. To address this problem, we develop a reconstruction of Kant's ethics in terms of system theory. We interpret Kant's three maxims of common human understanding as a unified perspectival structure: The first maxim requires unprejudiced rational autonomy, the second one refers to perspectival transformations, and the third maxim demands consistency across such transformations. We represent this structure using the formal language of lattice gauge theory, in which individual perspectives correspond to local degrees of freedom, perspective changes to symmetry transformations, and consistency against such transformations to invariance. On this basis, we reconstruct Kant's four formulas of the categorical imperative as self-referential conditions for the preservation of rational agency and the maintenance of the realm of ends. A healthy democratic open society is thus not defined by the absence of conflict, but rather by the preservation of autonomous judgment and by the capacity to revise one's own maxims through social interaction. Social structure can sustain moral agency, but cannot substitute individual moral responsibility.

Keyphrases

Agents, social networks, rationality, alignment, symmetry.

Contents

Introduction

Kant's Ethics

Kant's Maxims of Common Human Understanding	1
Individualistic Ethics	2
Interactional Ethics	3

Lattice Gauge Ethics

Individualistic Systems	4
Interactional Systems	6
Lattice Gauge Symmetries	8

Discussion

Citation	15
Acknowledgments	15
Affiliations	15
References	15

Introduction

Healthy democratic open societies exhibit the paradox of an inherent vulnerability (Arendt 1976; Taswell 2025b; Zhu et al. 2025). They presuppose individuals who are capable of judging for themselves, expressing dissent, and assuming responsibility for their actions (Etzioni 2000; Minkler 1989; Nyswander 1967; Popper 2013). Yet these capacities cannot simply be secured by external authority, because judgment ceases to be independent when it is merely prescribed.

This is what makes openness both possible and fragile. As Popper (2013, p. 165) argues, an open society differs from a closed, tribal, or collectivist order in that individuals are confronted with personal decisions. Openness therefore requires more than individual rights that are just guaranteed by social and political institutions. It also requires people who can think for themselves, accept criticism, revise their judgments, and recognize that others have the same capacity for judgment (Popper 2013, pp. 174, 179, 188, 482).

However, under conditions of uncertainty, complexity and rapid social change, the difficulty is not only that judgment becomes harder, but that individuals may become less willing to assume responsibility for their own choices. Responsibility may be cast outward, onto those easiest to identify and blame (Havel 1990, p. 35), so that complex social problems are recast as threats posed by migrants, outsiders, or other convenient scapegoats (Arendt 1976, p. 295). Or responsibility may be turned inward, as the social and political causes of distress are overlooked and individuals are encouraged instead to understand and manage these problems as matters of personal resilience and self-management. Both reactions can amount to a withdrawal from moral judgment: Responsibility is no longer critically examined, and the obligations we owe to one another cease to become objects of reflection.

It is this inward displacement of responsibility that becomes especially apparent in contemporary therapeutic culture, where depression, anxiety, and other forms of distress are increasingly framed as problems to be treated and managed at the level of the individual. The contemporary diffusion of *mindfulness* offers a representative example. Purser's critique of "McMindfulness" (Purser 2019) concerns the way mindfulness is repackaged as a tool of emotional regulation, resilience,

* Document created 2025-10-04, updated 2026-09-29.

† Correspondence to peter.beimgraben@b-tu.de.

and self-optimization, so that stress may be treated as a problem of individual adaptation rather than as a response to wider social conditions.

This tendency can be reinforced in neuroscientific terms. Claims about ‘changing the brain’ or ‘enhancing well-being’ may lend an appearance of *objectivity* to this individualized model of adaptation. In this sense, it approaches what Han (2017) calls “psychopolitics”: Subjects remain formally free while increasingly being encouraged to manage and optimize themselves. This narrowing of freedom has a further consequence: it weakens the relation between the individual and the common world in which judgment acquires political meaning. Arendt’s account of the “right to have rights” (Arendt 1976, p. 296) helps to illuminate what is at stake: Agency depends not only on formal rights, but on belonging to a world in which one individual’s words and actions can appear before others and can be judged by them (Arendt 1977, p. 146). Even when rights remain intact, this political capacity can weaken if individuals cease to judge for themselves or to consider how their judgments appear from different standpoints.

This does not require subordinating the individual to the community. Instead, Mariña’s reading of Kant’s “ethical community” (Kant 1793/1996, 6:94) suggests that autonomy is exercised within relations that shape — but do not determine — our possibilities of rational action (Mariña 2022). An ethical community can therefore neither be imposed through external coercion nor reduced to an aggregate of isolated individuals. Its members must remain autonomous while sustaining the conditions that make moral autonomy possible.

The problem of maintaining healthy democratic open societies can therefore be recast as a problem of interactional ethics: How can autonomous agents together sustain the conditions of possibility of moral autonomy, without undermining one another’s individual freedom? A healthy open society, on this view, is not one without conflicts, tensions, or uncertainties, but one in which judgment, criticism, alignment, mutual revision, and responsible action are guaranteed by self-preserving rational agency.

In our reconstruction, Kant’s three “maxims of the common human understanding” in the *Critique of the Power of Judgement* (Kant 1790/2000a, 5:294f) provide the crucial starting point for addressing this problem: “To think for oneself”. “To think in the position of everyone else”; and: “Always to think in accord with oneself”. We interpret these three maxims as a unified perspectival structure. The first requires unprejudiced rational autonomy the second refers to perspectival transformations, and the third demands consistency across such transformations. Reason is thus neither an internal property of an isolated subject nor a collective identity that dissolves individual differences, but an interactive order arising among autonomous perspectives.

This structure suggests a natural system-theoretical formulation in terms of local gauge freedom, symmetry transformations among perspectives, and eventually invariance across those transformations. Our system-theoretical reconstruction of Kant’s interactional ethics essentially draws upon the formal structure of lattice gauge theory — yet, not in order to reduce ethics to physics or mathematics, nor to claim that morality itself constitutes a kind of physical gauge field. Rather, we use the formal language of local gauge freedom, relational transformations, and invariants to represent individual perspectives as local degrees of freedom, perspective changes between agents as symmetry transformations in a social connection, and the structures preserved under such transformations as the common normative order that is preserved across different local standpoints (O’Neill 1989).

The question pursued in our study is therefore not how a mathematical model might solve the contemporary democratic crisis and the current threads upon human rights (Arendt 1976, p. 279), but rather a much more fundamental one: What kind of rational structure enables a society composed of free subjects to remain open while sustaining the conditions of its own openness? A healthy democratic open society can, on this view, be understood as a self-maintaining interactional system (Taswell 2025b, p. 1). Its openness is not guaranteed once and for all by external authority, but must be continually realized through the judgments, actions, and relations of its members as the condition of its own possibility, i.e. as Kant’s “realm of ends” (Kant 1785/2002, 4:438).

Kant’s Ethics

One of the most famous and mostly cited paragraphs on ethics is the beginning of Kant’s 1785 *Groundwork for the Metaphysics of Morals*, (Kant 1785/2002, 4:393):¹

There is nothing it is possible to think of anywhere in the world, or indeed anything at all outside it, that can be held to be good without limitation, excepting only a **good will**. Understanding, wit, the power of judgment, and like *talents* of the mind, whatever they might be called, or courage, resoluteness, persistence in an intention, as qualities of *temperament*, are without doubt in some respects good and to be wished for; but they can also become extremely evil and harmful, if the will that is to make use of these gifts of nature, and whose peculiar constitution is therefore called *character*, is not good. It is the same with *gifts of fortune*. Power, wealth, honor, even health and that entire well-being and contentment with one’s condition, under the name of *happiness*, make for courage and thereby often also for arrogance, where there is not a good will to correct their influence on the mind, and thereby on the entire principle of action, and make them universally purposive; not to mention that a rational impartial spectator can never take satisfaction even in the sight of the uninterrupted welfare of a being, if it is adorned with no trait of a pure and good will; and so the good will appears to constitute the indispensable condition even of the worthiness to be happy. (Emphases in the original)

Already this short fragment of one of the most influential philosophers of cultural history² poses several severe challenges of interpretation: What is “will”? What does it mean to have a “good” will? What are “understanding”, “wit”, or the “power of judgment”? What is meant by “character”, “happiness”, or “rationality”?

These very questions essentially refer to Kant’s metaphysical attitude of “transcendental philosophy” (Kant 1999, 21:17). Kant was acutely used to accurately distinguish between the concepts “transcendental”, on one hand, and “transcendent”, on the other hand, where the former refers to “something that indeed precedes experience (*a priori*), but that, all the same, is destined to nothing more than solely to make cognition from experience possible”, whereas the latter refers to the

¹We follow here the standard citation scheme of Kant’s works referring to “volume number:page number” of the German *Akademie-Ausgabe*, published by the Berlin Brandenburg Academy of the Sciences (online as *Bonner Kant-Korpus*: <https://korpora.org/kant/>)

²One of the world largest exhibitions on Kant’s life and work has recently been opened at the Kant-Museum in Lüneburg (Lower-Saxony, Germany): <https://ostpreussisches-landesmuseum.de/en/kantmuseum/>

supersensible that “cross beyond experience” (Kant 1783/2004, 4:373, footnote). Hence, the “transcendental” always refers to the “conditions of the possibility”, either “of the objects of experience” in epistemology, (Kant 1787/1998, B197),³ or to freedom and morality in ethics (Kant 1788/2015, 5:4, footnote).

According to Korsgaard, the “[f]ailure to grasp this difference or to grasp its depth and significance has bedeviled the English-speaking world’s reception of Kant” which has led to detrimental “misunderstandings of Kant’s ideas”, resulting into “misconstructions of his project [that] have been unusually extreme” (Korsgaard 1996, pp. xii – xiii). In order to avoid even more of such possible misunderstandings (for some pertinent examples cf. Ebbinghaus (1954); Farr (2002); Geiger (2010); Harrison (1957); Melnick (2002); O’Neill (1989); Rossi (1949); Steinberger (1999); Yaffe (1995)), we begin with a brief primer on Kant’s transcendental philosophy in this section.

Going back to the initial quotation from the *Groundwork* above, the main pillars of Kant’s philosophy are the three “higher faculties of cognition”, namely, “understanding”, the “power of judgment”, and “reason” (Kant (1787/1998, B169); cf. also Graben (2024)). First, “understanding” denotes the “faculty of concepts” (Kant 1790/2000a, 5:406), or, in other words the “faculty of rules” for knowledge (Kant 1787/1998, B171). Second, the “power of judgment” comprises the “faculty of subsuming under rules”, i.e., categorization (Kant 1787/1998, B171). And third, “reason” is considered as the “faculty of ideas” of thinking and acting (Kant 1790/2000a, 5:265f), or, equivalently, as the faculty “for the derivation from principles” (Kant 1790/2000b, 20:201), i.e., the capability of logical deduction and the justification of action, or the *sense for the sense*, so to speak (Gabriel 2018).

Each higher mental faculty contains one or more transcendental concepts providing the conditions of possibility of their respective domains: knowledge, judgement, and action. Understanding is firstly constituted by the transcendental “categories”, such as, among many others, “substance”, “causality”, or “community” (Kant 1787/1998, B106). Secondly, the (reflective) power of judgment is related to the transcendental notion of “purpose” (Kant 1790/2000a, 5:184). And thirdly, the “transcendental ideas” (Kant 1787/1998, B378) of (pure practical) reason are “God”, “freedom”, and “immortality” (Kant 1788/2015, 5:3). Although none of those metaphysical ideas can be immanently proven; and there is no possible knowledge about them, they might be postulated for enlightening and for cultivating humanity (Korsgaard 1996, pp. x – xi).

What is now a “good will” for Kant? First, he defines “an intelligence (a rational being)” as a “being capable of actions in accordance with the representation of laws”, and “the causality of such a being in accordance with this representation of laws is his will” (Kant 1788/2015, 5:125). The “good will” entails an “absolute value” for the existence of a rational being (Kant 1790/2000a, 5:443), such that it “has no determining ground; insofar as it can determine the capacity for choice” alone, i.e., “good will” is “practical reason itself” (Kant 1797/1991, 6:213). Following Korsgaard (1996, p. 123)), good will is hence fully efficacious rationality.

Kant’s Maxims of Common Human Understanding

In order to motivate our subsequent reconstruction of Kant’s ethics in terms of systems theory, we discuss his “maxims of the common human understanding”, outlined in the 1790 *Critique of the Power of Judgement* (Kant 1790/2000a, 5:294f). In his later *Metaphysics of Morals* (1797),

³Note that the citation scheme for the *Critique of Pure Reason* (A:1781, B:1787) here deviates from the Academy standard. One cites the first issue with “A” and the second one with “B”, instead.

Kant defines: “A maxim is a subjective principle of action, a principle which the subject himself makes his rule (how he wills to act)” (Kant 1797/1991, 6:225). In this sense, a *maxim* is a subjective representation of a law or a guideline in the mind of an individual agent for the justification of intended actions. Consequently, “action” always means for Kant that an agent behaves rationally by following a maxim. Thus, Kant distinguishes “actions” or “deeds” from unconscious behavior or animal instinct (Kant 1797/1991, 6:223, 6:444). Only *persons*, i.e. rational beings acting upon maxims, are subjects “whose actions can be imputed” to them (Kant 1797/1991, 6:223). In the *Groundwork*, cited above, Kant (1785/2002, 4:438) defines a “person” as a rational being who has “to take its maxims always from its own point of view but also at the same time from that of every other rational being as a universally legislative being”. Postponing the interpretation of the final phrase “universally legislative being” for the moment, Kant here clearly states that persons are characterized by their individual standpoints; his transcendental philosophy is essentially a theory of *perspectivity* (Kant (1785/2002, 4:450, 4:452f, 4:455, 4:458), cf. also Korsgaard (1996); O’Neill (1989); Stratton-Lake (1993)).

Now, we can understand Kant’s notion of the “common human understanding” as a “communal sense”, namely as

[...] a faculty for judging that in its reflection takes account (*a priori*) of everyone else’s way of representing in thought, in order as it were to hold its judgment up to human reason as a whole and thereby avoid the illusion which, from subjective private conditions that could easily be held to be objective, would have a detrimental influence on the judgment. Now this happens by one holding his judgment up not so much to the actual as to the merely possible judgments of others, and putting himself into the position of everyone else, merely by abstracting from the limitations that contingently attach to our own judging; (Kant (1790/2000a, 5:293f); emphases in the original)

Hence, the “communal sense” is constituted by the person’s capability of “putting himself into the position of everyone else” within a community of equal and equitable agents by perspective changing through the abstraction of subjective peculiarities. According to O’Neill (1989, p. 25), the communal sense constrains our understanding through “practices of communication, that can be shared in any possible community.”

This finding brings transcendental philosophy close to the Kant reconstruction of Cassirer (1922) and his interpretation of Einstein’s theory of relativity (Einstein (1920); cf. Giovanelli (2023); Graben (2011); Heis (2014)). Considering the principle of relativity in the form: “All [...] co-ordinate systems are essentially equivalent for the formulation of the general laws of nature”, (Einstein 1920, p. 90)⁴ requires that all systems of reference are equally suitable for the statement of general natural laws; none of such a system is singled out as being superior above any other one, and the formulas of general natural laws must be *invariant* under changing the perspective from one particular reference frame to another one. Cassirer (1922) considers this principle not as a special physical law but rather as a metaphysical, regulative principle that must be met by any possible candidate of a physical law (Giovanelli 2023;

⁴Einstein’s precise formulation (emphases in the original), referring to the general principle of relativity, says: “All Gaussian co-ordinate systems are essentially equivalent for the formulation of the general laws of nature.” We omit “Gaussian” above for the sake of simplicity. Yet, replacing “Gaussian” by “inertial”, yields the (special) principle of relativity, where an inertial system is one in non-accelerated motion (or relative rest).

Graben 2011; Heis 2014), i.e. as a “touchstone of truth” in Kant’s words (Kant 1787/1998, B675).

Now, Kant’s three maxims of common human understanding are stated as:

1. To think for oneself;
2. To think in the position of everyone else;
3. Always to think in accord with oneself. (Kant 1790/2000a, 5:294)

Kant elucidates maxim (1), the “maxim of the understanding”, as the “the maxim of the unprejudiced way of thinking [...] that is never passive”, where passivity means a “heteronomy of reason, [...] called prejudice; and the greatest prejudice of all is [...] superstition”. Contrariwise: “Liberation from superstition is called enlightenment” (Kant 1790/2000a, 5:294). The passivity of reason could be identified with what Arendt (1971, pp. 431, 434) calls “frozen thoughts”, resulting in conventionalism and even dogmatism, in the worst case (Arendt 1971; Graben 2016). And following O’Neill (1989, 25f): “Only those who think for themselves have any contribution to make to a debate or plan. Those who suppress their own voices do not reason; they are mere voiceless echoes, whose parroted words cannot be taken as expressions of judgment or as acts of communication.” Correspondingly, Popper (2013, p. 444) states the “very fact that rationalism is critical, whilst irrationalism must tend towards dogmatism [...]. Criticism always demands a certain degree of imagination, whilst dogmatism suppresses it.” Therefore, the first maxim advises one to be open-minded and also open for rational arguments. Relating these issues to transcendental perspectivity, we conclude that maxim (1) presents the freedom of an agent to deliberately decide and act from her individual perspective.

The second maxim (2) of “the power of judgment” describes an agent’s “broad-minded way of thinking” when “he sets himself apart from the subjective private conditions of the judgment [...] and reflects on his own judgment from a universal standpoint (which he can only determine by putting himself into the standpoint of others” (Kant 1790/2000a, 5:295). The apparent “universal standpoint” does neither refer to a bird’s eye view, nor to an absolutely divine perspective, but rather to the possibility to regard a judgement at stake from the standpoints of all other agents in the community. According to O’Neill (1989, p. 26): “Only those who try to think from the standpoint of everyone else and strive to listen to and interpret others and to see the point of their contributions are genuinely aiming to be ‘fellow workers’ and to avoid maxims to which others cannot agree. [...] Reasoned thinking is governed not by transcendent standards but by the effort to orient one’s thinking in ways that do not preclude its accessibility to others.” Moreover: “One who adopts a maxim of enlarged thought must therefore listen to what others are actually judging and communicating” (O’Neill 1989, p. 46). And finally, she states: “Once we seek to share others’ standpoints, and so become aware of incompatibilities between standpoints, further reflection may lead us toward reappraisals in which coherence is restored” (O’Neill 1989, p. 183). This idea of rational alignment is also acknowledged by Arendt (1977, p. 221): “That the capacity to judge is a specifically political ability in exactly the sense denoted by Kant, namely, the ability to see things not only from one’s own point of view but in the perspective of all those who happen to be present; even that judgment may be one of the fundamental abilities of man as a political being insofar as it enables him to orient himself in the public realm, in the common world”. And, according to Popper (2013, p. 431), one “could then say that rationalism is an attitude of readiness to listen

to critical arguments and to learn from experience. It is fundamentally an attitude of admitting that ‘I may be wrong and you may be right, and by an effort, we may get nearer to the truth’. [...] In short, the rationalist attitude [...] is very similar to the scientific attitude, to the belief that in the search for truth we need co-operation, and that, with the help of argument, we can in time attain something like objectivity.” Hence, the “fact that the rationalist attitude considers the argument rather than the person arguing is of far-reaching importance. It leads to the view that we must recognize everybody with whom we communicate as a potential source of argument and of reasonable information; it thus establishes what may be described as the ‘rational unity of mankind’” (Popper 2013, 431f). Note that the maxim of the “broad-minded”, or, “enlarged way of thinking” also relates to Habermas’ discourse ethics (Habermas 1990; Bordum 2005), where public discourse is conceived as a general thought experiment for finding consent within an open society. Therefore, we conclude that maxim (2) entails, within our framework of transcendental perspectivity, a principle of relativity, demanding the equivalence of all individual perspectives.

For Kant, maxim (3), to which he refers as to the maxim of “reason”, being “namely that of the consistent way of thinking, is the most difficult to achieve, and can only be achieved through the combination of the first two and after frequent observance of them has made them automatic” (Kant 1790/2000a, 5:295). In other words: “Consistency is the greatest obligation of a philosopher and yet the most rarely found” (Kant 1788/2015, 5:24). Following the third maxim, for O’Neill (1989, p. 26), “is no trivial requirement of logical form, but rather a never-ending task. The set of judgments that we independently form, then revise as we shift our standpoint to take account of others’ standpoints, will constantly change, and so may repeatedly fall into inconsistency.” Likewise, we could align the third maxim with the Socratic “I am I”, or his “being one” (Arendt 1971, pp. 439 – 442), for the “principle of agreement with oneself is very old; it was actually discovered by Socrates, whose central tenet [...] is contained in the sentence: ‘Since I am one, it is better for me to disagree with the whole world than to be in disagreement with myself’” (Arendt 1977, p. 220). Notice however, that Kant’s “consistency” must not to be confused with another notion of an “insane consistency, [...] the chief psychological support of totalitarian fiction” (Arendt 1976, p. 392), that is thriven “on this escape from reality into fiction, from coincidence into consistency” by totalitarian propaganda (Arendt 1976, p. 352): “Nothing matters but consistency. [...] The aggressiveness of totalitarianism [...] if it feverishly seeks to expand, it does so neither for expansion’s sake nor for profit, but only for ideological reasons: to make the world consistent, to prove that its respective supersense has been right” (Arendt 1976, p. 458). Relating consistency to transcendental perspectivity again, we find that maxim (3) must refer to the principle of invariance, as known from Einstein’s relativity theories (Einstein 1920, p. 90). Therefore, consistent thinking means that our thoughts could persist upon arbitrary perspective changes.

Clearly, the three maxims (1 – 3) of the common human understanding are principles of rationality. Therefore, they could be straightforwardly augmented by a fourth, supreme maxim:

4. Be rational! Be real!

Maxime (4) provides the condition of possibility of rational agency.

Individualistic Ethics

Now, we are prepared for our perspectivistic reconstruction of Kantian ethics, where we essentially follow Geiger (2015), who suggests

to consider “single agents and conceive of moral laws in the singular and determinately”, on the one hand, and to regard “a community of diverse moral lawgivers who each have as their legislative end the ideal of a comprehensive system of moral laws”, on the other hand (Geiger 2015, p. 396). Thus, we present first our account of *individualistic ethics* in this subsection, and that of *interactional ethics* in the following one.

For an individualistic ethics, Kant suggests two different formulations of the moral law, i.e. of the famous *categorical imperative*, that are discussed under the labels “formula of universal law” (FUL) and “formula of an end in itself” (FEI), sometimes also called the “formula of humanity” (Guyer 2002), that differ from each other through their different perspectives (O'Neill 1989, p. 142), namely either considering “self as agent and other as agent” (Stratton-Lake 1993, p. 335).

The Formula of Universal Law (FUL): The first formulation of the categorical imperative addresses an individual agent: “What ought I to do?” (O'Neill 1989, p. 142). Kant's response, known as the formula of universal law, says:

Act only in accordance with that maxim through which you can at the same time will that it become a universal law. (Kant 1785/2002, 4:421), cf. also Kant (1788/2015, 5:30), Kant (1797/1991, 6:226); emphasis in the original)

Interpreting this formula in the terms of the maxims of communal sense, discussed above, reveals that an agent who deliberately adopts a particular maxim (according to maxim (1)), i.e. a subjective rational principle for justifying action, must can will, “at the same time” of acting, that the chosen maxim may become a “universal law”, i.e. a law that equally commits all other rational agents “through” the adoption of that very maxim. In order to assess whether the agent's maxim could be willed as a universal law, the agent has principally to assume the perspectives of any other agent, according to maxim (2). Only, if the chosen maxim appears to be invariant under perspective change (according to maxim (3)), it could qualify as a universal law.

Hence, our reconstruction entails an interesting analogy between the FUL and Einstein's principle of relativity (Einstein 1920, p. 90). Whereas the latter is not a particular law of nature, but rather a methodological principle for testing the admissibility of possible candidates of physical laws, the former is not a particular moral law, but, by contrast, a transcendental principle for probing the suitability of law candidates, namely of subjective maxims, as universal moral laws. By iterating the three maxims of common human understanding: (1) free selection of possible maxims of action, (2) transporting ourselves into all possible perspectives (Kant 1785/2002, 4:453), and (3) consistent reasoning, we obtain the *invariant of morality*, namely what Kant calls “the mere form of universal lawfulness” (Kant 1793/1996, 6:3). In this sense, the FUL is nothing else than a principle of relativity. However, this does not lead to moral relativism at all, because only its invariant, *universal lawfulness*, vindicates general objective morality.

Finally, we give a particular example for the perspectival iteration process. Assume, an agent chooses the maxim of *mindful non-action* (Purser 2019, Conclusion, Wait-and-see-ism) prescribing the avoidance of any rational interaction with others. Could this maxim be generalized to a universal moral law? Apparently not, because it immediately leads into a “contradiction in conception”, when “the attempt to universalise the maxim entails an incoherent world” (Stratton-Lake 1993, p. 332). This particular maxim, not to act, implies the maxim, not to act according to any maxim, and hence to the maxim that one must not adopt any maxim at all, i.e. a world where nobody can act upon maxims, following

the categorical imperative. In such a world rational agency loses its sense.

From this point of view, the categorical imperative becomes *self-referential*, in analogy with Gödel's incompleteness theorems of meta-mathematics (Gödel 1931; Hofstadter 1979), an issue Kant was also acutely aware about, when he refers to the first maxim of common understanding as to the “maxim of reason's *self-preservation*” (Kant 1786/1996, 8:146, footnote); cf. Kant (1785/2002, 4:430), O'Neill (1989, p. 26)). Thus, we may rephrase FUL in the form of a “strange loop” (Hofstadter 1979) as a *strange imperative* as follows:

Act only in such a way that your actions can be the conditions of the possibility of subsequent actions.

Because “action” always refers to maxims and grounds in the mental representation of rational agents, the FUL of the categorical imperative must thereby be observed in order to be obeyable.

Interestingly, a quite similar formulation has been put forward as a “constructivist ethical imperative” (cf. Bagnoli (2021); Bagnoli (2026) for related constructivist approaches) by Foerster (2003, p. 282):

Act always so as to increase the number of choices (Foerster 2003, p. 227), cf. Pangaro (2011)).

Clearly, the number of choices counts the choices of possible action. Hence, an agent obeying von Foerster's imperative also acts in such a way that her actions become the conditions of the possibility of subsequent actions. Note that related forms of self-referential ethics refer to the “right to have rights” Arendt (1976, p. 296) and to the “the responsibility for its own responsibility” Havel (1989, p. 330).

The Formula of an End in Itself (FEI): Besides a maxim, Kant claims, rational action also requires an object that he identifies with an *end*: “An end is an object of free choice, the representation of which determines it to an action (by which the object is brought about). Every action, therefore, has its end” (Kant 1797/1991, 6:384f). The end, he states, is the “matter (an object of free choice) [...] of pure reason” (Kant (1797/1991, 6:380); Kant (1788/2015, 5:41)). Pure practical reason is, so Kant, our normative faculty which assigns value to the objects of desire (Kant 1790/2000a, 5:207, 5:209). Hence, practical reason has, according to the interpretation of Korsgaard (1996, p. 122), “a value-conferring Status”.

Kant's formulation of the categorical imperative through the ends in itself (FEI) now reads:

Act so that you use humanity, as much in your own person as in the person of every other, always at the same time as end and never merely as means. (Kant 1785/2002, 4:429), cf. Kant (1788/2015, 5:87); Kant (1797/1991, 6:395); emphasis in original).

In order to appropriately interpret this formulation, we first have to recognize that also FEI refers to action, and hence, to an agent's behavior that is entailed by the mental representation of a rational principle, i.e. by a maxim. Second, “humanity” must not be interpreted as mankind, i.e. the mass of human beings on earth, or as a biological species, but rather as the human essence, comprised by rationality (Kant 1785/2002, 4:430), cf. also Korsgaard (1996); Guyer (1995); Rollin (1976)). Humans are only a particular class of intelligent, rational beings (Kant 1788/2015, 5:87), Kant (1790/2000a, 5:233)). Kant deliberately “interchanges the terms ‘humanity’ and ‘rational nature’” in his writings (Korsgaard 1996, p. 110), for humans are the only rational beings we

currently know about. The next important concept is thirdly that of a “person”, again referring to a rational agent who is able to take a perspective and having an attitude toward a chosen perspective (Kant (1785/2002, 4:428, 4:438); Kant (1797/1991, 6:223)). Fourth, using humanity, “as much in your own person as in the person of every other”, again refers to maxim (2), i.e. the principle of relativity, of common human understanding, by advising the agent to deliberately adopt the perspectives of other agents. Fifth, FEI obligates the agent not to use oneself or other persons “merely as means”. Instead, agents are, sixthly, supposed to treat other persons “at the same time” of acting “as end”; in other words, humans, and more generally, rational beings, have to be treated as “ends in themselves” (Kant 1785/2002, 4:428). What does this mean?

The transcendental concepts of *ends* and *purposes* are the essential topics of Kant's *Critique of the Power of Judgement* (Kant 1790/2000a). Here, Kant (1790/2000a, 5:180) defines: “Now since the concept of an object insofar as it at the same time contains the ground of the reality of this object is called an end, and the correspondence of a thing with that constitution of things that is possible only in accordance with ends is called the purposiveness of its form”. Similarly: “For the represented effect, the representation of which is at the same time the determining ground of its production in an intelligently acting cause, is called an end” (Kant 1790/2000a, 5:426). Or, alternatively, “an end is the object of a concept insofar as the latter is regarded as the cause of the former (the real ground of its possibility); and the causality of a concept with regard to its object is purposiveness (*forma finalis*). Thus where not merely the cognition of an object but the object itself (its form or its existence) as an effect is thought of as possible only through a concept of the latter, there one thinks of an end. The representation of the effect is here the determining ground of its cause, and precedes the latter” (Kant 1790/2000a, 5:220).

These quotations indicate that Kant considers ends in the Aristotelian tradition of final causes where the effects precede their causes. In other words, ends are concepts of objects that precede their material realization which is entailed by the mental representation of its concept. Thus, the ideas of a house, of a fork, or of a violin, precede their material realizations in the minds of the architect, of the cutlery-maker, or of the instrument-builder, respectively. They are the final causes of the corresponding objects.

Another important conception of Kant's *Critique of the Power of Judgement* is that of a “*natural end*”, which is an “an organized and self-organizing being”, “if it is cause and effect of itself (although in a twofold sense)”, namely an “organized product of nature” in which “everything is an end and reciprocally a means as well” (Kant 1790/2000a, 5:370f, 5:373, 5:376). Apparently, Kant here already introduces the idea of *autopoiesis* of biological organisms (Maturana et al. 1998).

Now, it is straightforward to clarify the mysterious notion of an “end in itself” (Kant (1785/2002, 4:428); Kant (1788/2015, 5:87)). An end in itself must be realized in order to be realizable. Hence, it is that mental representation of a rational agent which is the condition of possibility for entertaining mental representations of rational agents at all. Thus, we arrive at another self-referential *gödelization* (Gödel 1931; Hofstadter 1979) of the categorical imperative:

Act so that the ends of your actions are the conditions of the possibility to entertaining ends.

Consequently, FEI essentially guarantees the self-preservation of humanity through moral action (Kant (1785/2002, 4:430); Kant

(1786/1996, 8:146, footnote); cf. O'Neill (1989, p. 26)): Human rationality as an end in itself must be preserved in order to be preservable. Therefore, also the FEI of the categorical imperative must be observed in order to be obeyable.

This interpretation is compatible with those of previous interpreters, such as Korsgaard (1996); Geiger (2015); Guyer (1995); Hill (1980); O'Neill (1989), or Stratton-Lake (1993), who all agree upon the point that our normative rational nature, our capability to set arbitrary ends makes rationality an end in itself (Kant 1797/1991, 6:381, 6:392).

Interactional Ethics

Sofar, we have reconstructed Kant's individualistic ethics what a single agent is ought to do (Geiger 2015), both from the agent's ego perspective, leading to FUL and from the perspectives of the affected alters, which entails FEI (O'Neill 1989; Stratton-Lake 1993). Next we switch to the perspective of a community or open society (Popper 2013; Etzioni 2000) and discuss Kant's interactional ethics (O'Neill 1989; Stratton-Lake 1993).

The Formula of Autonomy (FA): The first step towards interactional ethics comprises another iteration of the three maxims of common human understanding:

The ground of all practical legislation, namely, lies *objectively in the rule* and the form of universality, which makes it capable of being a law (at least a law of nature) (in accordance with the first principle [FUL]), but *subjectively* it lies in the *end*; but the subject of all ends is every rational being as end in itself (in accordance with the second principle [FEI]); from this now follows the third practical principle of the will, as the supreme condition of its harmony with universal practical reason, the idea of the will of every rational being as a will giving universal law. (Kant (1785/2002, 4:431), emphases in original)

Here, the objective “form of universality” is the invariant of maxim (3) under the relativity principle of maxim (2), obtained from the freedom of choosing subjective, individual ends, according to maxim (1). The resulting formula of autonomy (FA) says:

Do no action in accordance with any other maxim, except one that could subsist with its being a universal law, and hence only so that the will could through its maxim at the same time consider itself as universally legislative. (Kant (1785/2002, 4:434), emphases in original)

Therefore, every agent must conceive herself not only as the “author” of her subjective maxims (Kant 1785/2002, 4:431), but simultaneously as a universal legislator (Kant 1785/2002, 4:432, 4:438). Kant calls this principle “the principle of the autonomy of the will, in contrast to [...] heteronomy” (Kant 1785/2002, 4:433). Thus, *autonomy* means self-legislation, in opposition to *heteronomy* as alien-legislation (Kant 1785/2002, 4:431). Moreover, Kant relates autonomy also with his conception of *dignity*: “but that which constitutes the condition under which alone something can be an end in itself does not have merely a relative worth, i.e., a price, but rather an inner worth, i.e., dignity” (Kant 1785/2002, 4:435). Eventually, for “regarding the subject which is an end in itself: if that representation is to have its total effect on me, then its ends must as far as possible also be my ends” (Kant 1785/2002, 4:430).

Korsgaard (1996, p. 23) suggests the following interpretation: “To explain obligation, we need an imperative that binds us unconditionally.

But this means that moral motivation, if it exists, must be autonomous. There can only be one reason why human beings must obey the moral law, and that is that we give that law to ourselves". And furthermore: "In the negative sense, to treat another as an end in itself is to respect her autonomy — to leave her actions, decisions, and ends to her own choice. But this respect gets its most positive and characteristic expression at precisely the moments when we must act together. [...] If my end requires your act for its achievement, then I must let you make it your end too. Both what I choose and the way I choose it must reflect this constraint. You must be free to choose whether you will contribute to the success of my project or not [...] my own ends must be the possible objects of universal legislation, subject to the vote of all. And this is how I realize my autonomy. Paradoxically if you like, my ends and actions are most truly my own when they are chosen under the restrictions of a possible reciprocal relation — a kind of friendship — with everyone" (Korsgaard (1996, p. 193); cf. also Geiger (2015); Guyer (1995); Nelson (2008); Stratton-Lake (1993)). Thereby, the principle of autonomy establishes the foundation of interactional ethics.

Importantly, FA implies now that self-legislation and autonomy are crucial duties that cannot be abandoned in favor of heteronomy and obedience. This refutes Eichmann's completely absurd Kant interpretation that he plead at his Jerusalem trial in 1961 (Arendt (1963), cf. Brott (2016); Hanna (2024)). Arendt retorts: "According to Kant, no man has the right to obey".⁵ Thus, subordination never justifies abdication of power, and hence of autonomy.

The Formula of a Realm of Ends (FRE): In the *Groundwork*, Kant makes an interesting remark about the connection of FUL, FEI, and FA, discussed above, regarding the category table in the *Critique of Pure Reason* (Kant 1787/1998, B171). He states:

All maxims have, namely,

1. a form, which consists in universality, and then the formula of the moral imperative is expressed thus: 'That the maxims must be chosen as if they are supposed to be valid as universal laws of nature' [FUL];
2. a matter, namely an end, and then the formula says: 'That the rational being, as an end in accordance with its nature, hence as an end in itself, must serve for every maxim as a limiting condition of all merely relative and arbitrary ends' [FEI];
3. a complete determination of all maxims through that formula, namely 'That all maxims ought to harmonize from one's own legislation into a possible realm of ends[,] as a realm of nature' [FA]. A progression happens here, as through the categories of the unity of the form of the will (its universality), the plurality of the matter (the objects, i.e., the ends), and the allness or totality of the system of them. (Kant 1785/2002, 4:436)

The FRE then states that

⁵In the German original: "Kein Mensch hat bei Kant das Recht zu gehorchen" (Arendt (2007); our own translation).

⁶The comma, which is crucial in the German original, is not present in the edition of Kant (1785/2002). Here we follow Guyer (1995, p. 354) who suggests its preservation in the translation.

[...] every rational being must act as if it were through its maxims always a legislative member in a universal realm of ends. The formal principle of these maxims is: 'Act as though your maxim should serve at the same time as a universal law (for all rational beings)'. (Kant 1785/2002, 4:438)

What does this formulation mean? "By a realm", Kant understands "the systematic combination of various rational beings through communal laws" (Kant 1785/2002, 4:433).⁷ "Morality thus consists in the reference of all action to that legislation through which alone a realm of ends is possible" (Kant 1785/2002, 4:434). Every "rational being belongs as a member to the realm of ends if in this realm it gives universal law but is also itself subject to these laws" (Kant 1785/2002, 4:433).

Therefore, the "realm of ends" must be interpreted in system-theoretic terms as a community of interactive agents that are reciprocally connected through communication and operation networks, pursuing individual ends as well as collaborative projects. This corresponds with the interpretation of Korsgaard (1996, p. 212): "When we enter into relations of reciprocity, and hold one another responsible, we enter together into the standpoint of practical reason, and create a Kingdom of Ends on earth." Moreover, Kant (1785/2002, 4:434) argues that in "the realm of ends everything has either a price or a dignity. What has a price is such that something else can also be put in its place as its equivalent; by contrast, that which is elevated above all price, and admits of no equivalent, has a dignity." Here, he stresses again the dignity of human autonomy and rationality as an unconditioned end in itself. By contrast, conditioned ends, and also their means, have a "price", i.e. a market exchange value (Kant 1785/2002, 4:434f).

The realm of ends splits into two entangled subsystems. First, the "culture of skill is certainly the foremost subjective condition of aptitude for the promotion of ends in general" (Kant 1790/2000a, 5:431f). Second, there must be a "culture of training (discipline)", which "consists in the liberation of the will from the despotism of desires" (Kant 1790/2000a, 5:432). The "culture of skill" (or likewise of "prudence" (Kant 1790/2000a, 5:172)) can be identified with the "civil society" (Kant 1790/2000a, 5:432), where citizens are equally — and with equity — subjected to "public juridical laws (which are all coercive laws)" (Kant (1793/1996, 6:95), Kant (1793/1996, 6:98f)). Within the culture of skill, people are allowed to use each other also as means. When a person enters a taxi for the aim of transportation, she uses the chauffeur as a means for that end. This is not forbidden by the categorical imperative as long as the person does not use others "merely" as means for her ends.

By contrast, we have to identify the "culture of discipline" with an "ethico-civil" society in which agents "are united under laws without being coerced, i.e. under laws of virtue alone" (Kant (1793/1996, 6:95), Kant (1793/1996, 6:98f)). Here, "laws of virtue alone" are just the laws of moral autonomy. Following Stangneth (2000), we may call this ideal the "culture of sincerity", alluding to Kant's dedication in the 1797

⁷Note, that Kant always writes "Reich der Zwecke" in German and nowhere something like *Königreich der Zwecke* ("kingdom of ends"). Thus, the contemporary translation of Wood (Kant 1785/2002), using "realm of ends" seems most appropriate. Moreover, a webglimpse search for "Königreich", queried 07/26/2026, yielded only 3 hits, all related to real political entities of Kant's time (<https://korpora.org/cgi-bin/wg2/webglimpse.cgi>). Finally, Kant was an enthusiastic sympathizer of the US-American and French revolutions holding a strong republican attitude (Korsgaard 1996, pp. 32 – 34). In fact, in her book *Creating the Kingdom of Ends*, Korsgaard (1996, p. 23) writes: "The Kingdom of Ends is an ideal. It is 'a systematic union of different rational beings through common laws', a republic of all rational beings." How could a kingdom be a republic, we thus ask?

Religion Within the Boundary of Mere Reason:

Oh sincerity! You Astrea, who have fled from the earth to heaven, how are you (the foundation of conscience, and hence of all inner religion) to be drawn down from there to us again? I can admit, though it is much to be deplored, that straightforwardness (saying the *whole* known truth) is not to be found in human nature. But we must be able to demand *sincerity* (that *everything said* be said with truthfulness). (Kant (1793/1996, 6:190, footnote), emphasis in original)

In that late publication, Kant introduces another equivalent of the realm of ends, namely the “church invisible” (Kant 1793/1996, 6:101), where “Religion is (subjectively considered) [as] the recognition of all our duties as divine commands” (Kant 1793/1996, 6:153f). Note, however, that Kant considers the realm of ends, the ethical community, the culture of sincerity, or the church invisible as “obviously only an ideal”, and not as a political utopia (Kant (1785/2002, 4:433); cf. Popper (2013, 441f), Mariña (2022)).

What are the moral duties of the members of the realm of ends? Explicitly, Kant only identifies two “ends that are also duties”, namely “one’s own perfection and the happiness of others” (Kant 1797/1991, 6:385). The former, “one’s own perfection” refers to the idea of “culture” again:

Natural perfection is the *cultivation* of any *capacities* whatever for furthering ends set forth by reason. That this is a duty and so in itself an end, and that the cultivation of our capacities, even without regard for the advantage it affords us, is based on an unconditional (moral) imperative rather than a conditional (pragmatic) one, can be shown in this way. The capacity to set oneself an end — any end whatsoever — is what characterizes humanity (as distinguished from animality). Hence there is also bound up with the end of humanity in our own person the rational will, and so the duty, to make ourselves worthy of humanity by culture in general, by procuring or promoting the *capacity* to realize all sorts of possible ends, so far as this is to be found in man himself. In other words, man has a duty to cultivate the crude predispositions of his nature, by which the animal is first raised into man. It is therefore a duty in itself. ((Kant 1797/1991, 6:391f), emphasis in original)

Regarding the latter, “the happiness of others”, by contrast,

[...] it is more difficult to *do good*, especially if it is to be done not from affection (love) for others but from duty, at the cost of forgoing the satisfaction of concupiscence and of active injury to it in many cases. The reason that it is a duty to be beneficent is this: Since our self-love cannot be separated from our need to be loved (helped in case of need) by others as well, we therefore make ourselves an end for others; and the only way this maxim can be binding is through its qualification as a universal law, hence through our will to make others our ends as well. The happiness of others is therefore an end that is also a duty. ((Kant 1797/1991, 6:393), emphasis in original)

Thus, benevolence might result from the principle of relativity, discussed as maxim (2) above. Note, however, that Popper (2013, 602f,

footnote 2) is rather sceptical about that maxim, since “there is, from the ethical point of view, no symmetry between suffering and happiness, or between pain and pleasure. Both the greatest happiness principle of the Utilitarians and Kant’s principle ‘Promote other people’s happiness ...’ seem to me (at least in their formulations) wrong on this point which, however, is not completely decidable by rational argument.” Accordingly, he says:

But of all political ideals, that of making the people happy is perhaps the most dangerous one. [...] But [...] the attempt to make heaven on earth invariably produces hell. It leads to intolerance. It leads to religious wars, and to the saving of souls through the inquisition. And it is, I believe, based on a complete misunderstanding of our moral duties. It is our duty to help those who need our help; but it cannot be our duty to make others happy, since this does not depend on us, and since it would only too often mean intruding on the privacy of those towards whom we have such amiable intentions. The political demand for piecemeal (as opposed to Utopian) methods corresponds to the decision that the fight against suffering must be considered a duty, while the right to care for the happiness of others must be considered a privilege confined to the close circle of their friends. (Popper 2013, 441f)

Therefore, Popper (2013, p. 48) recommends, “perhaps not so much a claim to be made happy [...], but a claim not to be made unhappy, where it can be avoided. They have a claim to be given all possible help, if they suffer.”

Yet, Kant’s program to look for ends that must also be duties, now leads to another viable interpretation of the realm of ends in terms of systems theory. In section we have identified the ends in itself as those ends that must be realized in order to be realizable, a *strange loop* according to Hofstadter’s reconstruction of Gödel’s incompleteness theorems of metamathematics (Gödel 1931; Hofstadter 1979). Now, we suggest the same procedure for the realm of ends: An end that must also be a duty, is an end that must be sustained in order to be maintainable. But this should be the realm of ends itself, or, in other words: a *healthy democratic open society* (Minkler 1989; Nyswander 1967; Popper 2013; Taswell 2025b; Zhu et al. 2025).

Thus, our reformulation of FRE reads

Act so that your actions within a healthy democratic open society are the conditions of the possibility for the maintenance of a healthy democratic open society.

Lattice Gauge Ethics

In the last section, we have reconstructed Kant’s individualistic and interactional ethics in terms of self-referential systems. Action, as we have endorsed, is rational behavior that is reasonably grounded in maxims, namely subjective principles as mental representations of behavioral laws. In order to preserve rational agency in this sense, an agent is committed to the categorical imperative in either one of the four different formulas. The formula of universal law (FUL) must be observed in order to be obeyable subsequently. The agent has to realize the end in itself, i.e. rational agency, in order to make it subsequently realizable (FEI). The formula of autonomy (FA) states that the agent is only subjected to her own universal self-legislation. And finally, the realm of ends (FRE), as given as a healthy democratic open society, must be sustained for its own subsequent maintenance.

Now, in the present section, we suggest a system-theoretical toy-model to further elucidate our reconstruction. Such models are often employed in the social sciences to describe emergent deterministic-like collective behavior of a large number of interacting agents (Castellano et al. 2009; Llabrés et al. 2026). Here, however, we are interested in emergent morality which arises from free, autonomous deliberation, induced by Kantian interactional ethics.

Individualistic Systems

Let us consider an individual agent, x , entertaining a set of maxims M_x for the furtherance of arbitrary ends. What should the agent do? One maxim, e.g., could be: 'I should maximize my profit'; a second, e.g., 'I must not cheat'. Often, the agent has to resolve decision problems between conflicting maxims, e.g. when maximization of profit is only possible by cheating. Thus, it makes sense to model moral considerations in terms of mathematical decision theory (Tversky et al. 1981; Busemeyer and Townsend 1993). A prospect for a decision model is given as a probability distribution over possible consequences. As far as this distribution must be mentally represented in the consciousness of agent x , it could be regarded as a *belief state* Ψ_x . In the novel *quantum cognition* approach (Busemeyer and Bruza 2012; Khrennikov 2009; Pothos et al. 2013), belief states are vectors in a Hilbert space that allow the calculation of prospects from complex probability amplitudes, on the one hand, and the comparison of projections onto different subspaces, on the other hand. Probability amplitudes, in contrast to prospect probabilities, are suited for linear superposition for the description of complementary properties, as appearing in the *bounded rationality* framework (Tversky et al. 1981; Blutner et al. 2016). Recently, Caldas et al. (2026) have discussed moral judgement within this new framework. Interestingly, the puzzles of bounded rationality arise by not taking the social relativity principle (2) properly into account (Blutner et al. 2016).

For the aim of the present study, namely a system-theoretic reconstruction of Kantian ethics, we pursue a similar strategy. First of all, we assume for the sake of simplicity that the set of maxims M_x , the agent's degrees of freedom, be finite and does not change over time.⁸ Then, we could assign an ordinal number $q_m \in \mathbb{N}$ to every maxim $m \in M_x$. For the example above, we could assign number $q = 1$ to the first maxim of profit maximization, and number $q = 2$ to the second maxim of not-cheating. In general, $1 \leq q_m \leq n$ with $n \in \mathbb{N}$ the cardinality of M_x .

Next, the freedom to choose any particular maxim according to the maxim (1) of common human understanding, can be restored by introducing a preference order over M_x (Arrow 1950). Mathematically, this is realized by a (possibly time-dependent) permutation $\pi(q_m)$ of the original numbering q_m . We now refer to such a permutation as to the belief state $\Psi_x = \pi(q_m)$, which is a vector of n non-repeating, consecutive integers, i.e., $\Psi_x \in \mathbb{N}^n$. For our example agent, there are two possible belief states:

$$\Psi_{x1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \Psi_{x2} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \quad (5)$$

In state Ψ_{x1} the first maxim of profit maximization is higher ranked than the second maxim not to cheat, i.e. $1 \succ 2$, because $\Psi_{x1}(1) = 1$ and $\Psi_{x1}(2) = 2$, when *rank* ($\succ$) corresponds to the successive

coordinates of the maxim numbers q_m in the belief vector. By contrast, in state Ψ_{x2} the preference order is reversed, $2 \succ 1$, for $\Psi_{x2}(1) = 2$, but $\Psi_{x2}(2) = 1$.

This construction is supported by Kant's view that the rearrangement of maxims in a preference ordering is itself an action that is justified by a meta-maxim. Following Kant, there is a "propensity of the power of choice to maxims" that allows the "subordination" of one maxim under another one and a "reversal" of their ordering due to "a free power of choice" (Kant (1793/1996, 6:30, 6:36); cf. Guyer (1995, p. 376)).

Accordingly, we could express the change of an individual's preference order by a matrix operator $\vec{\pi}$, that applies to a belief state vector, e.g.

$$\vec{\pi} \cdot \Psi_{x1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \Psi_{x2}, \quad (6)$$

where the permutation matrix

$$\vec{\pi} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \rho(S_2)$$

belongs to a representation ρ of the symmetric group S_2 over a two-dimensional representation space. In the general setting with n maxims and belief states $\Psi_x \in \mathbb{N}^n$, the permutation operators belong to a representation of the symmetric group S_n .

Interactional Systems

In the next step, we consider a community of N agents $x, y, z, \dots$, each one possessing its own individual belief state, $\Psi_x, \Psi_y, \Psi_z, \dots$. In order to allow the comparison of those states, we assume that all agents entertain the same finite set of possible maxims M , though possibly in different preference orders. Then, there is one basic belief space, $\mathcal{H} = \mathbb{R}^n$, occurring in as many local replicas, $\mathcal{H}_x, \mathcal{H}_y, \mathcal{H}_z, \dots$ as there are agents in the system, namely $N \in \mathbb{N}$.

Following Kant again, those agents share the "universal capacity for the communication of the state of mind in the given representation" (Kant 1790/2000a, 5:217), since "[c]ognitions and judgments must, together with the conviction that accompanies them, be able to be universally communicated" (Kant 1790/2000a, 5:238). Therefore, agents must be connected through communication channels that form a communication network (Arendt (1976, pp. 433, 465); Mack (1995)), for "influencing human beings and their will" (Kant 1790/2000a, 5:172). Thus, agents form communities and societies through interaction: "Our political life rests on the assumption that we can produce equality through organization, because man can act in and change and build a common world, together with his equals and only with his equals" (Arendt 1976, p. 301).

Social networks can display different kinds of topologies, reflecting their connectivity patterns (Mizraji 2026). In graph theory, a finite social network $\Gamma = (V, E)$ is given through a set of N vertices V , describing agents, and edges $E \subset V \times V$, featuring their communication channels. Algebraically, the topology of a network of N vertices is described by its binary adjacency matrix, $\vec{A} \in M(N \times N)$, such that $A_{ik} = 1$, if there is an edge $k \rightarrow i$, and $A_{ik} = 0$, if no such edge is present at all.⁹

In Figure 1 we present four different kinds of such topologies for a small population of $N = 6$ individuals that have been investigated in our previous study on healthy aging in open societies (Zhu et al.

⁸This assumption seems to be at variance with the freedom of thinking, i.e. maxim (1), which implies the possibility to create arbitrary new maxims in imagination. However, it does not prevent free deliberation of action, as we show in the sequel.

⁹Note the convention for the adjacency matrix: first index: destination, second index: source.

2025). Figure 1(a) depicts a *segregated society* of strictly isolated individuals. This corresponds to the “atomized society” that Arendt (1976, pp. 235, 352) relates to the decrease of imperialism and with the rise of totalitarianism (Arendt 1976, 318ff). Its adjacency matrix equals the N -dimensional identity matrix

$$\vec{A}_{\text{seg}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = 1 \quad (7)$$

Note that each agent is only connected to herself, accounting for her personal identity and her conscious conscience, in our model.

The transition from a segregated to a *totalitarian society*, sketched in Fig. 1(b), is explained by the role of the “mass movements”, such as NSDAP in Germany and Bolsheviki in the Soviet Union (Arendt 1976, p. 323). “Totalitarian movements are mass organizations of atomized, isolated individuals. Compared with all other parties and movements, their most conspicuous external characteristic is their demand for total, unrestricted, unconditional, and unalterable loyalty of the individual member. This demand is made by the leaders of totalitarian movements even before they seize power” (Arendt 1976, p. 323). According to Arendt (1976, p. 325), “the totalitarian leader is nothing more nor less than the functionary of the masses he leads; he is not a power-hungry individual imposing a tyrannical and arbitrary will upon his subjects. Being a mere functionary, he can be replaced at any time, and he depends just as much on the ‘will’ of the masses he embodies as the masses depend on him. Without him they would lack external representation and remain an amorphous horde; without the masses the leader is a nonentity.”

However exchangeable the totalitarian leader might be, the network topology of totalitarianism is caricatured by that social network where one distinguished individual, the dictator, directly rules over all the others members in the mass. This is expressed by the adjacency matrix

$$\vec{A}_{\text{tot}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

that differs from $\vec{A}_{\text{seg}}$ in the first column, where $A_{k1} = 1$ for the presence of all edges $1 \rightarrow k$, $1 \leq k \leq N$. In terms of graph theory, $\vec{A}_{\text{tot}}$ is (though neglecting the self-connections) the adjacency matrix of a tree. Note that according to Arendt (1976, p. 314), totalitarianism also strives to “extinguishing individual identity permanently”, through the destruction of the identity links in the network (Arendt 1976, p. 435).

Another network topology, plotted in Fig. 1(c), is characteristic for a military “command-obedience” chain (Gil 2016), reflected by a single path through the network’s graph, to which we therefore refer as a

military society. Its adjacency matrix

$$\vec{A}_{\text{mil}} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

consists of the main diagonal of the identity matrix and a second diagonal, indicating the neighborhood edges $k+1 \rightarrow k$ throughout the path. Such a command-obedience chain is characteristic for *authority* which “rests neither on common reason nor on the power of the one who commands; what they have in common is the hierarchy itself” (Arendt 1977, 92f).

Finally, Fig. 1(c) displays what we call the “open society” (Popper 2013, p. 165) (or “democratic society” in the motivating study by Gil (2016)). In this case, the society’s graph is *strongly connected*, i.e., every node is connected to every other node through a path of sufficient length. Algebraically, this property is characterized by its adjacency matrix $\vec{A}_{\text{op}}$ being *irreducible*, indicating that there is an integer number $r \in \mathbb{N}$, such that $\vec{A}_{\text{op}}^r$ is strictly positive, $\vec{A}_{\text{op}}^r > 0$, i.e. $(\vec{A}_{\text{op}}^r)_{ik} > 0$ for all indices $1 \leq i, k \leq N$.

In our particular example from Fig. 1(c), the graph of the open society is even *completely connected*, in which case every node is connected to every other node through an edge in E . Therefore, its adjacency matrix

$$\vec{A}_{\text{op}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

consists of ones only, such that $\vec{A}_{\text{op}}^1 > 0$.

For the open network topology, we prove two important propositions. First, for any subset of agents $\{x_1, x_2, x_3, \dots, x_s\}$, one can construe a path $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_s$ throughout the graph. For that aim, we assume there are integers $r_1, r_2, \dots, r_{s-1}$, such that the matrix powers yield $(\vec{A}_{\text{op}}^{r_1})_{i_2 i_1} > 0$, $(\vec{A}_{\text{op}}^{r_2})_{i_3 i_2} > 0$, etc, indicating that vertices x_1 and x_2 are connected by a path of length r_1 , x_2 and x_3 are connected by a path of length r_2 , etc. Then, due to the irreducibility of the adjacency matrix, the matrix product $\vec{A}_{\text{op}}^{r_1} \cdot \vec{A}_{\text{op}}^{r_2} \cdot \dots \cdot \vec{A}_{\text{op}}^{r_{s-1}}$ exhibits a non-zero element $i_s i_1$ for the transition $x_1 \rightarrow x_s$, i.e. the path $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_s$. Moreover, by letting $x_s = x_1$, one also finds loops of arbitrary length, connecting every vertex with itself throughout the network.

Interactional ethics is challenged by the problem to identify universal moral values that align across the entire society. Clearly, such values could principally be imposed by a dictator in a despotic or totalitarian society. But this crucially depends on the “good will” of the leader; and most tyrannies were morally evil and completely corrupt in history (Arendt 1976; Popper 2013). A principal problem for morality was already recognized by Kant:

Human beings [...] mutually corrupt one another’s moral predisposition and, even with the good will of each individual, because of the lack of a principle which unites them, they deviate through their dissensions from the common goal

of goodness, as though they were *instruments of evil*, and expose one another to the danger of falling once again under its dominion. (Kant (1793/1996, 6:96), emphasis in original)

This issue is nicely illustrated by the so-called *Condorcet cycles* in political science (Arrow 1950; Bochsler 2010). Assume three agents, x, y, z and three maxims 1, 2, 3 in their respective belief spaces. We may interpret maxim 1 as the maxim of profit maximization and maxim 2 with the maxim of non-cheating, as above. Additionally, maxim 3 could be identified as the supreme rationality principle (4) from Sect. . Then the belief states

$$\Psi_x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \Psi_y = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \Psi_z = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \quad (8)$$

are not commensurable in terms of democratic decision finding, because x, y agree that $2 \succ 3$, while y, z agree that $3 \succ 1$, and x, z agree that $1 \succ 2$. Thus, a majority of agents (x, z) vote $1 \succ 2$, another majority (x, y) votes $2 \succ 3$, but nevertheless a third majority (y, z) votes $3 \succ 1$, resulting in a *cyclical majority paradox*.

Lattice Gauge Symmetries

Discussing the Condorcet cycle above, we have directly compared the belief states of different agents. This is problematic for the communication channels in the social network might be disturbed by noise or by other external influences, and they might be deforming by nonlinear transmission peculiarities. Therefore, the next step in our Kant reconstruction employs *lattice gauge theory* as an ultimate resort (Balian et al. 1974; Kogut 1979; Wilson 1974).

Local gauge theory has been invented as a mathematical framework for the treatment of local symmetries in physical systems (Eguchi et al. 1980; Jackson et al. 2001). For illustration we consider Einstein's relativity theory again, which describes symmetries of space and time. An inertial system provides a *global* perspective upon moving bodies and the special principle of relativity claims that all inertial systems are essentially equivalent for the formulation of the general laws of nature (Einstein 1920, p. 90). However, special relativity theory is not suitable to describe accelerated motion and motion in gravitation fields globally. Thus, Einstein saw a need to generalize the special relativity theory toward a general relativity theory, where coordinate systems can be freely chosen locally. Yet this possibility leads to globally curved space-time, subjected to the general principle of relativity, mentioned in Sect. (Einstein 1920, p. 90).

Using mathematical discretization techniques for space and time, a local gauge theory becomes a lattice gauge theory as a suitable model for discrete communication networks connecting a finite number of agents. Some previous ideas to this approach have already been outlined by Mack (1995); Mack (2001) and Graben (2016) (cf. also Graben (2023) for a rather different application).

Considering a social graph $\Gamma = (V, E)$ again, a *lattice gauge society* is firstly given by an assignment of belief state vectors $\Psi_x \in \mathcal{H}_x$ to the agents $x \in V$. Their state spaces are local copies $\mathcal{H}_x \cong \mathcal{H}$ of a unique vector space $\mathcal{H}$, called *fibers*. Moreover, we consider a symmetry group G , such that a representation ρ of G over $\mathcal{H}$ is given through $\rho(G) \subseteq \text{Aut}(\mathcal{H})$. Then, for every group element $g \in G$, $\rho(g)$ is a matrix operator acting on the group representation space $\mathcal{H}$. Secondly, one prescribes an assignment of group elements, $f_{yx} \in G$, to the edges $(x, y) \in E$. If an edge is bidirectional (indicated by double

arrows in the graph (Mack 2001)), then $f_{yx} = \bar{f}_{xy}$, where $\bar{g} \in G$ denotes the inverse of the group element $g \in G$. Note, however, that in general $f_{yx} \neq f_{xy}$ for different edges f_{yx} and f_{xy} .

Figure 2 displays a graph Γ , modeling a dialogue between two agents, $x, y \in V$, with edges $E = \{(x, x), (x, y), (y, x), (y, y)\}$. Each vertex $i \in V$ in Fig. 2(a) is surrounded by a circle, indicating a local fiber space $\mathcal{H}_i$. Correspondingly, we consider a group G with a representation $\rho(G)$ over $\mathcal{H}$ such that to each edge $(i, k) \in E$, a group element $f_{ki} \in G$ is assigned.¹⁰ Here, the self-connections are realized as the group's neutral element: $f_{xx} = f_{yy} = e$.

For our examples above, the representation space is that of preference orderings for maxims $\mathcal{H} = \mathbb{R}^n \supset \mathbb{N}^n$, and the group $G = S_n$, i.e. the symmetric group. Its representation $\rho(G)$ is then the group of permutation matrices. E.g. we may chose the lattice configuration

$$\Psi_x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \Psi_y = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \rho(f_{xy}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \rho(f_{yx}) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

with the additional identities $\rho(f_{xx}) = \rho(f_{yy}) = \mathbf{1}$.

In order to make any statements about the state of a society, we have to define *observables* that are related to projectors in quantum theory and in quantum cognition. Thus, we discuss another set of matrices that map a belief state Ψ_x onto a measurable real number as follows. A projector over $\mathcal{H} = \mathbb{R}^n$ is a matrix $\vec{P} : \mathcal{H} \rightarrow \mathcal{H}$ that is *idempotent*: $\vec{P}^2 = \vec{P}$.

Consider, e.g. the belief state Ψ_y above. The matrix

$$\vec{P}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is a projector, for

$$\vec{P}_1^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \vec{P}_1.$$

Then,

$$\vec{P}_1 \cdot \Psi_y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix},$$

hence $\vec{P}_1$, the *supreme projector*, projects onto the first coordinate of the belief state. The supreme maxim in the preference order of agent y in belief state Ψ_y turns out to be maxim $q = 2$, forbidding cheating. Correspondingly, another projector $\vec{P}_2$, projecting onto the second coordinate of a preference order would be

$$\vec{P}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Accordingly, we call $\vec{P}_k$ the projector onto the k -th coordinate of a preference order, when $1 \leq k \leq n$.

Top obtain a real, observable, measurement outcome, we use the adjoint state

$$\Psi_y^+ = (2 \quad 3 \quad 1)$$

¹⁰Note that we follow the convention for the notation of adjacency matrices in graph theory again.

for calculating

$$\sqrt{\Psi_y^+ \cdot \vec{P}_1 \cdot \Psi_y} = \sqrt{(2 \ 3 \ 1) \cdot \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}} = \sqrt{4} = 2.$$

Thus, $\max q = 2$ has supreme rank in the belief state Ψ_y .

In general, we call the matrix product $\Psi_x^+ \cdot \vec{P} \cdot \Psi_x$ the *expectation value* of the projector $\vec{P}$ in state Ψ_x . The square root of the expectation value of a projector yields the projection of the state onto a subspace.

Two states Ψ_x and Ψ_y at different sites x and y belong to different local vector spaces $\mathcal{H}_x$ and $\mathcal{H}_y$, such that their difference $\Psi_y - \Psi_x$ is not defined. However, they can be compared if they are connected through an edge $p = (x, y) \in E$ by applying a linear map from $\mathcal{H}_x$ to $\mathcal{H}_y$, the so-called *parallel transporter*, that can be motivated through Kant (1785/2002, 4:453),

$$\vec{\Phi}_p \Psi_x = \rho(f_{yx}) \Psi_x, \quad (9)$$

making the difference vector

$$\Delta_p \Psi_y = \Psi_y - \vec{\Phi}_p \Psi_x$$

well-defined afterwards. We call the belief states Ψ_x, Ψ_y *parallel* with respect to a path p , if the difference $\Delta_p \Psi_y = 0$.

Applied to our example, we determine the effect of the parallel transporter

$$\vec{\Phi}_p \Psi_x = \rho(f_{yx}) \Psi_x = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix},$$

such that

$$\Delta_p \Psi_y = \Psi_y - \vec{\Phi}_p \Psi_x = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Thus, the belief states Ψ_x, Ψ_y are obviously not parallel.

Since G is a group with neutral element e and inverse $\bar{g}$ for any $g \in G$, parallel transporters can be defined for any path p across vertices in the graph. The homomorphism $\vec{\Phi}_p \circ \vec{\Phi}_q = \vec{\Phi}_{pq}$, with $pq \in E^2$ as the concatenation of paths p, q defines a *connection* in lattice gauge theory. Note that the parallel transporter assigned to a fundamental loop (x, x) must be $\rho(e) = \mathbf{1}$, the unit matrix. Though, it is not necessarily the case that the parallel transporter $\vec{\Phi}_p = \mathbf{1}$ for any loop p through the graph. If $\vec{\Phi}_p \neq \mathbf{1}$ for a loop $p \in E^*$, the connection is called *frustrated* (Kogut 1979, pp. 683, 685). Here, E^* refers to the *Kleene hull* of the edge set, containing the edge concatenation products of arbitrary lengths. Frustrated connections in a lattice gauge society are essentially reflected by the appearance of Condorcet cycles and by Arrow's "Possibility Theorem of Social Welfare" (Arrow 1950; Bochsler 2010).

Computing the loop parallel transporters of Fig. 2(a), we obtain for the general case

$$\vec{\Phi}_p = \rho(f_{xy} f_{yx}) = \rho(f_{xy}) \cdot \rho(f_{yx})$$

with $p = x \rightarrow y \rightarrow x$. The matrix product $\rho(f_{xy}) \cdot \rho(f_{yx}) = \mathbf{1}$, only if $\rho(f_{xy}) = \rho(f_{yx})^{-1}$, and hence $f_{xy} = \bar{f}_{yx}$, as the inverse of

f_{yx} . But this means that the edges $(x, y), (y, x)$ must be bidirectional, which is apparently not the case in Fig. 2(a).

Inserting the particular examples above, we get

$$\vec{\Phi}_p = \rho(f_{xy}) \cdot \rho(f_{yx}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \neq \mathbf{1},$$

thus, the loop in the dialogue network of Fig. 2(a) is obviously frustrated.

We call an observable, induced by a projector $\vec{P}$, *invariant* if it does not change under parallel transport along different paths, connecting any two agents $x, y \in V$ through different paths $p, q \in E^*$ in the connection (Mack 1995), i.e.

$$(\vec{\Phi}_p \Psi_x)^+ \cdot \vec{P} \cdot \vec{\Phi}_p \Psi_x = (\vec{\Phi}_q \Psi_x)^+ \cdot \vec{P} \cdot \vec{\Phi}_q \Psi_x \quad (10)$$

is independent of the chosen communication channels p, q . In that case, the two agents x, y can partly agree about their respective preference orders upon a common ground. Yet, choosing first a loop p and secondly the elementary loop $q = (x, x)$, both connecting x with herself, entails

$$(\vec{\Phi}_p \Psi_x)^+ \cdot \vec{P} \cdot \vec{\Phi}_p \Psi_x = \Psi_x^+ \cdot \vec{P} \cdot \Psi_x \quad (11)$$

for all agents $x \in V$ and all loops $p \in E^*$. This, however, requires $\vec{\Phi}_p = \mathbf{1}$, meaning that the particular loop p must not be frustrated (Mack 1995).

In our lattice gauge ethics, a given assignment does not uniquely determine the state of the system. Instead, vertices and edges can be locally gauged according to the group transformation as follows (Mack 1995; Mack 2001): To each vertex x we attach a local group element g_x such that the state Ψ_x is transformed as

$$\tilde{\Psi}_x = \rho(g_x) \Psi_x. \quad (12)$$

Moreover, the assignment of the edge (x, y) is transformed according to

$$\tilde{f}_{yx} = g_y f_{yx} \bar{g}_x. \quad (13)$$

This reassignment is called *local gauge transformation* and is visualized in Fig. 2(b). Local gauge transformations correspond to the selection of local frames of reference in general relativity theory, leading to curvature of the underlying space-time.

To extend our example discussed above, we deliberately chose the following local gauge transformation:

$$\rho(g_x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{1}, \rho(g_y) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

such that the belief states transform as

$$\tilde{\Psi}_x = \rho(g_x) \Psi_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

and

$$\tilde{\Psi}_y = \rho(g_y) \Psi_y = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix},$$

while the gauged edges become in the respective matrix representation of the symmetric group S_3

$$\rho(\tilde{f}_{yx}) = \rho(g_y)\rho(f_{yx})\rho(\bar{g}_x) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\rho(\tilde{f}_{xy}) = \rho(g_x)\rho(f_{xy})\rho(\bar{g}_y) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Finally, we compute the gauged loop parallel transporter for our numerical example as

$$\tilde{\Phi}_p = \rho(\tilde{f}_{xy}\tilde{f}_{yx}) = \rho(\tilde{f}_{xy}) \cdot \rho(\tilde{f}_{yx}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Hence the loop is still frustrated after the given gauge transformation.

Importantly, the parallel transporter transforms covariantly under a local gauge transformation. This is proven through

$$\begin{aligned} \tilde{\Phi}_p \tilde{\Psi}_x &= \rho(\tilde{f}_{yx})\tilde{\Psi}_x = \rho(g_y f_{yx} \bar{g}_x) \rho(g_x) \Psi_x = \\ &= \rho(g_y f_{yx} \bar{g}_x g_x) \Psi_x = \rho(g_y) \rho(f_{yx}) \Psi_x = \rho(g_y) \tilde{\Phi}_p \Psi_x \end{aligned} \quad (14)$$

for an edge $p = (x, y)$. Because of the group properties, covariance extends straightforwardly to any path $p \in E^*$. Therefore, the group G is called the *gauge group*, or the gauge symmetry of the model, while the edge assignment is referred to as to the *gauge field*. The gauge symmetry combines the principles of locality and relativity in the framework of gauge theory.

Applied to the case of Fig. 2(b), Eq. (14) entails for the loop from x to

$$\begin{aligned} \tilde{\Phi}_p \tilde{\Psi}_x &= \rho(\tilde{f}_{xy}\tilde{f}_{yx})\tilde{\Psi}_x = \rho(g_x f_{xy} \bar{g}_y g_y f_{yx} \bar{g}_x) \rho(g_x) \Psi_x = \\ &= \rho(g_x) \rho(f_{xy} \bar{g}_y g_y f_{yx} \bar{g}_x g_x) \Psi_x = \rho(g_x) \rho(f_{xy} f_{yx}) \Psi_x = \rho(g_x) \tilde{\Phi}_p \Psi_x \end{aligned}$$

Crucially, this provides now a unique possibility to annul frustration, at least locally. To this end, we simply chose the local gauge

$$\rho(g_x) = \tilde{\Phi}_p^{-1}, \quad (16)$$

yielding $\tilde{\Phi}_p = 1$. This finding applies straightforwardly to any loop p from x back to x , for all agents x in the strongly connected network.

Thus, for our numerical example we set

$$\rho(g_x) = \tilde{\Phi}_p^{-1} = \tilde{\Phi}_p = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which remedies the frustration of the dialogue network Fig. 2.

Now, we are prepared for the final steps of our system-theoretic reconstruction of Kantian ethics. First, we identify the symmetries of our lattice gauge ethics with Kant's maxims of the common human understanding from Sect. . Maxim (1), to think for oneself, corresponds to the gauge freedom of an individual agent x to chose a local permutation $g_x \in G$ for deliberately changing her preference order. Maxim (2),

to think in the position of everyone else, reflects the local gauge symmetry of the communication network, and thereby the covariance law (14), expressing the relativity principle that all individual perspectives are fundamentally equitable, and no perspective is preferred against another one a priori. Then, maxim (3), always to think consistently in accord with oneself, requires that frustration can always be locally remedied by gauge transformations along loops connecting arbitrary groups of agents within the social network, that could be called *teams*. Consequently, there must be invariant observables, shared by all agents of a team in the lattice gauge society, according to Eq. (11). Note that the consistency maxim (3) can only be observed in network topologies containing loops. This is not the case for the network caricatures depicted in Fig. 1(a – c); only the open society Fig. 1(d) exhibits the necessary condition to comply with maxim (3).

Finally, we make the system self-referential again. By incorporating the rationality maxim (4), to which we assign some number $\aleph$ (with $\aleph \in \mathbb{N}$ in (8) above), into each agent's subjective preference ordering as the invariantly supreme maxim, all agents must, by means of this gödelization, agree upon a rational discourse of public affairs in an open society. The supreme rationality principle (4) and its restrictions, the maxims of common human understanding (1 – 3), and eventually the four different formulas of the categorical imperative, FUL, FEI, FA, and FRE, as its corollaries, are necessarily constitutive for a lattice gauge society. Only if every agent in the social network ranks maxim (4) above all other maxims and their respective ends, i.e., only if the supreme projector $\tilde{P}_1$ is invariant under local gauge transformations,

$$\sqrt{\Psi_x^+ \cdot \tilde{P}_1 \cdot \Psi_x} = \aleph, \quad (17)$$

for all agents $x \in V$, there exists a lattice gauge society that could be maintained above all. It must therefore necessarily be an open society with irreducible adjacency matrix in order to guarantee the existence of loops connecting teams as a condition of the possibility of the consistency maxim (3). This is vindicated by Kant's view: "The [moral] law rather imposes itself on him [i.e. man] irresistibly, because of his moral predisposition; and if no other incentive were at work against it, he would also incorporate it into his supreme maxim as sufficient determination of his power of choice, i.e. he would be morally good" (Kant 1793/1996, 6:36).

Summarizing our findings on interactional ethics, we obtain Tab. 1 of the rational maxims for a lattice gauge ethics. Therefore, we have reason to identify the lattice gauge society with Kant's realm of ends, with the culture of sincerity, with the church invisible, and ultimately with a healthy democratic open society. Good will in such a society of interacting agents is that will that invariantly ranks the moral law at supreme position of all the possible preference orders for tentative maxims (as in state Ψ_z in Eq. (8) above).

For concluding this section, we have eventually to address the issue of the "radical evil" (Kant 1793/1996, 6:38). Putting it as succinctly as possible, this theory says, that there is a "propensity" (Kant 1793/1996, 6:29, 6:37) in human nature to voluntarily neglecting the voice or reason not giving supreme priority to the rational moral law: "It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims" (Kant 1793/1996, 6:36). Or, in the words of the *Threepenny Opera*: "Your fine philosophy, good sirs, you may proclaim. / But till you feed us, right and wrong can wait!" (Brecht et al. 1964, p. 2.3.)¹¹ Hence, we could

¹¹Whereas the German original is even much more succinct and drastic: "Erst kommt

reconstruct this idea within the framework of lattice gauge ethics into a single formula, the *formula of radical evil* (*sic!*).

$$\sqrt{\Psi_x^+ \cdot \vec{P}_1 \cdot \Psi_x} \neq \aleph. \quad (18)$$

Note that Arendt (1976, p. 335) has criticized that the “effect of the work was exactly the opposite of what Brecht had sought by it. The bourgeoisie could no longer be shocked; it welcomed the exposure of its hidden philosophy, whose popularity proved they had been right all along, so that the only political result of Brecht’s ‘revolution’ was to encourage everyone to discard the uncomfortable mask of hypocrisy and to accept openly the standards of the mob.”

Discussion

“According to Kant, no man has the right to obey” (Arendt 2007) — Arendt’s acutely pointed formulation captures a central implication of our reconstruction of Kant’s interactional ethics: A rational agent cannot transfer responsibility for judgment to an external authority, but must regard the maxims of action as principles that she herself autonomously adopts and subjects to her own reflection. In the previous sections, we have developed a reconstruction of this structure in terms of system theory and lattice gauge ethics. Kant’s three “maxims of common human understanding” (Kant 1790/2000a, 5:294) were presented as local gauge freedom of thought, symmetry transformations between individual perspectives, and invariant consistency under such transformations. The rationality maxim was then placed at the supreme position of each agent’s preference ordering as a constitutive condition for the maintenance of a lattice gauge society as a healthy democratic open society (Minkler 1989; Nyswander 1967; Popper 2013; Etzioni 2000; Taswell 2025b; Zhu et al. 2025).

To further examine this reconstruction, we have to turn to a more difficult question: What happens when an agent retains the capacity to reason and nevertheless chooses evil? Kant’s notion of “radical evil” (Kant 1793/1996, 6:38) cannot simply be identified with irrationality. Irrationality concerns a failure of judgment itself, as the refusal of consistency, the rejection to revise manifestly contradictory judgments, or the denial to subject one’s own claims to rational discourse. An evil agent need not lose the capacity to reason about justifications, to reckon consequences of tentative deeds, or to coherently organize action plans. Rather, Kant describes the radical evil in terms of a reversal in the ordering of incentives: The agent still recognizes the moral law, but makes conformity to it conditional upon self-love, happiness, or private advantage. Moral requirements may therefore be followed when they coincide with self-interest, while self-love takes priority once the two come into conflict (Kant 1793/1996, 6:36). In this restricted sense, we may speak of *rational evil*. The term is not intended to suggest that evil receives rational justification. An agent may remain capable of calculation, planning, and strategic consistency while ordering her incentives in the wrong way. What is lost is not the capacity of rational judgement but rather the agent’s ability to reason effectively in pursuit of ends that place private incentives above the moral incentive. An agent may endorse both a maxim against deception and a maxim favouring private advantage. The decisive issue is, which of the two takes precedence in the agent’s subjective preference ordering in case of conflict.

Mariña’s social interpretation of Kant’s *Religion* helps placing this ordering within the relations in which moral agency actually develops

das Fressen, dann kommt die Moral” (Brecht et al. 2001).

(Mariña 2022). Human beings do not form their ideas of happiness, status, or self-worth in isolation, but through their dealings with others (Kant 1793/1996, 6:96). Kant’s discussion of comparative self-love is especially revealing in this respect. The wish not to stand beneath others may, through repeated comparison, turn into anxiety about relative standing and eventually into a desire for superiority, providing fertile ground for envy, rivalry, and hostility (Kant 1793/1996, 6:27, 6:93f). None of this dissolves individual responsibility. Social conditions do not themselves produce radical evil, since agents remain responsible for the maxims they adopt. They can, however, make particular incentives easier to reinforce and stabilize. Competition, insecurity, and persistent status comparison may make self-love increasingly dominant within an agent’s ordering of motives. Society is therefore not the source of evil, but it can amplify and stabilize the propensity to it.

This is where an interactional account becomes necessary. The preference ordering of each agent represents the way in which she ranks her maxims, whereas communication, perspective transformations, and coherency loops describe how judgments are exposed to the standpoints of others. Interaction may intensify comparison and rivalry, but it can also create conditions under a reconsidered and revised judgment. Frustration, then, should not be equated with evil itself. A social network may contain unresolved inconsistencies without its members being morally corrupt, just as a stable social order may remain deeply unjust. Mariña’s account of the ethical community becomes especially important at this point (Mariña 2022). Kant’s ethical community does not overcome radical evil by imposing the correct inner disposition; that would itself violate autonomy (Kant 1793/1996, 6:95). What it can change are the relations within which agents encounter one another. A community shaped by trust, sincerity, reciprocity, and mutual recognition of their dignity need not cast others primarily as rivals, threats, or objects of manipulation or domination. It cannot choose morally on behalf of its members, but it can create conditions less governed by competitive self-love.

This is also why Kant’s “church invisible” is at once inward and irreducibly social (Kant 1793/1996, 6:101). Its invisibility reflects the fact that moral disposition cannot be imposed or fully known in a third-person perspective from outside (Taswell 2025a; Graben 2025). Yet it cannot be realized in isolation, since trust, sincerity, reciprocity, and mutual recognition exist only among agents. Kant therefore offers neither obedience to an organization nor withdrawal into private conscience. The ethical community instead requires individuals to judge for themselves while relating to others as equally autonomous agents. From this perspective, our self-referential reconstruction of the realm of ends, suggested in the present study, takes on a more concrete meaning. A realm of ends cannot be reduced to a shared set of rules; it must be an order in which rational agency itself is sustained. When social relations repeatedly encourage agents to treat others as competitors or merely as tools to their own ends, self-love can take priority over moral considerations. Under such conditions, the reciprocal recognition for genuine autonomy begins to erode. An ethical community must therefore be reproduced through the conduct of its members by sustaining the conditions under which autonomous moral choice remains possible.

A healthy democratic open society is not one in which disagreement disappears or where all agents converge on the same preference ordering. What must be preserved is the possibility of autonomous judgment, the accessibility of relatively different perspectives to one another, the good will to avoid performative inconsistency, and eventually the capacity of agents to reconsider the ordering of their individual maxims

through social interaction. Self-maintenance therefore means more than preserving existing institutions. It means preserving the conditions under which individuals can continue to act as rational and autonomous agents. The present model also has a clear limitation. Our lattice gauge ethics assumes that the rationality maxim can retain an invariantly supreme position in agents' preference orderings. Radical evil, however, shows that agents may preserve substantial capacities for reasoning while failing to give practical reason priority. Social structure cannot replace individual moral responsibility, but neither does moral agency operate outside social relations.

In this sense, the moral of the *Threepenny Opera*, mentioned above, rather becomes a commonplace: The healthy democratic open society, preventing poverty and paternalism, is both, a necessary and sufficient condition for its own maintenance. Thus, we conclude that only good will, the worthiness to be happy, turns out to be a prerequisite for the pursuit of happiness.

Citation

Brainiacs Journal 2026 Volume 7 Issue 2 Edoc JD986920E.
Title: "A System-Theoretical Account to Kant's Interactional Ethics".
Authors: Peter beim Graben, Kiersten Xiangqi Zhu.
Dates: created 2025-10-04, received 2026-09-29, updated 2026-09-29, published 2026-10-01
Copyright: © 2026 Brain Health Alliance
Contact: peter.beimgraben@b-tu.de
NPDS: [LINKS/Brainiacs/BeimGraben2026STAKIE](#)
DOI: [10.48085/JD986920E](#)

Acknowledgments

We thank Serafim Rodrigues for valuable discussions. This study was partly supported by the EU Interreg Brandenburg-Polska project "ILB – IaaS – Innovation as a Service".

Affiliations

Peter beim Graben, peter.beimgraben@b-tu.de; Brandenburgische Technische Universität (BTU) Cottbus–Senftenberg, Communications Engineering, Cottbus, Germany, and Bernstein Center for Computational Neuroscience, Berlin, Germany.

Kiersten Xiangqi Zhu, k.zhu26@imperial.ac.uk; Imperial College, Department of Bioengineering, London, United Kingdom.

References

- [1] H. Arendt. *Eichmann in Jerusalem. A Report on the Banality of Evil*. New York (NY): Viking Press, 1963 (cited p. 7).
- [2] H. Arendt. "Thinking and moral considerations: A lecture." *Social Research* 38.3 (1971), pp. 417–446. URL: <http://www.jstor.org/stable/40970069> (cited p. 4).
- [3] H. Arendt. *The Origins of Totalitarianism*. New Edition with Added Prefaces. New York: Harcourt Brace & Company, 1976. URL: https://www.almendron.com/tribuna/wp-content/uploads/2015/06/Arendt_Hannah_The_Origins_of_Totalitarianism_1979.pdf (cited pp. 1, 2, 4, 5, 9, 10, 14).
- [4] H. Arendt. *Between Past and Future. Eight Exercises in Political Thought*. New York: Penguin, 1977 (cited pp. 2, 4, 10).
- [5] H. Arendt. "Hannah Arendt im Gespräch mit Joachim Fest. Eine Rundfunksendung aus dem Jahr 1964." *HannahArendt.net* 3.1 (2007). DOI: [10.57773/hanet.v3i1.114](https://doi.org/10.57773/hanet.v3i1.114) (cited pp. 7, 14).
- [6] K. J. Arrow. "A difficulty in the concept of social welfare." *Journal of Political Economy* 58.4 (1950), pp. 328–346. URL: <http://www.jstor.org/stable/1828886> (cited pp. 9, 11, 12).
- [7] C. Bagnoli. "The practical significance of the categorical imperative." In: *Oxford Studies in Agency and Responsibility*. Vol. 11. 2021, pp. 178–199 (cited p. 5).
- [8] C. Bagnoli. "Constructivism in metaethics." In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta and U. Nodelman. Metaphysics Research Lab, Stanford University, 2026. URL: <https://plato.stanford.edu/archives/sum2026/entries/constructivism-metaethics/> (cited p. 5).
- [9] R. Balian, J. M. Drouffe, and C. Itzykson. "Gauge fields on a lattice. I. General outlook." *Physical Reviews D* 10 (10 1974), pp. 3376–3395. DOI: [10.1103/PhysRevD.10.3376](https://doi.org/10.1103/PhysRevD.10.3376) (cited p. 11).
- [10] R. Blutner and P. beim Graben. "Quantum cognition and bounded rationality." *Synthese* 193 (2016), pp. 3239–3291. DOI: [10.1007/s11229-015-0928-5](https://doi.org/10.1007/s11229-015-0928-5) (cited p. 9).
- [11] D. Bochsler. "The Marquis de Condorcet goes to Bern." *Public Choice* 144.1 (2010), pp. 119–131. DOI: [10.1007/s11127-009-9507-y](https://doi.org/10.1007/s11127-009-9507-y) (cited pp. 11, 12).
- [12] A. Bordum. "Immanuel Kant, Jürgen Habermas and the categorical imperative." *Philosophy & Social Criticism* 31.7 (2005), pp. 851–874. DOI: [10.1177/0191453705057307](https://doi.org/10.1177/0191453705057307) (cited p. 4).
- [13] B. Brecht and K. Weill. *The Threepenny Opera*. Trans. by D. Vesey and E. Bentley. Foreword by Lotte Lenya. New York: Grove Press, 1964 (cited p. 13).
- [14] B. Brecht and K. Weill. *Die Dreigroschenoper*. Frankfurt am Main: Suhrkamp, 2001 (cited p. 14).
- [15] S. Brott. *Observations on the architecture of evil: A new reading of Eichmann on Kant*. University of Queensland. 2016. URL: <https://epri.uts.qut.edu.au/92553/> (cited p. 7).
- [16] J. R. Busemeyer and P. D. Bruza, eds. *Quantum Models of Cognition and Decision*. Cambridge University Press, 2012 (cited p. 9).
- [17] J. R. Busemeyer and J. T. Townsend. "Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment." *Psychological Review* 100.3 (1993), pp. 432–459. DOI: [10.1037/0033-295X.100.3.432](https://doi.org/10.1037/0033-295X.100.3.432) (cited p. 9).
- [18] I. F. R. Caldas, I. de Moraes Paim, K. T. F. Leite, et al. "Quantum cognition and the contextual dynamics of moral judgement." *Royal Society Open Science* 13.2 (2026), p. 251534. DOI: [10.1098/rsos.251534](https://doi.org/10.1098/rsos.251534) (cited p. 9).
- [19] E. Cassirer. "Einstein's theory of relativity considered from the epistemological standpoint." *The Monist* 32.1 (1922), 89–134a (cited p. 3).
- [20] C. Castellano, S. Fortunato, and V. Loreto. "Statistical physics of social dynamics." *Reviews of Modern Physics* 81 (2 2009), pp. 591–646. DOI: [10.1103/RevModPhys.81.591](https://doi.org/10.1103/RevModPhys.81.591) (cited p. 9).
- [21] J. Ebbinghaus. "Interpretation and misinterpretation of the categorical imperative." *The Philosophical Quarterly* 4.15 (1954), pp. 97–108. URL: <http://www.jstor.org/stable/2216381> (cited p. 3).
- [22] T. Eguchi, P. B. Gilkey, and A. J. Hanson. "Gravitation, gauge theories and differential geometry." *Physics Reports* 66.6 (1980), pp. 213–393. DOI: [10.1016/0370-1573\(80\)90130-1](https://doi.org/10.1016/0370-1573(80)90130-1) (cited p. 11).
- [23] A. Einstein. *Relativity: The Special and the General Theory*. London: Methuen, 1920. URL: <https://www.gutenberg.org/ebooks/36114> (cited pp. 3–5, 11).

- [24] A. Etzioni. "Creating good communities and good societies." *Contemporary Sociology* 29.1 (2000), pp. 188–195. URL: <https://ssrn.com/abstract=1438224> (cited pp. 1, 6, 14).
- [25] A. Farr. "Can a philosophy of race afford to abandon the Kantian categorical imperative?" *Journal of Social Philosophy* 33.1 (2002) (cited p. 3).
- [26] H. von Foerster. *Understanding Understanding: Essays on Cybernetics and Cognition*. New York: Springer, 2003 (cited p. 5).
- [27] M. Gabriel. *Der Sinn des Denkens*. Berlin: Ullstein, 2018 (cited p. 3).
- [28] I. Geiger. "What is the use of the universal law formula of the categorical imperative?" *British Journal for the History of Philosophy* 18.2 (2010), pp. 271–295. DOI: [10.1080/09608781003643568](https://doi.org/10.1080/09608781003643568) (cited p. 3).
- [29] I. Geiger. "How are the different formulas of the categorical imperative related?" *Kantian Review* 20.3 (2015), pp. 395–419. DOI: [10.1017/S1369415415000151](https://doi.org/10.1017/S1369415415000151) (cited pp. 4–7).
- [30] L. Gil. "What are the best hierarchical organizations for the success of a common endeavour?" *Mathematical Anthropology and Cultural Theory* 9.1 (2016). URL: <http://mathematicalanthropology.org/Pdf/MACTGi10416.pdf> (cited p. 10).
- [31] M. Giovannelli. "Relativity theory as a theory of principles: A reading of Cassirer's *Zur Einsteinschen Relativitätstheorie*." *Journal of the International Society for the History of Philosophy of Science* 13.2 (2023), pp. 261–296. DOI: [10.1086/726076](https://doi.org/10.1086/726076) (cited p. 3).
- [32] K. Gödel. "Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I." *Monatshefte für Mathematik und Physik* 38 (1931), pp. 173–198 (cited pp. 5, 6, 8).
- [33] P. beim Graben. "Naphtas Visionen. Perspektivität in der Naturwissenschaft." In: *Post-Physikalismus*. Ed. by M. Knap, T. Müller, and P. Spät. Freiburg / Br.: Alber, 2011, pp. 122–141 (cited pp. 3, 4).
- [34] P. beim Graben. "Comment on Gil's: What are the best hierarchical organizations for the success of a common endeavour." *Mathematical Anthropology and Cultural Theory* 9.2 (2016). URL: <http://mathematicalanthropology.org/Pdf/MACTbeimGraben0616.pdf> (cited pp. 4, 11).
- [35] P. beim Graben. "Gauge symmetries of musical and visual forces." *Journal of Mathematics and the Arts* 17.3–4 (2023), pp. 347–382. DOI: [10.1080/17513472.2023.2281895](https://doi.org/10.1080/17513472.2023.2281895) (cited p. 11).
- [36] P. beim Graben. "A neural network account of Kant's philosophical aesthetics." *Mind and Matter* 22.2 (2024), pp. 227–252. DOI: [10.5376/mm2024.227](https://doi.org/10.5376/mm2024.227) (cited p. 3).
- [37] P. beim Graben. "Commentary on "From the Absolute of Either-Or To the Uncertainty of And"" *Brainiacs Journal* 6.2 (2025), SE696E956. DOI: [10.48085/SE696E956](https://doi.org/10.48085/SE696E956) (cited p. 14).
- [38] P. Guyer. "The possibility of the categorical imperative." *Philosophical Review* 104.3 (1995), pp. 353–385. URL: <http://www.jstor.org/stable/2185633> (cited pp. 5–7, 9).
- [39] P. Guyer. "The derivation of the categorical imperative: Kant's correction for a fatal flaw." *Harvard Review of Philosophy* 10.1 (2002) (cited p. 5).
- [40] J. Habermas. *Moral Consciousness and Communicative Action*. Cambridge (MA): MIT Press, 1990. URL: <https://archive.org/details/moralconsciousne0000habe> (cited p. 4).
- [41] B.-C. Han. *Psychopolitics: Neoliberalism and New Technologies of Power*. London: Verso, 2017 (cited p. 2).
- [42] R. Hanna. *Defending Kant from his friends and his enemies*. againstprofphil.org/2024/05/26/. 2024. URL: <https://againstprofphil.org/2024/05/26/defending-kant-from-his-friends-and-his-enemies/> (cited p. 7).
- [43] J. Harrison. "Kant's examples of the first formulation of the categorical imperative." *The Philosophical Quarterly* 7.26 (1957), pp. 50–62. URL: <http://www.jstor.org/stable/2216346> (cited p. 3).
- [44] V. Havel. *Letters to Olga June 1979 – September 1982*. New York: Henry Holt and Company, 1989 (cited p. 5).
- [45] V. Havel. *Living in Truth. Twenty-two essays published on the occasion of the award of the Erasmus Prize to Václav Havel*. Ed. by J. Vladislav. London: Faber & Faber, 1990 (cited p. 1).
- [46] J. Heis. "Realism, functions, and the a priori: Ernst Cassirer's philosophy of science." *Studies in History and Philosophy of Science Part A* 48 (2014), pp. 10–19. DOI: [10.1016/j.shpsa.2014.08.003](https://doi.org/10.1016/j.shpsa.2014.08.003) (cited pp. 3, 4).
- [47] T. E. Hill. "Humanity as an end in itself." *Ethics* 91.1 (1980), pp. 84–99. URL: <http://www.jstor.org/stable/2380373> (cited p. 6).
- [48] D. R. Hofstadter. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York (NY): Basic Books, 1979 (cited pp. 5, 6, 8).
- [49] J. D. Jackson and L. B. Okun. "Historical roots of gauge invariance." *Reviews of Modern Physics* 73.3 (2001), pp. 663–680. DOI: [10.1103/RevModPhys.73.663](https://doi.org/10.1103/RevModPhys.73.663) (cited p. 11).
- [50] I. Kant. *Opus Postumum*. The Cambridge Edition of The Works Of Immanuel Kant. Cambridge: Cambridge University Press, 1999 (cited p. 2).
- [51] I. Kant. *Critique of Practical Reason*. Ed. by M. Gregor. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 1788/2015. URL: aghalibrary.com (cited pp. 3–6).
- [52] I. Kant. *Critique of Pure Reason*. Ed. by P. Guyer and A. W. Wood. 2nd. The Cambridge Edition of The Works Of Immanuel Kant. Cambridge: Cambridge University Press, 1787/1998 (cited pp. 3, 4, 7).
- [53] I. Kant. *Critique of the Power of Judgment*. Ed. by P. Guyer. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1790/2000 (cited pp. 2–7, 9, 14, 18).
- [54] I. Kant. *First Introduction to the Critique of the Power of Judgment*. Ed. by P. Guyer. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1790/2000 (cited p. 3).
- [55] I. Kant. *Groundwork for the Metaphysics of Morals*. Ed. by A. W. Wood. Yale University Press, 1785/2002 (cited pp. 2, 3, 5–8, 12).
- [56] I. Kant. *Prolegomena to Any Future Metaphysics*. The Cambridge Edition of The Works Of Immanuel Kant. Cambridge: Cambridge University Press, 1783/2004 (cited p. 3).
- [57] I. Kant. *Religion Within the Boundary of Mere Reason*. Ed. by A. W. Wood and G. di Giovanni. The Cambridge Edition of the Works of Immanuel Kant. Translated by George di Giovanni. Cambridge: Cambridge University Press, 1793/1996, pp. 39–215 (cited pp. 2, 5, 7–9, 11, 13, 14).
- [58] I. Kant. *The Metaphysics of Morals*. Ed. by M. Gregor. Texts in German Philosophy. Cambridge (MA): Cambridge University Press, 1797/1991 (cited pp. 3, 5, 6, 8).
- [59] I. Kant. "What does it mean to orient oneself in thinking?" In: *Religion and Rational Theology*. Ed. by A. W. Wood and G. di Giovanni. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1786/1996 (cited pp. 5, 6).
- [60] A. Y. Khrennikov. "Quantum-like model of cognitive decision making and information processing." *Biosystems* 95.3 (2009), pp. 179–187. DOI: [10.1016/j.biosystems.2008.10.004](https://doi.org/10.1016/j.biosystems.2008.10.004). URL: <http://www.sciencedirect.com/science/article/pii/S0303264708002062> (cited p. 9).
- [61] J. B. Kogut. "An introduction to lattice gauge theory and spin systems." *Reviews of Modern Physics* 51 (4 1979), pp. 659–713. DOI: [10.1103/RevModPhys.51.659](https://doi.org/10.1103/RevModPhys.51.659) (cited pp. 11, 12).

- [62] C. M. Korsgaard. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996 (cited pp. 3, 5–7).
- [63] J. Llabrés, R. Toral, M. S. Miguel, and F. Vázquez. “Reducibility of higher-order to pairwise interactions: Social impact models on hypergraphs.” *Physical Review Research* 8.2 (2026). DOI: [10.1103/PhysRevResearch.8.023113](https://doi.org/10.1103/PhysRevResearch.8.023113) (cited p. 9).
- [64] G. Mack. “Gauge theory of things alive.” *Nuclear Physics B* 421.3 (1995), pp. 923–925. DOI: [10.1016/0550-3213\(95\)00423-7](https://doi.org/10.1016/0550-3213(95)00423-7). URL: <http://www.sciencedirect.com/science/article/B6TVD-41B7B2K-8N/2/da77de27d3ebefec027bc8b6981b551c> (cited pp. 9, 11, 12).
- [65] G. Mack. “Universal Dynamics, a Unified Theory of Complex Systems. Emergence, Life and Death.” *Communications of Mathematical Physics* 219.1 (2001), pp. 141–178. DOI: [10.1007/s002200100397](https://doi.org/10.1007/s002200100397) (cited pp. 11, 12).
- [66] J. Mariña. “Radical evil, social contracts and the idea of the church in Kant.” *Kantian Review* 27.1 (2022), pp. 71–79. DOI: [10.1017/S1369415421000704](https://doi.org/10.1017/S1369415421000704) (cited pp. 2, 8, 14).
- [67] H. Maturana and F. Varela. *The Tree of Knowledge*. Boston: Shambhala Press, 1998 (cited p. 6).
- [68] A. Melnick. “Kants formulations of the categorical imperative.” *Kant-Studien* 93.3 (2002), pp. 291–308. DOI: [doi:10.1515/kant.2002.010](https://doi.org/10.1515/kant.2002.010) (cited p. 3).
- [69] M. Minkler. “Health education, health promotion and the open society: An historical perspective.” *Health Education Quarterly* 16.1 (1989), pp. 17–30. DOI: [10.1177/109019818901600105](https://doi.org/10.1177/109019818901600105) (cited pp. 1, 8, 14).
- [70] E. Mizraji. “The strange story of an almost unknown prime number counter: the Rafael Barrett formula.” *The Mathematical Gazette* 0.0 (2026), pp. 1–7. DOI: [10.1080/00255572.2026.2664683](https://doi.org/10.1080/00255572.2026.2664683) (cited p. 9).
- [71] W. Nelson. “Kant’s formula of humanity.” *Mind* 117.465 (2008), pp. 85–106. DOI: [10.1093/mind/fzn004](https://doi.org/10.1093/mind/fzn004) (cited p. 7).
- [72] D. B. Nyswander. “The open society: Its implications for health educators.” *Health Education Monographs* 22 (1967), pp. 3–15 (cited pp. 1, 8, 14).
- [73] O. O’Neill. *Constructions of Reason: Explorations of Kant’s Practical Philosophy*. Cambridge University Press, 1989 (cited pp. 2–6).
- [74] P. Pangaro. “Invitation to recursioning: Heinz von Foerster and cybernetic praxis.” *Cybernetics and Human Knowing* 18.3 (2011), p. 139 (cited p. 5).
- [75] K. R. Popper. *The Open Society and Its Enemies: New One-Volume Edition*. Ed. by A. Ryan and E. H. Gombrich. Princeton (NJ): Princeton University Press, 2013 (cited pp. 1, 4, 6, 8, 10, 14).
- [76] E. M. Pothos and J. R. Busemeyer. “Can quantum probability provide a new direction for cognitive modeling?” *Behavioral and Brain Sciences* 36.03 (2013), pp. 255–274. DOI: [10.1017/S0140525X12001525](https://doi.org/10.1017/S0140525X12001525) (cited p. 9).
- [77] R. E. Purser. *McMindfulness: How Mindfulness Became the New Capitalist Spirituality*. London: Repeater Books, 2019 (cited pp. 1, 5).
- [78] B. E. Rollin. “There is only one categorical imperative.” *Kant-Studien* 67.1–4 (1976), pp. 60–72. DOI: [doi:10.1515/kant.1976.67.1-4.60](https://doi.org/10.1515/kant.1976.67.1-4.60) (cited p. 5).
- [79] M. M. Rossi. “Book Review: *The Categorical Imperative: A Study in Kant’s Moral Philosophy* by H.J. Paton.” *Journal of Philosophy* 46.10 (1949), pp. 293–303. URL: <http://www.jstor.org/stable/2019922> (cited p. 3).
- [80] B. Stangneth. *Kultur der Aufrichtigkeit. Zum systematischen Ort von Kants Religion innerhalb der Grenzen der blossen Vernunft*. Würzburg: Königshausen & Neumann, 2000 (cited p. 7).
- [81] P. Steinberger. “The standard view of the categorical imperative.” *Kant-Studien* 90.1 (1999), p. 91. URL: <https://www.proquest.com/scholarly-journals/standard-view-categorical-imperative/docview/1294192554/se-2?accountid=17185> (cited p. 3).
- [82] P. Stratton-Lake. “Formulating categorical imperatives.” *Kant-Studien* 84.3 (1993), pp. 317–340. DOI: [doi:10.1515/kant.1993.84.3.317](https://doi.org/10.1515/kant.1993.84.3.317) (cited pp. 3, 5–7).
- [83] C. Taswell. “From the Absolute of Either-Or to the Uncertainty of And.” *Brainiacs Journal* 6.2 (2025), DE9EB766D. DOI: [10.48085/DE9EB766D](https://doi.org/10.48085/DE9EB766D) (cited p. 14).
- [84] C. Taswell. “Who defends democracy and decency? Let’s maintain standards for maintaining standards.” *Brainiacs Journal* 6.3 (2025), D4MF33921. DOI: [10.48085/D4MF33921](https://doi.org/10.48085/D4MF33921) (cited pp. 1, 2, 8, 14).
- [85] A. Tversky and D. Kahneman. “The framing of decisions and the psychology of choice.” *Science* 211.4481 (1981), pp. 453–458. DOI: [10.1126/science.7455683](https://doi.org/10.1126/science.7455683). URL: <http://www.sciencemag.org/content/211/4481/453.abstract> (cited p. 9).
- [86] K. G. Wilson. “Confinement of quarks.” *Physical Reviews D* 10.8 (1974), pp. 2445–2459. DOI: [10.1103/PhysRevD.10.2445](https://doi.org/10.1103/PhysRevD.10.2445) (cited p. 11).
- [87] G. Yaffe. “Freedom, natural necessity and the categorical imperative.” *Kant-Studien* 86.4 (1995), p. 446 (cited p. 3).
- [88] K. X. Zhu and P. beim Graben. “Public health and safe communities require open societies.” *Brainiacs Journal* 6.3 (2025), Z6E2BB73D. DOI: [10.48085/Z6E2BB73D](https://doi.org/10.48085/Z6E2BB73D) (cited pp. 1, 8, 9, 14, 18).

Table 1: Isomorphism between Kant’s maxims of common human understanding (Kant 1790/2000a, 5:294) and lattice gauge ethics.

Maxim	Kant	Lattice Gauge Ethics
1	“To think for oneself”	Local gauge freedom
2	“To think in the position of everyone else”	Local gauge symmetry
3	“Always to think in accord with oneself”	Local gauge invariance
4	Be rational! Be real!	Maintenance of lattice gauge society

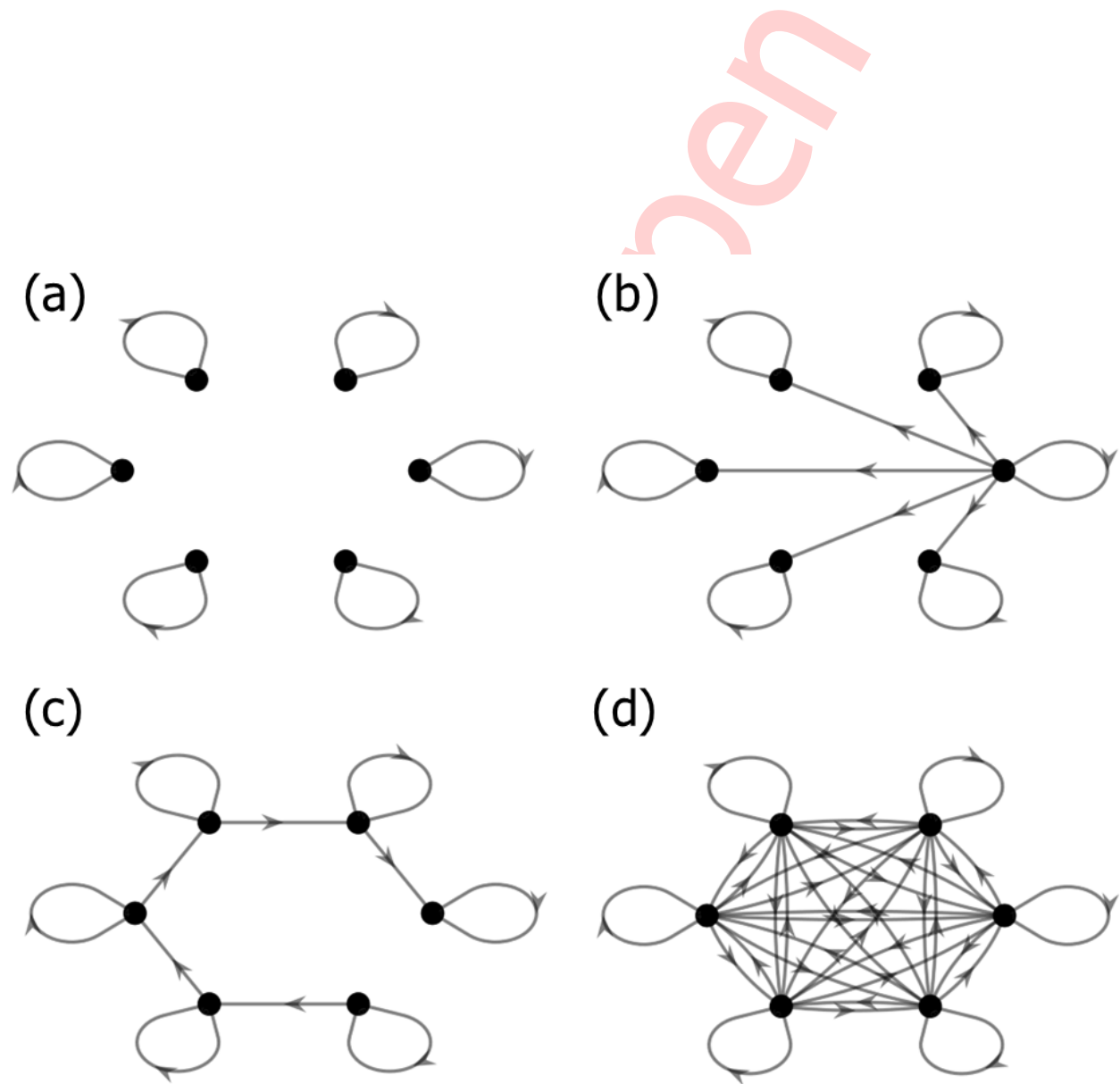
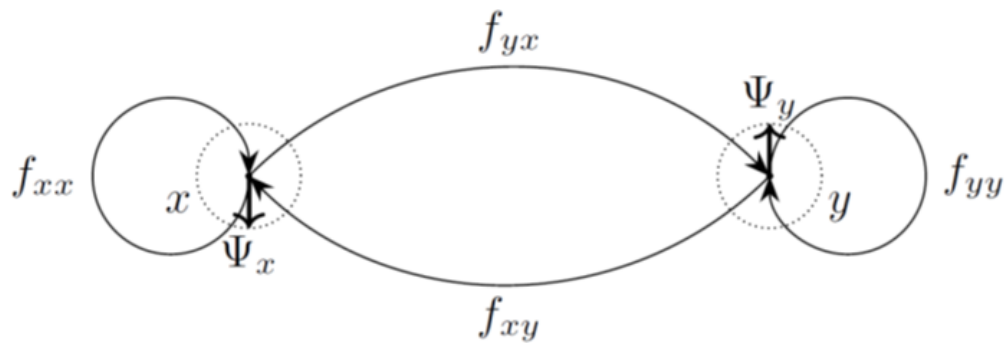


Figure 1: Schematic social network topologies. (a) A *segregated* society of isolated individuals. (b) A *totalitarian* society, dominated by a dictator. (c) A *military* command-obedience chain. (d) An *open* society. Number of individuals $N = 6$ for illustration (cf. Zhu et al. (2025, Fig. 2)).

(a)



(b)

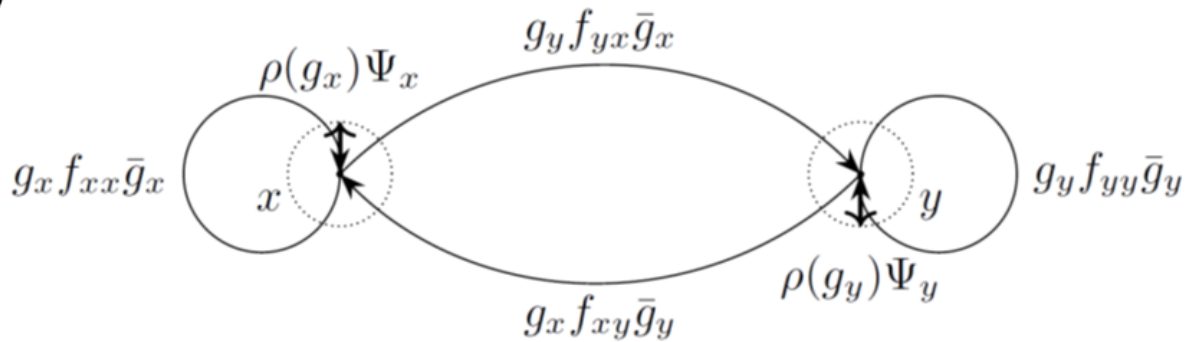


Figure 2: Lattice gauge model on a binary lattice Γ . (a) To vertices x, y of Γ vectors of its associated fiber spaces $\mathcal{H}_x, \mathcal{H}_y$ are attached; correspondingly, an element of a gauge group G is assigned to each edge of Γ . (b) The configuration after local gauge transformation: Vertex states are transformed through a representation of the gauge group acting on fiber spaces; edges are locally gauged.