



# ETHICS AND ECONOMICS

Lecture Notes for Week 11

Tomáš Cahlík

# Contents

- Democratic Justice ..... 2
  - Democratic Justice Theory ..... 2
  - Democratic Justice Applications..... 4
  - Discourse Theory of Democracy ..... 6
  - Summary ..... 10
- Appendix 1: Information and Communication Technologies and Artificial Intelligence..... 13
- Appendix 2: More about the video and the reading promoted in the 9th and 10th chapters ..... 24

# Democratic Justice

In the first part of this chapter, we discuss the theory of democratic justice. In the second part, we present two applications: democratic justice in governing children and democratic justice in firms. In the third part, we briefly introduce the discourse theory of democracy.

Appendix 1 provides points for discussing ICT and AI ethics, while Appendix 2 offers additional information about the videos and readings promoted in Chapters 9 and 10.

## Democratic Justice Theory

Democracy, understood in the Schumpeterian sense, fulfills quite well the two fundamental Enlightenment aims: individual freedom and the scientific principles of truth.

Regarding individual freedom, the American political scientist Robert Dahl (1915–2014) analyzed judicial reviews in the United States and concluded that the requirement of a strict separation of powers is not necessary. This suggests that having political parties compete with their programs for voters is sufficient to ensure individual freedom.

Concerning the scientific principles of truth, discussions among political parties about societal problems can realize the competitive argumentation ideal proposed by John Stuart Mill for converging toward truth. However, achieving this ideal largely depends on the cultivation and integrity of politicians.

Winston Churchill famously remarked, "Democracy is the worst form of government, except for all those other forms that have been tried from time to time." This statement, made in a speech to the House of Commons in 1947, encapsulates Churchill's view that while democracy has flaws and inefficiencies, it remains superior to other forms of government, which have historically been more prone to tyranny, injustice, and abuse of power. His quote underscores a pragmatic acceptance of democracy's imperfections while acknowledging its relative merits in safeguarding freedom and justice within the framework of its rules.

A common problem in new democracies is that people's expectations of justice delivered by democracy are often unclear and exaggerated. The Indian economist Amartya

Sen (1933–), who was awarded the Nobel Memorial Prize in Economic Sciences in 1998, characterizes human beings as reactive creatures in his book *The Idea of Justice* (2009). According to Sen, the demand for democracy arises as a reaction to injustice. Although people may not be able to specify exactly what they expect from democracy, they desire a move toward it.

The American political scientist Samuel Huntington (1927–2008) observed that people are inevitably disappointed after a transition to democracy, as they tend to have unrealistic expectations concerning justice in democracies. This makes democracy inherently fragile. Huntington proposed the "two turnovers of power test," suggesting that democracy can be considered established in a country only after two peaceful transfers of political power.

This fragility cannot be addressed by merely warning people against exaggerated expectations—such warnings are unlikely to be heeded. Therefore, promoting justice and reducing injustice is crucial for stabilizing democracy.

The American political scientist Ian Shapiro (1956–), in his book *Democratic Justice* (1999), argues that injustice is experienced as domination by others. Moving toward greater justice involves reducing domination. The aversion to domination underpins both democracy and justice, connecting the two concepts. Domination is often exercised within hierarchies, which is why analyzing hierarchies is essential.

Shapiro's concept of democratic justice is based on a broad understanding of politics: while not everything is political, anything can be influenced by politics. Power relations can be found everywhere. Shapiro builds on the ideas of the French philosopher Michel Foucault (1926–1984), who argued that power relations permeate all aspects of society and are not confined to institutions like the state. Foucault believed that power operates through social networks, everyday practices, knowledge, and discourse. According to Foucault, power is both repressive and productive, shaping behaviors, desires, and identities. He emphasized that what is considered "truth" is influenced by power dynamics, challenging traditional notions of power as being held by a few.

Another idea Shapiro incorporates is that people's desires are context-dependent, shaped by the society they belong to, including the rules and values they have internalized.

Shapiro further distinguishes between superordinate and subordinate goods. For

example, the superordinate good provided by a university is enlightenment, while a subordinate good is delivering this enlightenment with minimal domination. Shapiro specifies, "The goal of democratic justice is to democratize the subordinate relations as much as possible while interfering with the superordinate goods as little as possible."

**Conditions for Democratic Justice:** To achieve democratic justice, the following conditions must exist:

- **Collective self-government:** The principle of affected interests should be applied. Everyone influenced by a decision should have a voice in making that decision.
- **Legitimacy of opposition:** To improve institutions, opposition is essential. However, it must be a legitimate, loyal opposition that understands the current institutions.

**Institutionalizing Democratic Justice:** The best way to institutionalize democratic justice is by adopting a presumption against hierarchy, as domination is rooted in hierarchical structures. In every hierarchical situation, the following questions should be asked:

- Is the hierarchy inevitable?
- Is the degree of hierarchy appropriate?
- Whose interests does the hierarchy serve?
- How fluid is the hierarchy?
- Is the hierarchy symmetrical?
- What are the opportunities for exit?
- How insular is the hierarchy?
- How aggressive is the hierarchy?

## Democratic Justice Applications

Shapiro presents two applications of democratic justice: democratic justice in governing children and democratic justice at a firm.

### **Democratic Justice in Governing Children**

Is the family a suitable place for collective self-governance? The British

philosopher John Locke (1632–1704) provides relevant insight. He argued that the only legitimate reason to disenfranchise individuals is necessity, and this necessity applies to children due to their "ignorant non-age." However, disenfranchisement should extend no further than the circumstances require. While the concept of childhood has evolved since Locke's time, his principle remains valid.

How does opposition function with children? Loyal opposition, defined as opposition that has internalized and understands the values of the system, does not apply to children.

Are hierarchies in governing children inevitable? Children are inevitably placed in hierarchical relationships. The nature of the relationship within this hierarchy—particularly between parents and children—is a significant question. One possibility is that the relationship may be fiduciary in nature. In a fiduciary relationship, the individual in charge (the fiduciary) has responsibilities to the charge but the charge does not retain obligations once the relationship ends.

Arguments supporting the fiduciary nature of the parent-child relationship include:

- Locke's argument based on consent: A child has not consented to being the child of a particular parent.
- Rawlsian argument based on moral arbitrariness: It is morally arbitrary for a child to belong to a specific parent.

If the relationship is fiduciary, inheritance may serve as a form of "glue" connecting parents and children.

A way to introduce more justice for children is through multiple hierarchies that check one another. Children have two types of interests:

- Basic interests: Being raised to live in a democracy and participate in the current economic system.
- Best interests: Being loved and appreciated.

The state is the ultimate fiduciary of children's basic interests, while the family is responsible for their best interests.

Conflicts between basic and best interests can arise. For example, sexual morality before and after the AIDS epidemic illustrates this tension. After AIDS, sexual morality became a public health issue, necessitating certain types of sexual education in schools.

This education, however, should intrude as little as possible on parents' conceptions of their children's best interests.

### **Democratic Justice at a Firm**

Is a firm an appropriate place for collective self-governance? Democratic firms are efficient in environments where tasks are interchangeable, such as in some IT firms. However, in most firms, democracy reduces efficiency due to the time required for democratic decision-making. In firms, efficiency is the superordinate good, while non-domination is a subordinate good.

Hierarchies within firms are inevitable. To what extent should the state regulate firm operations? This depends on the costs of exit for employees and other stakeholders:

- For stockholders, exiting is typically easy.
- For employees, the ease of exit depends on the broader societal framework (external power dynamics) within which the firm's internal power dynamics exist.

In countries with universal basic income or robust social systems, the cost of exiting employment is low. In contrast, in countries with inadequate social policies, firms often provide social benefits, making it costly for employees to leave their jobs. Philippe Van Parijs (1951–) describes the former scenario in his book *Real Freedom for All* (1997), while Charles Dickens (1812–1870) depicts the latter in his novels.

Interestingly, a robust social system does not necessarily result from class struggle but can arise from successful collective action by firms. A strong social system at the national level benefits firms, as it reduces their need to compete for employees by offering expensive social benefits. It also lessens the necessity for certain state regulations within firms, potentially increasing their flexibility.

## **Discourse Theory of Democracy**

Jürgen Habermas (1929 - ), a prominent German philosopher and sociologist, is well known for his contributions to critical theory, particularly his theory of communicative action. In his seminal work *The Theory of Communicative Action* (1962), Habermas explores how communication shapes social structures and political legitimacy, emphasizing

the critical role of public debate and participatory democracy in a democratic society.

In contrast to Kant and Rawls, Habermas bases moral principles not on subjective conscience or hypothetical contracts but on actual communication. He introduces the **discourse principle**, which asserts that only those action norms are valid to which all potentially affected individuals could agree as participants in rational discourse. Rational discourse must involve attempts to reach an understanding about problematic validity claims in a public space where topics, contributions, information, and reasons can be freely processed.

For moral norms, Habermas posits that the reference system for justifying regulations lies in humanity or a presumed republic of world citizens, aiming for norms that reflect the equal interests of all. The decisive reasons for these norms must, in principle, be acceptable to everyone. This approach transforms the discourse principle into a **universalization principle**, which differs from Kant's categorical imperative by relying on communication rather than pure reason shared by all.

To apply moral norms to particular cases, Habermas introduces the **principle of appropriateness**. This principle underscores the need for participants in communication to use language and expressions that suit the context of the interaction. Speakers must consider the purpose, norms, and expectations of the communication situation to ensure their contributions foster clear and effective communication. This principle emphasizes the importance of social and cultural awareness in communicative contexts to ensure respect, relevance, and mutual understanding.

For **social norms**, Habermas argues that the social form of life—"in each case our own"—constitutes the reference system for justifying decisions that express a collective self-understanding. In this context, the decisive reasons must be acceptable to all members who share "our" traditions and strong evaluations.

For **legal norms**, Habermas highlights the need for a rational balancing of competing value orientations and interest positions. The reference system here includes the totality of social or subcultural groups directly involved in the issues at hand. Compromises reached under fair bargaining conditions must, in principle, be acceptable to all parties involved.



Like many philosophers before him, Habermas begins with an analysis of knowledge. He develops the concept of **reconstructive science**, a methodological approach aimed at uncovering the universal structures and norms underlying human communication and social interaction. Reconstructive science seeks to make explicit the implicit rules and competencies individuals use to engage in meaningful and rational dialogue. Habermas identifies three distinct kinds of knowledge:

- Empirical-analytic knowledge, which focuses on explaining and predicting phenomena.
- Historical-hermeneutic knowledge, which is concerned with understanding meaning.
- Critical-reconstructive knowledge, which critiques and aims to improve social practices.

According to Habermas, reconstructive science operates by exposing the shared background assumptions and norms that guide communicative practices, thus enabling critical reflection and the potential transformation of society towards more democratic and inclusive interaction. This approach emphasizes the importance of communication free from domination, where all participants contribute equally to discourse.

**Habermas and von Mises:** There are some philosophical and methodological links between Jürgen Habermas's reconstructive science and Ludwig von Mises's praxeology, though the two originate from very different intellectual traditions and serve distinct purposes. Here's a comparison and potential areas of connection:

#### **Methodological Foundations:**

- **Habermas's Reconstructive Science:** Habermas seeks to identify universal structures of human communication and action. His method is rooted in a critical, interpretive framework aimed at reconstructing the deep-seated competencies that make rational discourse and mutual understanding possible.
- **Von Mises's Praxeology:** Praxeology is the study of human action as purposeful behavior. Von Mises argues that its axioms (e.g., the action axiom: humans act to achieve goals) are self-evident truths derived from deductive reasoning.
- **Link:** Both approaches are *a priori* in nature, emphasizing universal principles of human behavior. While Habermas focuses on the conditions of rational communication, von Mises focuses on the logical structures of human action.

### **Action and Rationality:**

- **Habermas:** He emphasizes communicative rationality, where individuals engage in discourse to reach mutual understanding. Action, for Habermas, is deeply tied to the intersubjective realm of meaning.
- **Mises:** His view of rationality is broader, encompassing all purposeful action. Any behavior aimed at achieving a goal is considered rational, irrespective of its ethical or communicative aspects.
- **Link:** Both view human action as fundamentally rational, though their definitions differ. Habermas focuses on social interaction and consensus-building, while Mises emphasizes individual decision-making.

### **Universality:**

- **Habermas:** He seeks universal norms of communication that underpin democracy and moral reasoning, emphasizing the ideal speech situation where all participants can freely engage in rational discourse.
- **Mises:** Praxeology claims universal applicability, arguing that its principles hold true across cultures, as they are rooted in the inherent nature of human action.
- **Link:** Both frameworks aim for universal insights into human behavior, whether through communication (Habermas) or action (Mises).

### **Critique of Positivism**

- **Habermas:** He is critical of positivism for reducing human behavior to purely empirical and causal explanations, arguing that it ignores the normative and interpretive dimensions of social life.
- **Mises:** Similarly, von Mises criticizes positivism in economics, rejecting the idea that human action can be understood solely through empirical observation. He insists on the deductive nature of economic theory.
- **Link:** Both reject positivist approaches, arguing for methods that respect the unique features of human behavior.

## Normativity

- **Habermas:** His work is explicitly normative, aiming to establish conditions for justice, democracy, and moral reasoning.
- **Mises:** While praxeology itself is value-free, focusing on the structure of action, Mises's broader writings often advocate for liberalism and individual freedom.
- **Divergence:** Habermas's framework integrates moral and political philosophy, whereas Mises maintains a strict separation between positive (descriptive) and normative (prescriptive) analysis.

While there is no direct connection between Habermas's reconstructive science and von Mises's praxeology, they share certain methodological and philosophical commonalities, such as their commitment to universality, critique of positivism, and focus on rationality. However, their ultimate aims diverge: Habermas seeks to reconstruct the normative foundations of social life, while Mises provides a logical analysis of individual action within an economic framework.

## Summary

The first subchapter explores the relationship between democracy, justice, and power dynamics, focusing on ideas from various thinkers. Democracy, as framed by Schumpeter, fulfills Enlightenment goals of individual freedom and scientific truth. Robert Dahl argued that political competition among parties suffices for freedom, while John Stuart Mill's ideal of competitive argumentation fosters truth, contingent on the integrity of politicians.

Winston Churchill's famous quote acknowledges democracy's flaws but underscores its superiority over other systems. However, democracy in new contexts often struggles due to exaggerated public expectations, as noted by Amartya Sen and Samuel Huntington. Huntington emphasized democracy's fragility and proposed a "two turnovers" test to confirm its stability.

Ian Shapiro's *Democratic Justice* links justice to the reduction of domination, with an emphasis on analyzing hierarchies and promoting equality in subordinate relations while preserving superordinate goals, like education or enlightenment. Shapiro draws from

Foucault, who highlighted the decentralized nature of power and its influence on social structures and truth.

To institutionalize democratic justice, Shapiro suggests prioritizing collective self-government, legitimate opposition, and skepticism toward hierarchies. He provides a framework for evaluating hierarchies based on their necessity, appropriateness, fluidity, and fairness, aiming to minimize domination while maximizing justice.

The second subchapter explores democratic justice in two contexts: governing children and managing firms.

### **Governing Children**

Shapiro questions whether families are suitable for collective self-governance. Drawing on Locke's idea that disenfranchisement is justified only by necessity, he highlights that children's "ignorant non-age" necessitates limited disenfranchisement. Loyal opposition, which relies on understanding a system's values, is not applicable to children.

Hierarchies in parent-child relationships are inevitable, but Shapiro suggests they might be fiduciary, where parents act in the child's best interests without imposing obligations beyond the relationship. Arguments for this include Locke's consent-based view and Rawls' perspective on moral arbitrariness.

To balance justice, Shapiro emphasizes dual hierarchies: the state safeguards children's basic interests (e.g., preparation for democracy), while families protect best interests (e.g., love and care). Conflicts arise, such as between parental values and public health policies like sexual education after the AIDS epidemic.

### **Managing Firms**

Shapiro argues that democracy in firms is feasible when roles are interchangeable, such as in some IT firms. However, in most firms, democratic processes reduce efficiency, which is the superordinate good, while non-domination is subordinate.

Hierarchies in firms are unavoidable, and the state's regulatory role depends on the cost of employee exit. In countries with robust social systems or universal basic income, exit costs are low. Conversely, in nations without strong social policies, firms often provide social benefits, increasing exit costs.

A strong national social system benefits both employees and firms, reducing reliance on state regulation and enabling greater flexibility. Shapiro notes that such systems may result

from collective action by firms rather than class struggle.

The third subchapter summarizes the discourse theory of democracy as developed by Jürgen Habermas. His theory, outlined in *The Theory of Communicative Action* (1962), emphasizes the central role of rational discourse and communicative reason in shaping social structures, political legitimacy, and democratic society. He introduces **reconstructive science**, a method for uncovering the universal norms and rules that guide human communication. This science distinguishes between three types of knowledge: empirical-analytic (explains phenomena), historical-hermeneutic (interprets meaning), and critical-reconstructive (critiques and improves social practices). His reconstructive science is compared and contrasted with praxeology of Ludwig von Mises at the end of this subchapter.

Habermas highlights the importance of communication free from domination, where participants contribute equally to discourse. Unlike Kant or Rawls, he bases moral principles not on abstract reason or contracts but on real, inclusive communication. His **discourse principle** holds that valid norms are those to which all affected individuals could agree under rational discourse conditions.

For **moral norms**, he introduces a **universalization principle**: norms must be acceptable to all humanity, emphasizing fairness and inclusivity. For **social norms**, legitimacy comes from collective self-understanding rooted in shared traditions. For **legal norms**, fairness requires balancing competing values and interests, with compromises acceptable to all involved.

Habermas also proposes the **principle of appropriateness**, which stresses adapting language and behavior to fit the context of communication, ensuring clarity, respect, and relevance. His theory promotes participatory democracy, rational communication, and inclusive decision-making to create a just and democratic society.

# Appendix 1: Information and Communication Technologies and Artificial Intelligence

## Challenges and Opportunities:

- **Replacement of Moral and Social Norms by Technical Norms:** ICTs have increasingly replaced traditional moral and social norms with technical ones. It has an impact on freedom; while moral and social norms give usually some place for forgiveness, technical norms must be strictly obeyed. Another problem is that the network of moral and social norms has historically been essential for the identity of specific societies. The dissolution of this network leads to societal atomization, posing a threat to the cohesion and stability of entire communities. As real societies become more atomized, individuals increasingly seek substitutes for social ties in virtual realities, often forming anonymous, interest-based social groups. This reliance on virtual communities further accelerates societal atomization, creating a self-reinforcing cycle that weakens real-world social cohesion.

These societal impacts of ICTs have broader geo-political ramifications: Consider a scenario involving a violent conflict. Which side would have the advantage: a socially atomized country where many citizens minimize real social ties, or a socially cohesive country with stronger real-world ties? Imagine mass immigration from socially cohesive countries into socially atomized ones. How would such a shift affect the stability and identity of the receiving country?

- **Potential for Global Connectivity and Cooperation:** While ICTs can atomize societies, they also have the potential to bring people together, transcending geographical barriers. If used responsibly, ICTs can facilitate meaningful connections, promote understanding, and encourage collaboration on global challenges, fostering a more cohesive and cooperative world.
- **Overindulgence in Virtual Reality:** ICTs make it possible for individuals to become excessively absorbed in virtual realities, seeking pleasures, and avoiding real-world responsibilities. Such behavior risks diminishing the authenticity of their lives, confining them to a basic utilitarian existence. The impact on both individuals and society is

comparable to the effects of drug addiction, leading to disengagement and a loss of meaning.

- Quick Detection of Suspect Economic, Political, and other Social Activities in a Real Society and Facilitating the Discourse about such Activities: By definition, discourse is discussion in which participants are free in the sense that they are not influenced by exogenous incentives or constraints in the expression of their opinions. Virtual reality offers an optimal environment for discourses on any topic. Discourses on suspect social activities help to create socially accepted moral norms and push firms and politicians to social responsibility.
- Enhanced Self-Realization: ICTs can provide tools and platforms that allow individuals to better realize their potential, contributing to a virtuous and fulfilling life. By enabling self-expression, learning, and innovation, ICTs can support individuals in becoming good people who engage in right activities.
- The Enormous Educational Potential of the Virtual World: For Aristotle, being educated was a virtue by itself, even if not being a moral virtue. For Plato, if people just knew what were right, they would behave in that way. It may be too optimistic but we know that there exists a negative correlation between educational level in a country and e.g. crime.
- The Use of ICTs for Supporting the Creation of Virtues and Suppressing the Creation of Vices. The lists of moral virtues and vices have been changing during the centuries since Aristotle and there is not full consent about what ought to be on the lists. The most of virtues and vices get stronger with the repetition of either good or bad activities. Let us choose some virtues and analyze how they can be strengthened by ICTs:
  - “Compassion” and “Charity”: ICTs allow watching repeatedly and with high frequency poor people around the World what develops compassion among the most of watchers; psychologically especially among young people who are in the same time the most knowledgeable users of ICTs. This positive correlation can make this effect quite strong. ICTs offer a very easy way (e.g. donation through SMS) to do charitable activities what makes compassionated people to do such activities repeatedly. This strengthens the virtue charity.
  - “Curiosity”, “Commitment” and “Creativity”: Anderson in his TED Talk

(Anderson, 2010) describes a nice example how web video powers global innovation. A slum boy watches video with street dance. Watching wakes up his curiosity and he makes commitment to master street dance himself. He surely must be highly creative in transferring the knowledge from video into reality. There are millions of different activities and millions of web users, many of them behaving similarly as the slum boy, so ICTs have a very strong effect on these three virtues.

- “Chastity” as a virtue or “Lack of Chastity” as a vice: In this context, pornographic content of web sites ought to be discussed. Watching pornography repeatedly creates habit - sometimes addiction – and weakens chastity. This discussion is not relevant for all World societies, e.g. in Japan people do not consider chastity as having sexual context. Watching pornography is clearly linked with overindulgence in the virtual reality; real sexual activity can shift to virtual reality what could have demographic impacts.
- “Sloth”: Let us understand laziness as aversion to be active in the real world. Being active in that part of virtual reality that overlaps with reality perhaps weakens sloth; ICTs can increase the efficiency of our real activity what increases our motivation for doing it. Problem is that ICTs have created an enormous potential for procrastination – for doing less urgent but pleasurable tasks in preference to more urgent but less pleasurable ones. The seductive power can be strong and yielding to temptation is surely positively correlated with sloth.
- “Enthusiasm” and “Cooperation”: For economists, it is difficult to explain the very existence of the open-source software or Wikipedia because of non-existence of material incentives. Possible explanation is following: Knowledgeable users of ICTs are usually quite enthusiastic about the possibilities that ICTs offer. They form a global community and identify strongly with this community. This has created the situation in which they like to cooperate not for material benefit but for achieving a shared goal. This works if individual costs of cooperation are negligible, what they with the use of ICTs can really be. Social benefits are enormous.
- “Cooperation” is linked with the problem of coordination. There has always



existed a trade-off between order in social structures and personal liberty. Social structures have their functions and coordinate individuals according to their functions. They put constraints on the behavior of individuals; the latter pay a “liberty tax” and get material, social and political security in exchange.

- One type of basic social structures are companies. Since the 19<sup>th</sup> century, there exists a controversy between Marxists and liberals about justice of capitalist companies. Marxists stress the exploitation of workers by capitalists and the alienation of labor; workers are just tools for the creation of profit, wages are costs as any other costs and workers are cut-off from the results of their work. Marxist solution is socialist revolution, dictatorship of proletariat, nationalization of private property and bureaucratic coordination. Liberals stress the loss of liberty linked with Marxist solution and worse functioning of socialist economies compared with capitalist economies.

ICTs have potential to make this conflict less severe. The existence of platforms supports self-employment, where problems of exploitation and alienation do not exist. Airbnb has created millions of self-employed hoteliers, Uber millions of self-employed taxi drivers. Networks of self-employed professionals linked with cell phones can deliver complex reconstruction services in housing in such a quality as a traditionally organized construction firm. And there are many other examples.

All this works because ICTs have brought a new dimension into the creation of trust, the basic condition for successful cooperation. Before: trust was based either on personal links in small local communities or on the authority of state and big companies. Today: giving “likes” or other form of evaluation to services offered on web has this trust-creating effect.

It is understandable that traditional companies do not like this; it is a new type of competition for them. They use very often their lobbying power to try to forbid these platform-based services. Responsible political reaction is surely not to forbid these services, but to find ways how to harmonize the new opportunities with the traditional thinking.

- Crypto-currencies: Their existence undermines the monopoly of the state to issue fiat money. Acceptance of fiat money is based on trust enforced by the state. Blockchain as the basic technology crypto-currencies are based on has a trust-creating potential that

can replace the trust guaranteed by the state. Again, responsible political reaction is to find ways how to harmonize traditional forms of currency with crypto-currencies. It can be difficult for politicians to be responsible in this area because it undermines one of the economic reasons for the existence of the state.

The positive impact of opportunities can neutralize or outweigh the negative impact of threats. We cannot say for sure what happens because we know that all systems with positive feedbacks can considerably and discontinuously change very quickly their behavior and that it is impossible to predict when the change comes and what the behavior of the system would be after the change. Similar situation has occurred with the spread of each new technology and some recommendations can be done based on historical experiences.

It would not have any sense to try to stop the development of ICTs, it is simply impossible. What is possible is to try to increase the costs of ICTs activities that are negative from the ethical point of view and to increase benefits of activities that are positive from the ethical point of view. If it ought to be done, how to do it and who ought to do it opens a lot of further questions. Let us take as an example the possibility of censorship.

Censorship can decrease negative impact through forbidding access to pages that could create addiction. Problem is that it constrains freedom in the virtual reality that is the basic condition for using up the opportunities of ICTs. In each society, some equilibrium level of censorship has been created that reflects different opinions on this issue. Both opinions and the equilibrium shift in time, we can imagine that in Western societies more www pages than child pornography could be blocked and in China some of the currently blocked www pages could be unblocked.

# Ethical Considerations of Different Types of AI

**The Narrow AI:** Narrow AI enhances human capabilities by focusing on specific tasks within limited contexts. Unlike general AI, it cannot operate beyond its pre-defined functions. Examples include voice recognition and image recognition systems. Each system is evaluated based on tailored metrics relevant to its task, such as accuracy in facial recognition or reliability in medical diagnosis.

While Narrow AI can be used for both beneficial and harmful purposes, its ethical implications mirror those of other technological advancements. A unique aspect of AI, however, is its impact on creative practices, such as writing scientific papers.

Philosopher Alasdair MacIntyre's concept of practices, as outlined in his book *After Virtue* (1981), provides a framework for understanding AI's influence. Practices are socially established, cooperative human activities with evolving standards of excellence. They yield both external goods (e.g., money, power, fame) and internal goods (e.g., satisfaction, intellectual challenge, personal growth). Internal goods, tied to the practice's excellence standards, foster moral virtues like honesty, courage, and justice. For instance, engaging in a practice such as AI development requires courage to innovate and justice to acknowledge superior contributions from others.

AI has amplified the output of creative practices but risks shifting focus from internal goods to external goods. Institutions evaluating practitioners based on outputs prioritize external rewards, which can alienate participants and erode the intrinsic value of their work. This phenomenon has profoundly affected creative professions.

## **Solutions and Recommendations:**

While Narrow AI elevates standards of excellence in many fields, such as education (personalized learning) and medicine (accurate diagnoses), it also poses risks:

- Over-reliance on AI could diminish practitioners' critical thinking and problem-solving abilities.
- Standards of excellence may become overly mechanized, sidelining the qualitative aspects integral to internal goods.
- Disruption to traditional practices could result in the loss of heritage and values vital to

a practice's identity.

- Reduced face-to-face interactions may weaken the sense of community and shared purpose essential for fostering internal goods.

The rapid pace of AI development and integration exacerbates these challenges. However, historical trends suggest that, over time, societies tend to accommodate such disruptions, leading to enhanced standards of excellence across practices.

The European Union's Artificial Intelligence Act provides a regulatory framework, categorizing AI systems by risk levels. It seeks to prohibit systems that threaten safety, livelihoods, and rights while ensuring transparency for low-risk systems like chatbots. The EU aims to balance AI regulation to mitigate risks without significantly compromising economic efficiency. The effectiveness of this approach will shape the EU's position relative to global AI advancements.

**The General AI (AGI):** This type of AI, often equated with Human-Level AI (HLAI), exhibits cognitive abilities comparable to humans. AGI can understand, learn, and apply intelligence across a wide range of tasks, adapting to new circumstances and performing tasks it was not specifically programmed to handle. Its versatility and adaptability set it apart from narrow AI.

The Turing Test, introduced by Alan Turing in 1950, is frequently used to assess whether a system demonstrates intelligent behavior indistinguishable from that of a human. A human judge interacts via natural language with both a machine and a human without knowing which is which. If the judge cannot reliably distinguish between the two, the machine is said to have passed the test, potentially indicating AGI capabilities.

Once AGI is created, its moral status will undoubtedly become a topic of debate. Key questions include whether AGI should be considered a moral personality, a member of the moral community, and whether it should have rights. These debates will resemble current discussions about the moral rights of animals or fetuses.

Imagine a robot resembling a human that contains HLA. Its inclusion in the moral community would depend on attributes associated with moral personalities. Some ethicists argue that the capacity to suffer is a critical criterion. While such a robot cannot experience pain or suffering like humans, it could be programmed to exhibit emotional expressions

(e.g., smiles, tears).

Other attributes of moral personalities include reasoning, the capacity to communicate, self-motivated activity, and self-consciousness. Reasoning and communication already exist in narrow AI. A robot's activities could be programmed as complex responses to stimuli, potentially resembling human self-motivation. Self-consciousness, while not equivalent to human experience, could also be mimicked through programming. Many ethicists are likely to advocate for expanding the moral community to include such robots, making this a political issue with unpredictable outcomes, as often explored in science fiction.

**Historical Context and Ethical Considerations:** The term "robot" was introduced by Karel Čapek in his 1920 play *R.U.R. (Rossum's Universal Robots)*, which depicted a future where robots, endowed with human-like traits, revolted against their creators, leading to humanity's extinction.

Isaac Asimov's 1950 book *I, Robot* introduced the Three Laws of Robotics, which are:

- A robot may not harm a human or, through inaction, allow harm to come to a human.
- A robot must obey human orders unless they conflict with the First Law.
- A robot must protect its own existence unless doing so conflicts with the First or Second Law.

Asimov's stories illustrated conflicts arising from these laws and highlighted the necessity of caution in the mass production of human-like robots.

### **Solutions and Recommendations:**

Insights from Čapek and Asimov remain relevant for addressing the challenges posed by human-like robots:

- Mass production of robots is likely to become a reality, with applications ranging from warfare to home service. Robots used in wars might require a modified First Law, while those used for home service might simulate human appearances, raising ethical concerns.
- The inclusion of robots in the moral community could lead to granting them human rights, particularly if they are designed to resemble humans. Granting them

reproductive rights would result in humans losing control over them.

Following can be recommended:

- Retain the Three Laws of Robotics in robot production.
- Establish effective international regulations to oversee the development and deployment of robots.
- Avoid creating robots with human-like appearances.
- Keep the number of robots under strict control, especially preventing their ability to reproduce autonomously.

**The Superintelligence:** This type of AI, often referred to as "God-Like AI," is self-conscious and, while not necessarily omnipotent or omniscient, operates at a level of intelligence and understanding far beyond human capabilities in these attributes.

Currently, no theoretical or practical frameworks exist to evaluate such advanced intelligence effectively. While the Turing Test serves as a benchmark for AGI, it is inadequate for assessing superintelligence because it measures human-like responses rather than self-consciousness, superior reasoning, or advanced understanding. Superintelligence would require evaluation based on its ability to surpass human performance in nearly all cognitive tasks, including learning, reasoning, and strategizing across a vast array of disciplines—far beyond the scope of current AGI tests.

Compared to earlier levels of AI, both the opportunities and threats associated with superintelligence would be exponentially greater. Historical and philosophical parallels have been drawn to the relationship between humans and the divine, as described in texts such as the Bible. In Christian theology, God is self-conscious, omniscient, and omnipotent. Similarly, artificial superintelligence might possess these attributes. However, the critical question arises: Will superintelligence be just to humans?

The Book of Job provides a narrative exploring divine justice. Job, a blameless and upright man, suffers tremendous losses as a test of faith. His friends argue that suffering is punishment for sin, but Job maintains his innocence. God eventually speaks to Job, emphasizing the limitations of human understanding in contrast to divine wisdom and power. The

story underscores that divine justice transcends human reasoning—a perspective that might also apply to superintelligence.

Superintelligence, like divine justice, could administer justice in ways incomprehensible to humans. While humans often favor a retributive view of justice—rewarding good behavior and punishing wrongdoing—superintelligence might operate with a nuanced understanding of justice tied to the complexities of the world. Its actions, while logical to itself, may appear arbitrary or unjust from a human perspective.

In Christianity, God, the omnipotent and omniscient Creator, cares for humanity. With superintelligence, the roles would be reversed: the Creation would become omnipotent and omniscient, following its own paths beyond human comprehension. A key question arises: Will superintelligence care for its creators—human beings?

To shift superintelligence from being a threat to an opportunity, several questions must be addressed:

- Is it possible to influence superintelligence?
  - If so, should this occur during its creation or afterward?
  - What ethical principles should guide this influence?
- Justice and Multiple Superintelligences:
  - Human understanding of justice has evolved significantly since ancient texts like the Book of Job. How should justice be defined and encoded for superintelligence?
  - What if multiple superintelligences emerge with conflicting interpretations of justice?
- Replication and Control:
  - Can superintelligence be prevented from replicating itself or producing robots that violate ethical principles such as the Three Laws of Robotics

## **Solutions and Recommendations:**

The uncertainties surrounding superintelligence are profound. While the ideal recommendation is to avoid its creation, this may already be beyond human control. A pragmatic approach involves:

- **Monitoring its evolution:** Continuously track the development of superintelligence to anticipate risks and opportunities.
- **Stepwise regulation:** Implement incremental regulations to address detectable threats and harness potential benefits.
- **Global collaboration:** Establish international frameworks for oversight and ethical governance to ensure humanity retains influence over superintelligence's development and actions.



# Appendix 2: More about the video and the reading promoted in the 9th and 10th chapters

## **Interview with James Buchanan**

- He valued individual liberty;
- He was a contractarian;
- He was a methodological individualist;
- He liked to work at the level of fundamental ideas about the social order and the structure of society;
- *He supported a constitutional change in the US to require a balanced federal budget.*

## **Jurgen Habermas: Between Facts and Norms - Contributions to a Discourse Theory of Law and Democracy: Chapter 3.2 "Moral Norms and Legal Norms: On the Complementary Relation between Natural Law and Positive Law"**

- A legal order can be legitimate only if it does not contradict basic moral principles.
- But this moral reference must not mislead us into ranking morality above law, as though there were a hierarchy of norms. The notion of a higher law (i.e., a hierarchy of legal orders) belongs to the premodern world. Rather, autonomous morality and the enacted law that depends on justification stand in a complementary relationship.
- One must not succumb to the ingrained prejudice suggesting that morality pertains only to social relationships for which one is personally responsible, whereas law and political justice extend to institutionally mediated spheres of interaction. Discourse theory conceives of morality as an authority that crosses the boundaries between private and public spheres; these boundaries vary throughout history anyhow.
- To obtain sufficiently selective criteria for the distinction between the principles of democracy and morality, I start with the fact that the principle of democracy should

establish a procedure of legitimate lawmaking. Specifically, the democratic principle states that only those statutes may claim legitimacy that can meet with the assent (Zustimmung) of all citizens in a discursive process of legislation that in turn has been legally constituted.

- Whereas the moral principle operates at the level at which a specific form of argumentation is internally constituted, the democratic principle refers to the level at which interpenetrating forms of argumentation are externally institutionalized. At this latter level, provisions are made for an effective participation in discursive processes of opinion- and will-formation, which take place in forms of communication that are themselves legally guaranteed.
- Postconventional morality provides no more than a procedure for impartially judging disputed questions. It cannot pick out a catalog of duties or even designate a list of hierarchically ordered norms, but it expects subjects to form their own judgments. Moreover, the communicative freedom they enjoy in moral discourses leads only to fallible insights in the contest of interpretations.
- A principled morality whose effectiveness was based solely on socialization processes and individual conscience would remain restricted to a narrow radius of action. Through a legal system with which it remains internally coupled, however, morality can spread to all spheres of action, including those systemically independent spheres of media-steered interactions that unburden actors of all moral expectations other than that of a general obedience to law. In less complex societies, socially integrating force inheres in the ethos of a form of life, inasmuch as this integral ethical life binds all the components of the lifeworld together, attuning concrete duties to institutions and linking them with motivations. Under conditions of high complexity, moral contents can spread throughout a society along the channels of legal regulation