

# Final Report

17-117-S "Intrapersonal conflicts of values: scientific and philosophical understanding of the present self's concern for her future self "

CHF 178'400.-

Dr. Eleonora Viganò, University of Zurich, Institute of Biomedical Ethics and History of Medicine

## *Project topic*

The project deals with the **relationship in which the individual stands with her future self**, with a focus on moral intertemporal choices or diachronic intrapersonal conflicts, namely choices in which what we consider the right thing to do at present opposes that which in the future we may consider the right thing to do (for instance, career choices or advance healthcare directives (ADs)). Moral intertemporal choices are a subspecies of intertemporal choices (ICs), which are decisions entailing a trade-off related to the magnitude of an outcome and the time at which it is received. The typical behavior observed in ICs is temporal discounting (TD), i.e. the devaluing of the reward as a function of its delayed nature, which makes people prefer the short-term, smaller gratification over the long-term, larger one.

## *Project description and results*

The project is divided into two parts, one is descriptive and the other is normative.

The descriptive part concerns the neurobehavioral studies investigating **how individuals perceive their future selves in ICs**, the **descriptive theories on the subjective perception of the future self**, and the delineation of a **model of agent taking moral ICs**.

Several studies at behavioral and neural levels showed that in ICs, *individuals tend to perceive their future selves as another person*, i.e. as distinct from them; the more the individual perceives her future self as different from her, the steeper her TD rate is. *The way the individual mentally represents her future self influences ICs*.

Three descriptive theories have been proposed to explain the connection between TD and the mental representation of the future self. Each theory attributes TD to a different aspect of the perception of the future self: psychological continuity with the future self (in the future-self continuity theory), mental simulation, which is the adoption of her perspective (in the mental simulation theory), and mental construal of the future self (in the Construal Level Theory). I conducted the first review on these theories and identified their interrelations. *I found that construal level of the future self is a component of mental simulation; future-self continuity can result from mental simulation and affects the likelihood that simulation of the future self occurs, as well as the construal level of objects; and the latter indirectly affects future-self continuity through mental simulation*.

*I outlined a model of the relevant entity involved in the present-self/future-self relationship, namely the agent or self*, which is empirically plausible, i.e. compatible with the findings on the perception of the future self. This agent *necessarily extends over a certain period of time* – otherwise, her plans and commitments were not possible – *but she can last less than a person's life and her actions are based on a hierarchical set of normative principles of action* (e.g. values, commitments). This model can explain, in case of transformative experiences, how a distinct entity, i.e. the agent, may emerge from a physically identical body, i.e. the person, and how a person can have different selves throughout her life.

The normative part of the project regards the development of a correct – or **moral** – **relationship between the present and the future self** and consisted in the **analysis of the morally relevant characteristics of this relationship**, the **elaboration of a normative theory of moral ICs** or diachronic interpersonal conflicts, the **application of this theory to a kind of moral ICs**, i.e. **ADs**, and the **theoretical design of a decision-support tool for moral ICs** that enhances people's ability to adopt their future selves' perspective.

The individual's tendency to perceive her future self as another person, the possibility that moral ICs give rise to a transformative experience, and my model of the agent are at the basis of *my interpretation of moral ICs as interpersonal conflicts between two agents existing at different times*. Furthermore, as the *present-self/future-self relationship* have a structure of power, freedom, and knowledge similar to the contemporary people/future generations relationship, I considered the first relation *as a special case of intergenerational ethics*. *The morally relevant characteristics of the present-self/future-self relationship are the asymmetry of decisional power between the present and the future selves, the indeterminacy of the future self's identity and existence, the present self's ignorance of the future self's identity and existence, the future self's veto power*

against the present self's long-term plans, and *the strong causal connection between the present self and the future self*.

*The normative theory of moral ICs* derives from the features of the present-self/future-self relationship. It is the set of moral requirements that establish the correct relationship between the present and the future selves. This theory *attributes moral responsibility for the future self to the present self*, even when they are two numerically distinct selves. Also, it *protects the future self's vulnerability* due to her temporal location *through the right to an open present*, which is the right to pursue one's normative principles of action. The corresponding duty of this right is the present self's duty of preserving the future self's agency, namely not restricting the future self's freedom as a practical agent. Finally, this theory *justifies the future self's use of veto power in case the latter is different from the earlier self*.

Once the patient with dementia is assessed as partially or not competent, she cannot exercise her veto power on her earlier self's decisions about ADs. *I divided the progression of Alzheimer's disease into three morally relevant phases*, on the basis of the kind of autonomy that the Alzheimer patient retains: presence of source and planning autonomy (phase 1), presence of source autonomy (phase 2), and absence of both forms of autonomy (phase 3). I argued that *as long as the patient possesses source autonomy* (phases 1-2), which consists in having normative principles of actions, *ADs can be analyzed as interpersonal conflicts* between the pre-dementia self and the post-dementia self, as both are agents. Therefore, *the normative theory of moral ICs applies to phases 1-2*, albeit with a further distinction, as in phase 2 the patient can decide her present but not her future medical treatments. Hence, *ADs about dementia should be adapted to the moral characteristics of each phase* of the disease.

An individual tends to identify with her future self when the biases of mental simulation that make the individual perceive the future self as another person are reduced. *Virtual reality (VR) enhances the user's ability to mentally simulate the future self* by increasing the vividness of the simulation of the future self and presenting a low-level mental construal of the future. In this way, the individual's perceived distance or difference from her future self is reduced. In collaboration with Tyebally Fang and Spitale, *I provided the features and recommendations for designing a VR decision-aid suitable for a kind of moral ICs faced by prospective mothers: prenatal testing*.

#### ***Major achievements of the project till August 2020***

- Book contract with Routledge for a monograph entitled "*Moral Choices for Our Future Selves: A Moral and Empirical Theory of Prudence*" to be published in the series Routledge Focus on Philosophy.
- Published article: "Moral and Vocational Dilemmas Meet the Common Currency Hypothesis: a Contribution to Value Commensurability", *Review of Philosophy and Psychology*, 11 (2020), with E. Lombardi Vallauri (second author).
- Published article: "In AI we trust incrementally. A multi-layer model of trust to analyze human-artificial intelligence interactions", *Philosophy & Technology*, 33(3) (2020), with A. Ferrario and M. Loi (first and second authors).
- Published article: "The individual's later self is less autonomous and a stranger: the impact of time in advance directives", *Bioethica Forum*, vol. 11 (4) (2018).