Renáta Kišoňová

Department of Theory of Law and Philosophy of Law

Comenius University, Bratislava

renata.kisonova@flaw.uniba.sk

THE IDENTITY AND GENDER PROBLEM IN THE CONTEXT OF AI

Abstract

My paper[1] focuses on the problem of identity and gender in the context of artificial intelligence and human rights. What identity does belong to AI? How can the society solve the gender problem regarding AI? I analyse the terminology identity, gender, sex in interdisciplinary point of view, considering philosophical, ethical, cognitive and biological aspects of identity.

Keywords: *AI, identity, gender, sex, transgender, Simone de Beauvoir, Human rights*

My paper focuses on the problem of identity and gender in the context of artificial intelligence and human rights. What identity does belong to AI? How can the society solve the gender problem regarding AI? I analyse the terminology identity, gender, sex in interdisciplinary point of view, considering philosophical, ethical, cognitive and biological aspects of identity. I will start for relief with asimply analogy: Siri, the artificially intelligent digital assistant trapped inside Apple's iPhone, won't tell you what its gender is if you ask. "I exist beyond the human concept of gender," Siri says. When I ask if it's sure it's not a woman, Siri says, "Sorry, I don't really know." Gender, identity and sexual orientation are fundamental characteristics of an individual's sexual identity and refers to a person's innermost concept of self as male, female or something else and can be the same or different from one's physical

---

sex. The establishment of gender identity is a complex phenomenon and the diversity of gender expression argues against a simple or unitary explanation. For this reason, the extent to which it is determined by social vs biological factors continues to be debated actively[2]. We may contemplate with Simone de Beauvoir: "If the female function is not enough to define woman, and if we also reject the explanation of the "eternal feminine," but if we accept, even temporarily, that there are women on the earth, we then have to ask: What is a woman?"[3] The biological basis of gender identity cannot be modelled in animals and is best studied in people who identify with a gender that is different from the sex of their genitals, in particular transsexual people. Transgender people express their gender identities in different ways. Some of them use their dress, behavior, and mannerisms to live as the gender that feels right for them, some of them take hormones and may have surgery to change body so it matches their gender identity. Some transgender people reject the traditional understanding of gender as divided between male and female, so they identify just as transgender, or genderqueer, genderfluid, or something else[4]. There is a difference from being transgender and from sexual orientation. Transgender people may identify as heterosexual (straight), homosexual (gay or lesbian), bisexual, asexual, or otherwise, or may decline to label their sexual orientation. The term *transgender* is also distinguished from *intersex*, a term that describes people born with physical sex characteristics "that do not fit typical binary notions of male or female bodies". Why is a discussion about it so important for society? I will analyse this issue deeplier, from philosophical and interdisciplinary point of view. First of all, let me focus on the term society. The term society is etymologically related to the Latin term "societas", which is translated as community, unity, fellowship, the verb sociare is translated as connect, associate, band together, accompany, share, take part. Adjective socius means joint, allied and noun socius can be translated as a species, companion, participant, or even an ally. The term "society" can be, therefore, understood in various meanings: a group of individuals, a set of social phenomena, the sum of relationships among people, interdependent society, community, humanity

---

[2] Roselli, C. E. (2017). *Neurobiology of gender identity and sexual orientation.* In: Journal of Neuroendocrynology. 30 (7).
[3] Beauvoir, S. (2010). de. *The Second Sex.* Vintage Books, New York, p. 25.
[4] https://www.plannedparenthood.org/learn/gender-identity/transgender

as a whole, state, organized community of people who reside in certain territory and are connected by cooperation, a certain form or system of social relations, a certain stage of human history development. Vast majority of authors in the field of social philosophy are of the opinion that the reason for the formation of society was co-operation. How come that co-operation could evolve in nature, when *acts of crookedness* are much more common? One of the answers is offered by computer models–for example, the "sentence for sentence" program. It is built on a simple principle–every opponent that could swindle can be forced into co-operation by making the first helpful step and subsequently we repeat every other step after our opponent – if he co-operates, we co-operate, if he swindles, we swindle, as well[5]. The presented model has its weaknesses – for the model to be functional the game should not have only one round, after which we would not meet our opponent again; in that case the most logical behaviour is to swindle and get lost. It is difficult to determine the level of remuneration for the swindle – it should not be too high in order not to destroy the relationship between opponents, nor too low, so that opponents get the impression that there is nothing to play for. These principles apply, as stated by Koukolík, for business, marriage, and political relationships – for any type of relationship[6]. Of course, man is not the only a being that lives in a group that calls for cooperation. Koukolík describes behaviour of lions in a group – they hunt in a group, they bring up their cubs together and protect their territory against alien lions. Their cooperation is based on a relatively complicated model – they co-operate, if a great effort is required in order to catch a pray, if it can be obtained easily, they do not co-operate. Females nurse the cubs together, but in case of delay attack by male, they prefer to defend their own cubs first. Adult females live in groups of 3-6 members, but sometimes as many as 18 females group together. Zoologists R. Heinsohn and C. Packer have examined the behaviour of female lions in case of threat to their common territory[7]. They have discovered that some of the females fight for the group and themselves with no regard to the threat and crooked behaviour of their companions. Others co-operate only in case of real emergency and

---

[5] Koukolík, F. (1997). *Mravenec a vesmír–O hvězdách, atomech,* životě *a vědcích*. Vyšehrad.
[6] Ibid.
[7] Heinesohn, R., Packer, C. (1995). *Complex Cooperative Strategies in Group – Territorial African Lions*. in: Science, 269, p.1260-1262.

the reminders of the group are "swindlers". Similar clashes for resources, territory or sexual partner among animals can be observed among human beings. Ordinary experience is sufficient for a man to find out that people behave in one way (both the man and higher vertebrates) with own relatives and differently with those, who they are not related to. Kinship is one of the essential organizational principles of all societies – plants, animals, humans. Relatives share a part of their genes, and therefore, if a man wants to increase the chances for survival of our genes, it seems advantageous to help relatives. Helping does not mean inbreeding – for close relatives the interaction of destructive alleles, which are carried by every living organism, is highly probable. The usual result is affected offsprings. The negative genetic consequences can be present even in offsprings of genetically very distant parents. Therefore, a man produces offspring with individuals who are not genetically too close, nor too distant[8]. The question is how to recognize a relative? For example, the bank swallow leaves the chicks that are unable to fly, in previously prepared corridors and therefore it feeds only its own offspring. When they learn to fly and they mingle with other groups, they produce characteristic sounds, by which their parents are able to recognize their own offspring. There is a variety of mechanisms for the recognition of own offspring from the others – e.g.: scent (wasps, bees, mice, humans – for example, see the experiment described by Koukolík[9]). We still have not addressed the question regarding the development of cooperation in nature. J.M. Smith and E. Szathmáry described the evolution of life in eight stages–initially there were simple molecules that were able to replicate; they formed more complex and collaborative groups and became the basis for the development of molecules that are the carriers of heredity. These were not destroyed in their struggle for resources and then came the formation of bacteria, protozoa, two sexes and multicellular organisms, colonies, primates – which formed a man. Each evolution level is more complex and contains more information than the previous one. It would seem that cooperation was the basis for evolution of all more complex life forms. If certain genes begin to multiply in a complex organism at the expense of entity, malignant tumour that destroys the body develops and without proper

---

8   Koukolík, F.(1997). *Mravenec a vesmír–O hvězdách, atomech,* životě *a vědcích*. Vyšehrad.
9   Ibid. p. 107-108

treatment the organism dies. It may be some sort of universal principle – the development of complex systems from genes to humans (language, differentiated culture, science, etc.) requires a sort of necessary feedback and control, that would limit opportunities of an individual in favour of the entire society, of course, to the extent that is not harmful to the individual or the entity. For millions of years, nature has been looking for such trade-offs – species, which did not cooperated became extinct and vice versa[10]. Not only cooperation and continual pursue for optimal compromise between opportunities (or freedom) of individual and good of all society are present in human society, but, as pointed out by many authors (whether in the field of social philosophy, psychology, anthropology, ethnology etc.), "fight for recognition" is present, as well. The process of integration of an individual into the society and vice versa, the impact of environment on the development of personality of the individual is called socialization. Two types of socialization can be distinguished: primary socialization – from the birth of a child as a part of education in a family that communicates the basic social standards and patterns of behaviour, and secondary socialization, which can be seen as a lifelong process following the primary socialization (school, work, etc.)[11].

The most widely known and experimentally studied mechanism of social cognition is a phenomenon known as *theory of mind*. This phenomenon indicates that the individual has a presumption of the contents of mind of another person–understands that other people have mental states that govern their behaviour[12]. An individual may be aware that at any moment the content of these mental states may be different from his own and from objective reality. (For example, there is a following trio of people – a woman, her husband and a stranger. The woman thinks that the stranger is boring; stranger believes that the woman consider him as extremely attractive; the husband suspects that the stranger believes that the husband's wife wants to run away with him. The stranger who does not know the relationship between the husband and the wife has the incorrect assumption as to the state of wife's mind; husband drew the correct conclusion on the contrary, because he understands signals of his wife and is able to interpret them correctly, and he has the so-called theory of mind.)

[10]   Ibid.
[11]   Nakonečný, M. (2009). *Sociální psychologie.* Praha: Academia.
[12]   Barrett L., Dunbar R., Lycett, J. (2007). *Evoluční psychologie* člověka. Praha: Portál.

On prejudices in social manner

Perception of others, social situations and interpretation of how people perceive each other, impressions they form of themselves (this actually starts the social interaction – what impression an individual form of others, that way he treats them). This formation of impression of others has been identified as "social perception". Today, the term is usually replaced by the term "social cognition", although several authors still use the accepted term *social perception*.

For example, S.T. Fiske and S.E. Taylor refer to the concept of social cognition as to how people form impression (make sense) of other people and themselves[13].

Social perception includes the ways of how people perceive themselves and other people in social situations; what judgments they create about themselves and the others.

It is not only the perception of social events, but also the interpretation of perceived reality.

Evidently, man perceives other people subjectively; his perception and assessment is influenced by social factors, expectations, experience and affectivity. Social perception is marked by errors, bias, facilitating, and so called "logical errors" that do not always give entirely reliable picture of the other person[14]. As P. Tavel (2012) states, social perception is a compromise between what a person expects based on their experience and what actually occurs in the outside world. The basic factors that influence social perception of persons are mentioned by P.F. Secord and C.W. Backman:

1. Information – physical appearance, expressive and other motoric symptoms, verbal behaviour
2. Variables – prior knowledge and feelings of perceived stimuli
3. Impression of stimulating person – attribution of personality traits, current feelings about the perceived person.

The aforementioned authors refer to three factors that shape our daily formation of impressions of other persons:

1. amount of achievable information about the perceived person
2. range of interactions between the sentient and the perceived

---

[13] Nakonečný, M. (2009). *Sociální psychologie.* Praha: Academia.

[14] Tavel, P. (2012). *Chyby a nepresnosti pri hodnotení ľudí.* In: Ostium, 1, vol. 8. (http://www.ostium.sk/index.php?mod=magazine&act=show&aid=244)

3. level of well-established relationship between the sentient and the perceived

The first two factors are closely related – the more time people spend together, the better they get to know each other; the third factor often causes distortion of impressions – nice individuals are being overestimated and vice versa[15].

It turns out that the accuracy of perceptual judgments is an extremely important element in our daily lives. An individual who is consistently unsuccessful in understanding others, has a difficult life (some authors see the inability as a cause of many mental illnesses). The psychologists include among the main characteristics of personality involved in interpersonal cognition: social sensitivity and social intelligence. Differences in the perception of others tend to be particularly influenced by:age (age significantly influences the selection of features that we notice about other people); intersexual differences (women formulate different types of judgments from men; women are more conscious of personal characteristics, men perceive the social role of assessed individuals more); personality traits (they play a special role in social perception); on the part of the assessor the most important include: empathy, authoritarianism, hostility or aggression, cognitive complexity, sociability; intelligence (there is a link between intelligence and perceptual validity of judgments); the influence of professional orientation (e.g.: after long-term practice of certain profession we attain "professional bias"; especially interesting are the professions where the entity is specially trained in social perception); the amount of information that we have about the other person.

The Mechanisms Leading to Distorted Perceptions and Assessment of Others:

- HALO-EFFECT: effect of the first impression; a systematic error in the evaluation of humans, in which a single personality trait is so prominent that the other traits are pushed into the background;
- PROJECTIONS: ascription of personal motives, strengths and patterns of behaviour to others;
- EFFECT OF NOVELTY: the latest information has more significant impact on the formation of impressions of others;

---

[15] Nakonečný, M. (2009). *Sociální psychologie.* Praha: Academia.

- PROMINENCE IN BEHAVIOUR: phenomenon of apparent correlation – inclination to ascribe a prominent behaviour to a person of physical prominence;
- POPULARITY OR UNPOPULARITY OF THE PERCEIVED: (popularity evokes an inclination to a more positive perception, while the opposite trend raises unpopularity);
- EFFECT OF LENIENCY: positive assessment of people to whom we have a positive attitude (especially parents);
- ERROR OF CENTRAL TENDENCY: tendency to avoid extreme evaluation; qualities of others are perceived in central neutral zone, fear of extreme evaluation;
- ERROR OF CONTRAST: alien qualities are perceived preferentially;
- PREJUDICES: qualities are ascribed to a person based on his/her affiliation to specific social group.

The term halo effect was used for the first time by E. L. Thorndike. The term is derived from the word "halo", which means nimbus, glory, aureole, therefore sometimes so-called "halo effect" is being used as a term. G. Nawratil and B. Rabaioli-Fischer describe the following example: imagine that you would see a person coming with a halo on the street, and only on this basis, you should compile a list of his qualities[16].

Prejudice could be simply characterized as previously formed attitude towards an object, which is manifested regardless of the individuality of the object. An individual does not consider the bias in the assessment of other persons any alternative explanation; the results of his interpersonal evaluations are predetermined[17].

Prejudice does not have to be only negative. For example, philosopher H.G. Gadamer spoke of so-called positive prejudice – tradition, into which all people are born. Prejudice, pre-understanding is a precondition for the possibility of any understanding – understanding cannot be exempted from assumptions. A crucial part of understanding is the tradition – a person who seeks to understand the subject is linked to the object of understanding by language, which in turn is linked with tradition[18].

[16] Tavel, P. (2012). *Chyby a nepresnosti pri hodnotení ľudí.* In: Ostium, 1, vol. 8. (http://www.ostium.sk/index.php?mod=magazine&act=show&aid=244)
[17] Hayes, N. (1998). *Základy sociální psychologie.* Praha: Portál.
[18] Gadamer, H. G. (2010). *Pravda a metoda.* Praha: Triáda.

Negative prejudice, which is often present in the society and has very nega-tive consequences, is ethnic prejudice. K. Lorenz saw the aggression associated with prejudice as a basic and necessary instinct related to the protection of the territory[19]. Likewise, prejudice against outside groups was interpreted by Dawkins[20], who argued that prejudice is a kind of biological instinct that compels individuals to protect fellows that share their genes. This theory also corresponds with the concept of prejudice in socio-biology by E.O.Wilson, which he based on the observation of ant communities. Wilson emphasized the importance of kin selection over other factors[21].

Philosopher and sociologist T.W. Adorno argued that the basis of prejudice is a certain type of personality and upbringing that makes certain types of people to incline towards prejudice against outside groups. These are primarily people whose parents enforce harsh discipline that caused the child's aggressive feelings, but it also did not allow the child to express these feelings so they internalized and were expressed towards other objects. Particularly against minority groups and all people who are socially different.

In addition to ethnic prejudice, psychologists, anthropologists, aesthetes and other scholars are interested in prejudice related to the perception of attractive and unattractive people. With the ancient platonic identification of beauty and goodness the positive qualities are generally ascribed to more attractive people. So, attractive people are, in a sense, born under a lucky star. They suffer less from loneliness, social anxiety and embarrassment in the public. We like to help them; we like to relate to them; it even turned out that it is more difficult to lie to an attractive person than to a less attractive one. Attractive individuals are preferred as sexual partners; they have a more positive opinion about themselves in terms of skills and mental health. Attractiveness is often connected with intelligence – attractive individuals are perceived to be more intelligent, etc.[22]

After interdisciplinary analyses of society and socialization, let me focus on biological determinism of sex, gender and identity: Most people ordi-narily seem to think that sex and gender are coextensive: women are human

[19] Lorenz, K. (2003). *Takzvané zlo.* Praha: Academia.
[20] Dawkins, R. (1989). *The Selfish Gene.* Oxford Paperbacks.
[21] Hayes, N. (1998). *Základy sociální psychologie.* Praha: Portál.
[22] Blažek, V., Trnka, R. (2009). *Lidský obličej.* Praha: Karolinum.

females, men are human males. Many feminists have historically disagreed and have endorsed the sex/ gender distinction. Provisionally: 'sex' denotes human females and males depending on *biological* features (chromosomes, sex organs, hormones and other physical features); 'gender' denotes women and men depending on *social* factors (social role, position, behaviour or identity). The main feminist motivation for making this distinction was to counter biological determinism or the view that biology is destiny. A typical example of a biological determinist view is that of Geddes and Thompson who argued that social, psychological and behavioural traits were caused by metabolic state. Women supposedly conserve energy and this makes them passive, conservative, sluggish, stable and uninterested in politics. Men expend their surplus energy and this makes them eager, energetic, passionate, variable and, thereby, interested in political and social matters[23]. These biological 'facts' about metabolic states were used not only to explain behavioural differences between women and men but also to justify what our social and political arrangements ought to be. Simone de Beauvoir famously claimed that one is not born, but rather *becomes* a woman, and that "social discrimination produces in women moral and intellectual effects so profound that they appear to be caused by nature"[24]. Commonly observed behavioural traits associated with women and men, then, are not caused by anatomy or chromosomes. Rather, they are culturally learned or acquired. In order to distinguish biological differences from social/psychological ones and to talk about the latter, feminists appropriated the term 'gender'. Psychologists writing on transsexuality were the first to employ gender terminology in this sense. Until the 1960s, 'gender' was often used to refer to masculine and feminine words, like *le* and *la* in French[25]. One way to interpret Beauvoir's claim that one is not born but rather becomes a woman is to take it as a claim about gender socialisation: females become women through a process whereby they acquire feminine traits and learn feminine behaviour. Masculinity and femininity are thought to be products of nurture or how individuals are brought up. They are *causally constructed* . AI has a significant and profound impact on the way that people are perceived and treated in society. Yet, the design and implementation

---

[23] https://plato.stanford.edu/entries/feminism-gender/
[24] Beauvoir, S. (2010). de: Beauvoir, S. de.: *The Second Sex.* Vintage Books, New York.
[25] https://plato.stanford.edu/entries/feminism-gender/

of AI perpetuates a vicious cycle. The technology captures and reproduces controlling and restrictive conceptions of gender and race which are then repetitively reinforced: Gender relations can be thought of as materialised in technology, and masculinity and femininity in turn acquire their meaning and character through their enrolment in working machines. In light of the problems outlined above, some aspects of technology in particular need to be challenged: The reproduction of stereotypical gendered bodies in robotics. Currently, there is a disjunction between the theoretical 'suspension of gender' which could be promising for destroying restrictive gender stereotypes, and the encapsulation of physical gender stereotypes in technology and robotics. The use of language, interaction and communication in relation to these systems. This refers to both the use of gendered 'voices' and 'responses' which these systems produce, in addition to the use of gendered pronouns and syntax[26]. What is gender? A simple question with no single answer. Gender can be understood through a multitude of perspectives: a subjectively held self-identity, a self-presentation to others, a social construct defined and maintained through performative acts, and a demographic imposed by society. In the context of computer vision, we may ask how the design and use of facial analysis and image labeling systems collapse these perspectives into a singular worldview: presentation equals gender. Forms of self-presentation are encoded into computational models used to classify these presentations. When classifying gender, designers of the systems in various contemporary studies and research chose to use only two predefined demographic gender categories: male and female. As a result, these presentations are recorded, measured, classified, labeled, and databased for future iterations of binary gender classification. These gender classification models are then bundled up for commercial use, often in the form of cloud-based services, providing an infrastructure that third-parties can use to create or augment their own services. In the process, these services propagate a reductionist view of gender provided by the underlying infrastructure. Self-identity is not used by computer vision systems. After all, it cannot be seen. As these societal changes proliferate, AI-driven conclusions have become more than a gender identity concern. Some AI experts and members of the transgender community are

---

[26] https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00011/full

worried about the potential for serious repercussions if gender recognition, as it exists today, is put to use for more complicated and sensitive tasks, whether it be using AI to help screen job candidates or nab criminal suspects. People are beginning to accept that gender may not be split between man and woman. I will finish my consideration with quotation from interview with robotic and AI scientists: "What we've continued to discover is that people identify their gender in different ways because gender, unlike sexual biological characteristics, is socially constructed"[27].

27 https://towardsdatascience.com/why-we-really-need-to-be-thinking-about-ai-and-gender-e2f96219f61c

References

Barrett L., Dunbar R., *Lycett, J.* (2007). *Evoluční psychologie* člověka. Praha: Portál.

Beauvoir, S. (2010). *de: Beauvoir, S.* de.: *The Second Sex.* Vintage Books, New York.

Blažek, V., Trnka, R. *(2009). Lidský obličej.* Praha: Karolinum.

Dawkins, R. (1989). *The Selfish Gene.* Oxford Paperbacks.

Gadamer, H. G. *(2010) Pravda a metoda.* Praha: Triáda.

Hayes, N. (1998). *Základy sociální psychologie.* Praha: Portál.

Heinesohn, R., Packer, C. *(1995). Complex Cooperative Strategies in Group – Territorial African Lions*. In: Science, 269, p.1260-1262.

Koukolík, F. (1997). *Mravenec a vesmír–O hvězdách, atomech, životě a vědcích.* Vyšehrad.

Lorenz, K. (2003). *Takzvané zlo.* Praha: Academia.

Nakonečný, M. (2009). *Sociální psychologie.* Praha: Academia.

Tavel, P. (2012). *Chyby a nepresnosti pri hodnotení ľudí.* In: Ostium, 1, vol. 8.