Modelling the Social Contract: An agent-based Model of Hobbes' Contract Theory

Pierucci, Federico; Kruk, Natalia; Cecconi, Federico

Preprint / Preprint
Arbeitspapier / working paper

Modelling the Social Contract An agent-based Model of Hobbes' Contract Theory

Federico Pierucci, pieruccifederico@gmail.com
Natalia Kruk, nd.kruk@student.uw.edu.pl
Federico Cecconi[1], federico.cecconi.istc.cnr.it

ABSTRACT: The aim of this paper is to build a computational model that presents the effects of social dynamics such as evolution on populations applying the theory of Hobbesian Social Contract, social learning and norm diffusion. The phenomenon we are studying is the so-called tragedy of the commons, in which individual agents, having open access to a resource unconstrained by common social structures, act according to their own self-interest, seeking to maximise their own profits. Developing the theoretical framework and agent-based model, we applied to our artificial environment the norm associated with altruism, which modifies agents' behaviour during the simulation, which in turn affects the distribution of wealth. Through the behavioural space, we show that under certain circumstances specified in the parameter, it is possible to obtain a social contract and, as a result, a state of equilibrium. We show that individuals who have obtained the norm are able to enter into a contract, resulting in a rising wealth of the population and a more equilibrium distribution, while if they do not, more inequalities emerge. However, our model is a simplification of Hobbes' theory, admittedly, our agents can spontaneously establish cooperation but there are no complex structures, such as psychological ones, or moral cognition. We believe that this is a skeletal description of the Hobbesian social contract, in which self-interested individuals without obligation to cooperate agree to abide by a norm and its benefits. Depending on how profitable cooperation is (due to the redistribution) and the number of altruistic agents, the community members work for the common good.

Introduction

From the earliest models of segregation (Schelling, 1971) to the first iterated prisoner's dilemma simulations (Axelrod, 2006), to the more complex simulations of state warfare and artificial societies (Epstein and Axtell 1996), agent-based modelling has been applied a wide variety of fields of questions. Having observed the success of agent-based model simulations in the broad area of social sciences (Gilbert, Troitzsch 2005), we believe that Agent-Based Modelling is a valuable tool to produce new strands of research, in which philosophical investigations would benefit from borrowing methods and strengths from other disciplines. In fact, while the paradigm of experimental philosophy has gathered significant interest from the scientific community, its application has been mostly reserved for the problem of moral philosophy. On top of that, even if scholars have applied game theory and rational choice theory to many problems of political philosophy there is still a gap in the application of quantitative methods to political problems. The goal of this article is to offer a proposal, and an initial implementation, that begins to fill this gap. Hence, the purpose of this article is to conduct a first investigation of a canonical problem in political philosophy, but from a previously unexplored angle. In creating our computational model, we followed and modified the model of Thomas Hobbes' famous social contract as he describes it in his "Leviathan." This allows us to open the field to a larger research question: how can we experimentally evaluate if a model for a social contract, as intuitively persuasive, internally consistent, and formally correct as it may seem, is really well-founded? Can the expected

[1] This technical report was produced as an activity of the Labss, The Laboratory of Agent-Based Social Simulation (LABSS) is an international research group working at the intersection of cognitive, social, and computational sciences (https://labss.istc. cnr.it). The work is currently undergoing evaluation and review

results in terms of the emergence of cooperation and equality, which are assumed to be true in the model, be proven empirically? Therefore, we take a broader view of this question: given a set of different scenario states, it can be shown whether the predicted cooperative equilibrium (social contract), directly follows from the hypotheses that describe the initial conditions (and whether it follows in a deterministic fashion, or a stochastic one). We believe that this is achievable by implementing an agent-based modelling framework and building real-world simulations of it, examining the emergent cooperative behaviour in those artificial societies (Epstein 1994) that are described in the model, and extracting empirical data by running said simulations. The simulation environment will be treated as an experimental space in which falsifiable propositions about the phenomena in question will be extracted and will be valuable for the empirical investigation of a specific contractual scenario.

The **state of the art** research for the computational tractability of the contractarian position is yet to be explored, having only one instance of it developed - precisely, on Hobbes's contract (Long 2019). However, there is vast research on Agent-Based Models that accounts for the formation and diffusion of social norms, in which we can find different types of cognitive architectures, such as BOID (Broersen et al. 2001, 2002), the EDA model (Felipe & Liu 2000) or the EMIL project (Andrighetto et al. 2007, 2010a, 2010b, 2013); different types of modelling frameworks such as OPERA (Dignum 2004), ISLANDER (Esteva et al., 2002), OperettA (Aldewereld, Dignumand 2010) and, recently, the NorMAS-ML (Freire et al., 2019); also, computational platforms such as AMELI (Esteva et al., 2004). In more recent times, the normative apparatus of NORMAS (Normative Multi-agent system) has been extensively reviewed (Mahoud et al. 2014) in its theoretical and computational aspects. We can also observe a constant stream of new models, such, only to name a few, research on designing norms and analyzing group dynamics (Aldewereled et al. 2016), on the norm-diffusion mechanism from sanctioning systems (Vinitsky et al. 2021), on the formation of coalitions in cooperative-games settings (Vernon-Bido, Collins 2020), and on the synthesis and resolution of normative conflicts (Riad e Golpayegani, 2021). All those resources have been produced to describe and explain how a set of independent agents, in an artificial environment, can build, recognize, internalize and enforce a set of norms and values.

On top of that, there is a solid amount of literature that allows for a formal treatment of the problems of social contracts thanks to the theoretical and mathematical tools of game theory, such as Gauthier (1986) and Binmore (1994, 1998). Especially for Hobbes's social contract, there has been by far the most consistent research endeavour, representing the "war of all against all" by the prisoner dilemma (Gauthier 1969, Rawls 1971, McLean 1981, Kavka 1983), the assurance game (Gauthier 1994, Shaver 1990, Moheler 2009) and the assurance dilemma (Kavka 1986).

Theoretical Prerequisites

In order to shape such a class of simulations, with respect to those constraints, a set of **theoretical prerequisites** are mandatory:

- A theory that accounts for the generation and production of moral statements from a set of universal primitives needs to be implemented. The *Universal Moral Grammar* project (UGM), by Mikhail (2007), is best suited for reaching the first desired condition: from a set of deontic primitives and syntactical structures, - as in Chomsky's Generative Grammar (Chomsky 1957, 1965) - every human agent can show complex moral cognition and is capable of producing moral sentences and judgements. UGM optimally fits the needs of this research program, making it possible to build an artificial agent that is capable of autonomous moral evaluation, given that s/he has access to those primitives and to a set of rules for producing well-structured moral statements.

- A structure that allows for a complete and exhaustive analysis of the pre-contractual condition (the so-called "State of Nature"). *Behavioural Game theory* is a well-supported and empirically grounded framework that makes it possible to represent the "state of nature" as a set of strategic interactions in which agents operate (Gauthier 1993). The most salient feature is the framing of the contract as a "*Public Good Game*"**, a social dilemma in which the formation of the Contract or the contribution to the public good, advantages collectively all agents,** but freeriding is, in a self-interested perspective, the most rational choice to take from an individual point of view. Empirical research in Behavioural Game Theory permits to building scenarios and rationality profiles that are compatible with the variety of behaviours that human beings show in such games.

· A theory for the structure of norms as a set of emerging proprieties in the context of the state of nature. Bicchieri's theory of norms (2006) will be used. Here, norms are **defined as a mechanism that bridges individual expectations and collective behaviours,** producing an equilibrium state once normative and empirical expectations are matched in a group of human agents, thanks to the presence of a normative system. This allows us to formally describe – and to model – the social contract as a game in which a macro-equilibrium (statical or dynamical, deterministically or stochastically gained that it may be) is reached between all the agents (Skryms 1996), and that changes the structure of the original public good game, making less rational, from a consequentialist perspective, any type of freeriding behaviour. The author presents also a cognitive model (Bicchieri 2006, 32) for the reasoning behind norm selection and recognition, making her theory very well suited for being formally implemented into ABM simulations without losing empirical grounding.

· A theory for the procedures that account for the internalization of a norm system, say, the capacity of the agent to consider a norm as a cognitive token that modifies agent's behaviour. From the research of Conte and Castelfranchi (1996) to more recent literature on the process of *immergence (*Andrighetto & Conte 2012), the possibility of constructing a normative agent that reasons not only through the standard utility function but also by the adjunct possibilities offered by a cognitive architecture that uses norms as an element of the cognition processes is well explored. In a feedback loop in which norms are present in a society, and cannot be completely reduced by their descriptive elements (emergence), *immergence* accounts for the process of internalising them, and operate as **a** ***proximate cause*** **of the action** (Andrighetto, Conte et al. 2013). **and accepted from a given "natural" condition, one fundamental need for a framework that allows for testing them is to have a decentralized system of heterogeneous and autonomous agents that, formerly independent and unbounded from the contractual obligation, are equipped with some form of rationality and expectations, decide whether or not to join given contract, to negotiate it, to partake in the formation of it or to rely on purely selfish means of existence, without being the case of any collective computation that guide his/her behaviours toward the desired end**. That is to say, the social contract needs to be treated as an ***emergent phenomenon*** which cannot be described using the elements that describe the system. Agent-Based Social Simulation, as various models of social phenomena show, suits the requirement of this condition, offering well-tested methods and design procedures that grant this autonomous agent capacity of computation.

The expected result would be a refined **theoretical framework** that could be applied to different contract theories, making them *controllable* and *falsifiable***.** On top of that, a **detailed simulation model** of what emerged from the in-lab experiments will be provided, to generalize the results and to control parameters and initial conditions. Those will be paired with extensive documentation and with a set of results, with different parametrization, that would allow us to better grasp the internal dynamics and outcomes of the aforementioned theory. The presence of documentation will allow the non-expert in computational methods to understand the actual process without the formalism that simulation models are imbued with.

For the scope of this paper, the simulations will be conducted in the **NetLogo** language and programming environment.

General Theory

We describe here the **meta-theory** of social contracts, declaring the necessary elements that, for every single contract, permit the construction of the above-mentioned model. This will bridge the theoretical section of the project with the operational one, building an actual working framework that can be used with different contracts and different programming environments, allowing for variations in the replicability and scalability of the model.

This meta-theory will be defined as "*Type-Contract-Theory*". It states what are the fundamental proprieties that agents, the set of interactions between them, and the world where the model takes place, that a computational model of a social contract need to have to allow the possibility of an instantiation of it by artificial means, thanks to the use of ABM. Those will be then considered as the variables that will consider in the actual series of simulations.

In order to offer a clear, yet simplified, formal definition of a social contract theory, we define that, in a general form, as an N-uple ∑ **= [***N*,*P*,*A***]** that yields

C

as the outcome. Every item in ∑ is a set of parameters that, under certain values, produce as consequence C= N***N***

(which is an adequate description of a contractual agreement) where:

N***N***

is a state of nature, a "no agreement position" (Gauthier 1969) or an "original position" (Rawls 1971). It represents the "natural" condition in which individuals live, or return if no social contract has been reached between the agents. It is defined by the lack of any form of mutual obligation on a set of social norms.

P***P***

is the first set of parameters

[P1, P2, P3 … Pn]

that define the **world proprieties** of a given state of nature. By world proprieties I mean the core features that represent the original condition, and that differentiate a single contract theory from the others. For instance, in Hobbes' theory, the original condition is considered to be one in which individuals have no access to personal property (Moloney 2011), while in Locke's theory there is a model of a market economy and monetary mechanism, albeit a rudimentary one (Locke 1696).

A

is a second set of parameters

[A1, A2, A3 … An]

that define the **agent proprieties**. By agent proprieties, we mean all of the defining qualities that represent them as being able to rationally deliberate, their disposition to cooperate, and their ability to think strategically. In general, we consider them as being a set of variables that will define the agent in every model, and be interpreted in every simulation model.

C is the end state that is produced by the contract when the agreement between agents is made or not. It represents the final state of a social system in which there is one group of agents that have agreed upon a set of fundamental norms.

Game Theory

Explanation of the model

The model that will be shown in this section aims to assess the emergence and diffusion of cooperation norm in a public good mechanism, implemented in a Sugar-Scape environment.

Extension of Sugarscape model towards simulation of the emergence of cooperation (social contract).

The artificial world used in this paper is an extension of the second model in the NetLogo Sugarscape [that] suite implements Epstein and Axtell's Sugarscape Constant Growback model (...). It simulates a population with limited, spatially-distributed resources available. It differs from Sugarscape 1 Immediate Growback in that the grow back of sugar is gradual rather than instantaneous. (Li & Wilensky 2009). Another important aspect of the model is the competitive nature of the agents. This is a reference to the game, where each seeks to maximize its profits, aiming for the source of the greatest wealth, changing its strategy to win- have the most sugar.

Thus using the core of Epstein and Axtell's model, we outlook on the emergence of cooperation (social contract) in our artificial society. We designed a simulation in which agents can decide to contribute to a central repository by donating part of their wealth and then studied which **set of parameters accounts for the formation of a group of cooperative agents**. As a consequence of this, the amount of wealth in the world increases, linearly, with the wealth the agents have saved in the repository, making cooperation a collectively beneficial endeavour. To address this problem, we have constructed a set of cognitively simplification agents who can observe and imitate each other, and who can gain norm that enforce contribution without any prior incentive or design.

Competitive nature of the agents

Methodology

The following sections provide information on the methodology, showing how a conceptual framework of the theory of social contract was developed into a computational model, which consists of agents (inhabitants), the environment (a two-dimensional grid) and the rules that govern the interaction of the agents with each other, as well as with the environment. (You can download detailed pseudocode, writing to federico.cecconi@istc.cnr.it).

Agents: they constitute rational, formerly independent, and unbounded from the contractual obligation individuals seeking to maximize their profits. The agents are given the ability to engage in activities such as collecting, storing and donating sugar, norm exchange (getting convinced), learning, moving and eating. In our simplification model, among agents there are no cognitive differences (in this case they are homogeneous), but because of other differentiating features, as various levels of their prosperities, they form a heterogenous society. Furthermore, unlike in the Hobbesian model, in our simulation, there is no aggression, nor force and no punishment- agents are guided only by profit and observation- their behaviour is dependent on external factors and norm emergence. For the whole simulation, there is a constant number of agents (500) - they don't hatch nor die.

- every agent has:
    - **sugar-** a metaphor for resources; agents collect, store and donate sugar during the simulation;
    - **metabolism** (random in the range 1-4)- how much sugar does the agent lose each tick;

    - **vision-** scope of the world seen by agents;
    - **vision-points-** scope of agent movement;
    - **altruism-** the agent is altruistic if its altruism > 0.75; it means that he makes a donation;
    - **donation** = sugar * donation amount
    - **donated?** (-1; 1); if the donation > 1;
    - **sugarHistory-** a list of collected sugar;
    - **sugarDifference** = sugar - donation;
    - **payoff** (-1; 0; -1) = the mean of sugarDifference; helps agents learn, as a result (**payoff result**), it increases or decreases altruism by 10%;
    - **intensity** (the range 0-1);
    - **norm** (0;1);

  - every agent does:
    - **eat** = set sugar (sugar - metabolism + psugar);
    - **move-** move turtles at patches without other agents in their vision;
    - **gets convinced** = if norm of agent < 0, and the importance of the other agent's norm > 0.8, then the first one gain the norm of the second one;
    - **learns** = if public goods > public goods threshold and redistribution is on, then learning procedure is possible, in effect it increases or decreases the level of altruism in the world (by 10%);
    - **set donated?-** if agent donated, set donated?=1; if not, set donated?=-1;
    - **set the-importance-of-the-norm;**
    - **set the-salience-of-the-norm;**
    - **increase-the-importance-of-the-norm** - it is possible if there was no donation done before;
    - **change-the-norm-salience;**

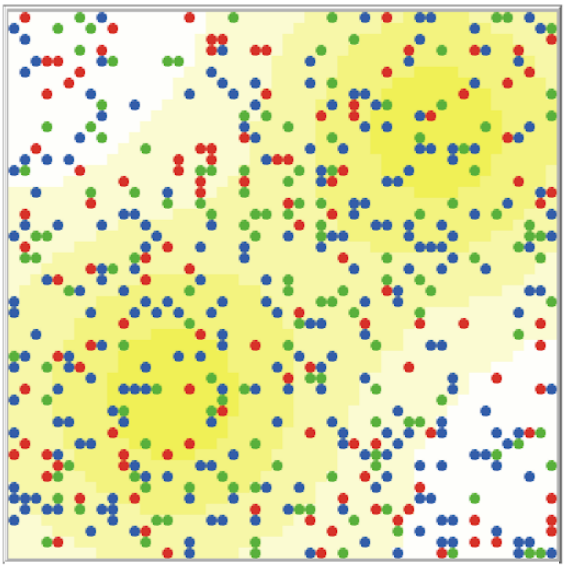

Figure 1 In order to facilitate simulation tracking, according to the agent's altruism level, the agents can occur in 3 different colours: red (non-altruistic), blue (neutral), green (altruistic).

- **Norm**: agents can't form the norm, it is "given", not "created". The norm operates as a ***proximate cause* of the action** (Andrighetto, Conte et al. 2013). **and can be accepted from a given "natural" contract [it is not mandatory], moreover, it can be partaken in the formation of it or relied on purely selfish means of existence, without being the case of**

any collective computation;

- norm has:
    - **The-importance-of-the-norm (0-1); -** it can decrease or increase during the simulation; it matters only if the turtle don't' have the norm; if the importance of the agent's norm is greater than 0.8, then agent gets the norm (0.8 is a simulation parameter);
    - **the-salience-of-the-norm**

- **Environment**: a two-dimensional 50×50 grid; grid cells (patches) have a sugar level, based on "sugar-map.txt" file; ; the spatial distribution of sugar varies: there are two clusters of maximum sugar capacity for cells (radial yellow), which gradually spread circularly to the point of sugarless cells (white ones); colour is changing proportional to the capacity of the cell.

- **Patches**: grid cells;

    - patch has:
        - **psugar -** the amount of sugar on the patch;
        - **max-psugar-** the maximum possible amount of sugar on the specific patch;

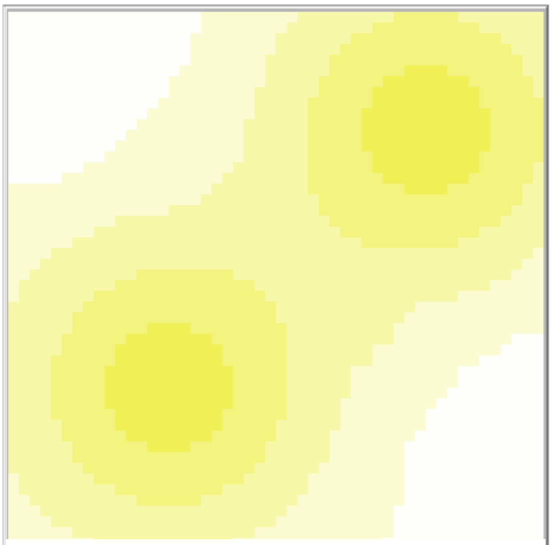

Figure 2. Spatial distribution of sugar capacities in the Sugarscape. Cells are coloured according to their sugar capacity: sugarless cells are white, whereas cells with maximum capacity contain a yellow circle; colour is changing proportional to the psugar in the cell.

- **Cooperation** is the result of the social contract, emerging under the specific conditions, related to the presence of the norm, sugar, altruism, donation and public goods threshold;
- **Information Management Activities**: collecting, storing, donating sugar; gaining the norm; convincing;

Outcomes

Our simulation showed how different strengths of the norm influence the level of altruism, which is related to the donation and distribution of wealth in the population. Agents in our model have the capability to learn by observing [altruistic] acts of other agents in the world. Also, we observed how the public goods threshold- the number of sugar in the world impacts the number of normative and altruistic agents and its dynamic.

To do this study, we used NetLogo Behaviour Space, trying various values of public goods thresholds (publicgoods_th) and norm effect (norm_effect). To generate the data, we experimented with 100 executions, stopping the simulation at 500 ticks. We then graphed the data, evaluating the norm strength effect by examining how it influenced the level of sugar, altruism and the distribution of wealth (sugar), according to the public goods threshold.

To assess the impact of the norm on the population, we first performed a test with mechanism, in which agents could only adapt their actions by observing the behaviour of others.

To generate the data, we experimented with two executions of 20 and 5 repetitions, stopping the simulation at 500 ticks. We then graphed the data, evaluating:

- **run number**
- **altruism-modality**
- **publicgoods_th**
- **count agents with [norm = 1]**
- **count agents with [altruism > 0.75]**
- **mean [sugar] of agents**
- **mean of [altruism] of agents**
- **list of [sugar] of agents**

Results

How does the strength of the norm impact both, the mean of altruism and mean of sugar of the whole population for the specific public goods threshold?

By pivot table, we obtained two histograms illustrating the influence of the norm's strength on both, the mean of altruism and mean of sugar of the whole population, in context of the specific public goods threshold. By overlying the results of all norm effects that we have examined (0.2; 0.25; 0.3; 0.35; 0.4), we obtained the most interesting results for the extreme values- for 0.2 and 0.4 parameter.

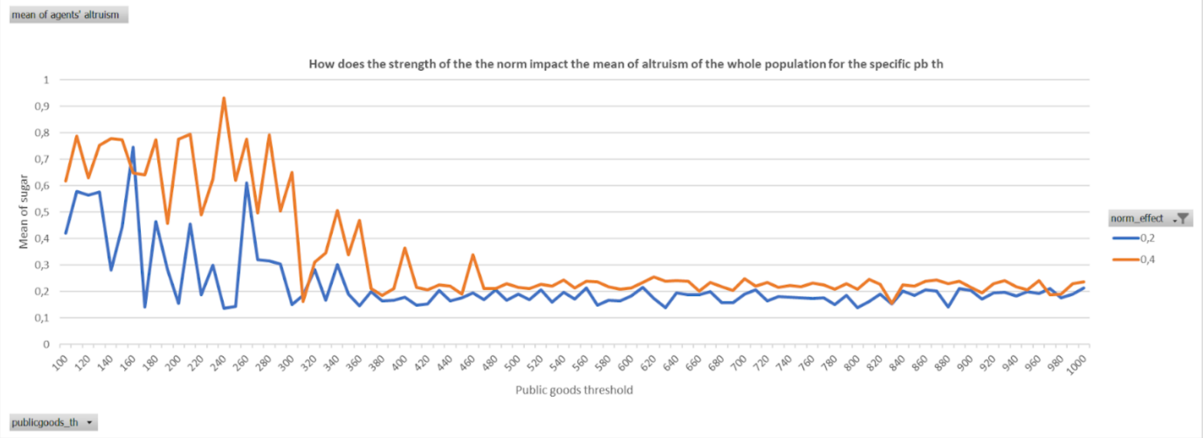

Figure 3 Line graph above shows the dynamic of the mean of altruism between norm effect 0.2 and 0.4- which are extreme values (the interval was 0.05).

According to our model parameters, altruism started from 0.75 (the maximum value is 1). On the X we have included a public goods threshold- from 100 to 1000 (with an interval of 10). Till public goods reach around 300, both norm effects strengths had very different dynamics. 0.4 (orange line) had a greater difference in intervals- the mean of altruism varied from 0.75 (lower threshold of altruism) to 0.1 (no-altruism), while 0.2 (blue line) varied from 0.6 to 0.45.

It looks like there were completely different dynamics of altruism dependent on the norm effect. The 0.2 norm effect has never reached the threshold for altruism (0.75) while the 0.4 norm effect has undergone “phase transitions” from altruism to the disappearance of altruism at around 300-400 of the public good.

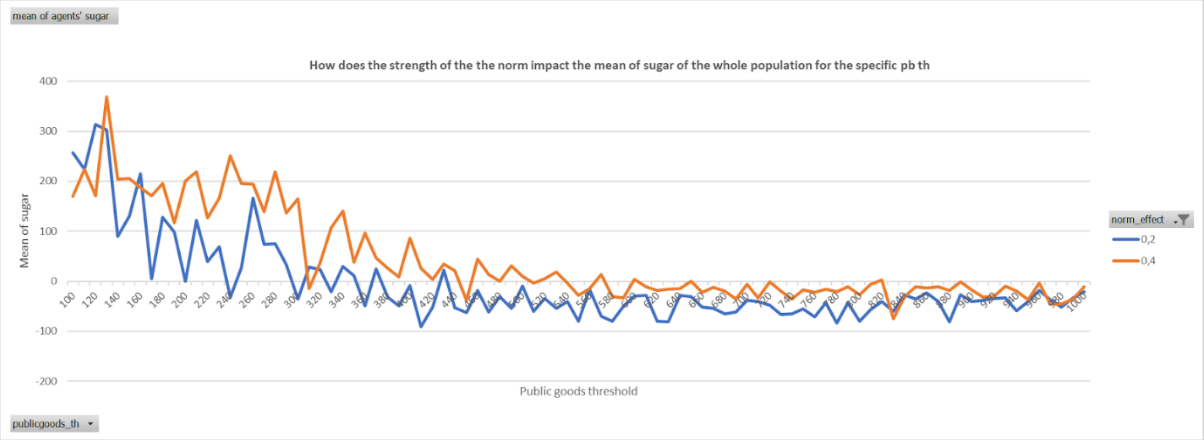

Figure 4 Line graph above shows the dynamic of the mean of sugar between norm effect 0.2 and 0.4

The 0.4 norm effect was at more or less higher level up of sugar till around 300 public goods parameter, and then dynamically fell to the level of 0. This sudden decline followed by short-term acute growth, and then gradual dwindling, has emerged for the same value of the public goods threshold as for the histogram illustrating mean of altruism for the same norm effect (0.4).

The 0.2 norm effect started at the same level of sugar as 0.4 but quickly had descended to the levels around 0. This suggested different characteristics of the dynamics of the sugar depending on the norm effect’s strength.

Wealth distribution for the specific public goods threshold

By comparing the results from different parameter combinations, we obtained 9 models (3 values for the public goods threshold and 3 of the strength of the norm). Interesting values were the outcomes of norm effect for 0.2, 0.3, 0.4; and for parameters of 100; 200; 300; 400 for the public goods threshold.

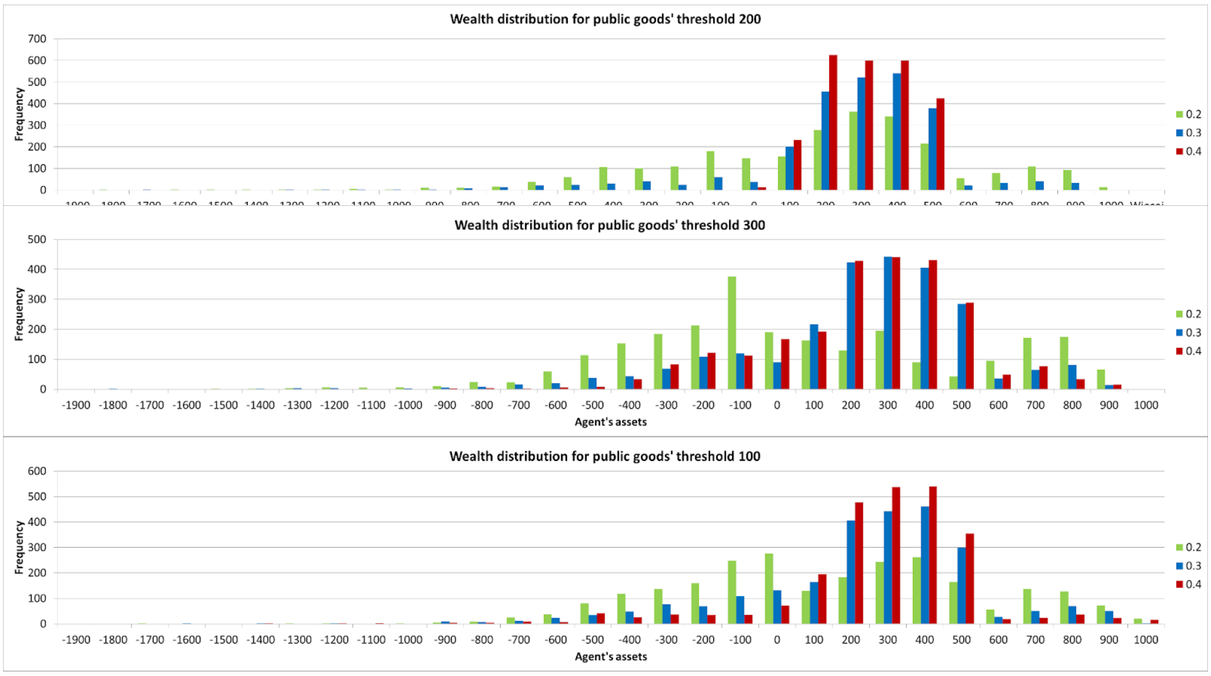

Figure 5 By overlaying distribution charts for each of the norm strength, we obtained the general image of wealth distribution for the specific public goods threshold.

Wealth distribution among normative vs non-normative populations for the specific norm effect and public goods threshold

Due to the focus on the normative aspect of our model, from the perspective of our study, the results of the simulation had two scenarios: the whole population could be either normative, which means that all the agents had a norm or non-normative - none of them had adopted the norm. At some threshold occurred rapid grow one of these groups, until they dominated the population.

By using SPSS, we obtained two graphs for each combination of parameters, which illustrated the wealth distribution for both, the normative and the non-normative population, which showed the necessary conditions for the equilibrium state to emerge. The most interesting results we obtained for 0.2 norm effect and 300 public goods and according to our data, the normative world occurred when the number of altruistic entities in the population was greater than 80% of the population.

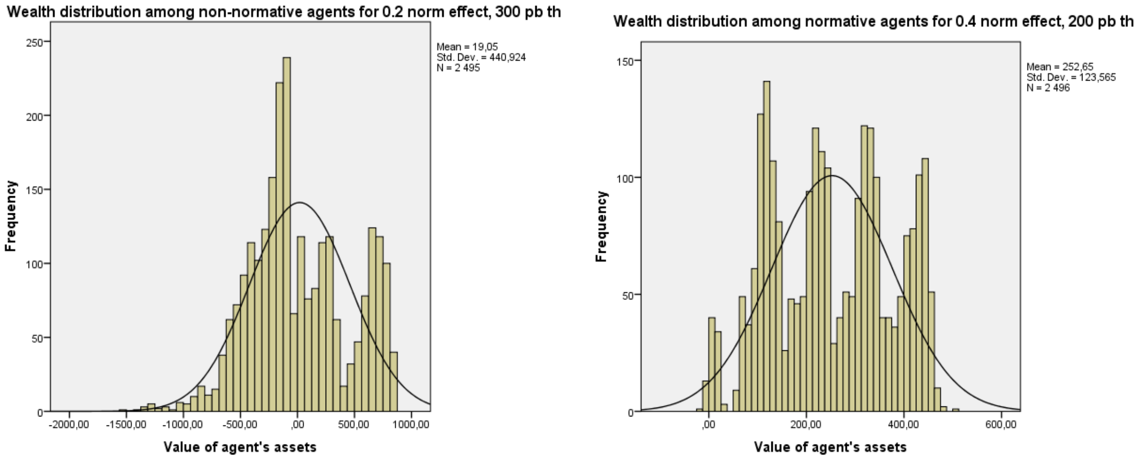

Figure 6 While every combination of norm effect and public goods threshold values had both histograms- distribution among normative vs non-normative agents, for parameters 0.2 norm effect and 300 public goods threshold and 0.4 norm effect, 200

public goods threshold, we obtained only single ones. We obtained, sequentially: all non-normative populations and only normative ones

Discussion

Our model presents the effects of social dynamics such as evolution (change of the strategy due to existing conditions) on populations. The phenomenon studied by us is the so-called tragedy of the commons. In economics and in an ecological context, this term means a situation in which individual users, who have open access to a resource unhampered by shared social structures, formal rules, charges, fees, or taxes that regulate access and use, act independently according to their own self-interest and, contrary to the common good of all users, causes depletion of the resource through their uncoordinated action in the case that there are too many users related to the available resources.

The best scenario for our model is the one, in which the majority of agents have more positive values of their resources (collected value/agent's assets), rather than values below 0 (then agents are "in debt"). In our model, we aim to establish the best combination of the tested parameters that will result in a state of equilibrium, i.e., a smaller variety of wealth distribution.

The most interesting results emerged for: 0.2 norm effect and 300 public goods threshold; and 0.4 norm effect and 200 public goods threshold. Both of those cases resulted in a single scenario, for the first one the outcome was the whole normative population, and for the second set of parameters- the non-normative one. In contrast, the other parameters' combinations resulted in two histograms- the first illustrated wealth distribution among the normative population; and the second among the non-normative population. Considering the results from all simulations with non-normative populations as the outcome, in each of them we obtained a much greater number of indebted agents and higher inequalities. However, in each simulation with a normative population, as a result, most of the agents had resources greater than 0. Also interesting is that the difference between the maximum and minimum values was always less than 600. This interpretation indicates that normative worlds are more favourable and the equilibrium state can emerge.

Significant inequalities in wealth distribution appeared for all the combinations of parameter 0.2 of the norm effect, especially for the threshold of 300 public goods, where there was also the largest number of negative values and the largest spread, in addition, it was the entire non-normative population.

Therefore, for the 0.4 norm effect and 200 public goods threshold, every agent had a norm. **It seems that for this set of tested parameters, a quasi-equilibrium state emerges. It means that if everyone adopts the norm, then the distribution is more clustered within the positive range of values and there are smaller inequalities among members of the population.**

For both, 0.3 and 0.4 norm strength, the number of wealth agents (with a positive number of assets) increases, which can be related to the fact, that most of the results of this combination of parameters were the normative populations. The effect of the norm with a parameter 0.3, for the thresholds of public goods at 100 resulted in 60% of the normative population; for 200 it was 80%, and for 300- **again 60% (also appeared a decline in the population wealth)**. For the 0.4 norm parameter and public goods threshold at 100, the outcome of all simulations was 80% of normative worlds; for the public goods threshold at 200 it was 100%; for 300 **- it was again lower at 60% (same as above)**. The distribution had a similar dynamic for those both sets of parameters and seemed to be very favourable due to the small number of indebted agents (those with resources below 0). Furthermore, most of the values for public goods threshold 200 and 400, were clustered between 200 and 500 [of collected values], which means that for these specific parameters, a state of quasi-equilibrium may emerge. However, the outcome is quite different for the threshold of 300, where the percent of normative populations as simulations results were smaller, moreover, 0.3 and 0.4 norm strength created debt more often in comparison to other wealth distribution histograms. Rapid grow until they dominate the population

Thus our model we outlined the change of the agents' behaviour in light of their experience (observation and redistribution), obtaining the answer what kind of behaviour can emerge when players can evolve- change their strategy. According to these results, it may be concluded, that the norm strength has an influence on the distribution of wealth. To put it briefly: The agents are pragmatic by seeking to maximalise their gains, they will tend to adopt the norm, because the normative population means a wealthy society with smaller inequalities.

Conclusions

In conclusion, having observed the success and the variety of agent-based models simulation conducted productively in the vast area of social sciences (Gilbert, Troitzsch 2005), Agent-Based Modelling is a precious and valuable way of producing a new field of research in which philosophical scrutiny would benefit from borrowing methods and strengths coming from other disciplines.

The ***impact*** of this research will be the next step in fostering cross-discipline analysis of philosophical problems. Norms are indispensable objects if we aim to understand how societies and individuals reason and act, both taken as individual elements and as a whole. The goal of this proposal is to shed light on what could seem, at a first glance, a quite abstract aspect of the vast amount of conceptual problems in political theory. However, offering a novel and empirically grounded model of political behaviour, it could also represent, outside of Academia, a valuable tool for understanding the decisions of different social actors, and could be considered an advancement in the modelling capabilities of voters, communities, institutions, taxpayers, and economic, political agents overall.

References

Aldewereld, H., & Dignum, V. (2011). OperettA: Organization-oriented development environment. In *Languages, Methodologies, and Development Tools for Multi-Agent Systems: Third International Workshop, LADS 2010, Lyon, France, August 30–September 1, 2010, Revised Selected Papers 3* (pp. 1-18). Springer Berlin Heidelberg.

Aldewereld, H., Dignum, V., & Vasconcelos, W. W. (2016). Group norms for multi-agent organisations. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, *11*(2), 1-31.

Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., & Villatoro, D. (2013). Punish and voice: punishment enhances cooperation when combined with norm-signalling. *PloS one*, *8*(6), e64941.

Andrighetto, G., Campenni, M., Cecconi, F., & Conte, R. (2010). The complex loop of norm emergence: A simulation model. In *Simulating Interacting Agents and Social Phenomena: The second world congress* (pp. 19-35). Springer Japan.

Andrighetto, G., Villatoro, D., & Conte, R. (2010). Norm internalization in artificial societies. *Ai Communications*, *23*(4), 325-339.

Andrighetto, Giulia, and Rosaria Conte. "Cognitive dynamics of norm compliance. From norm adoption to flexible automated conformity." *Artificial Intelligence and Law* 20 (2012): 359-381.

Bicchieri, C. (2006). The rules we live by.

Binmore, K. (1998). The complexity of cooperation. *Journal of Artificial Societies and Social Simulation*, *1*(1).

Binmore, K. G. (1994). *Game theory and the social contract: just playing* (Vol. 2). MIT press.

Boella, G., Verhagen, H., & van der Torre, L. (2007). 07122 Abstracts Collection--Normative Multi-agent Systems. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Bratman, M. (1984). Two faces of intention. *The Philosophical Review*, *93*(3), 375-405.

Broersen, J., Dastani, M., Hulstijn, J., & van der Torre, L. (2002). Goal generation in the BOID architecture. *Cognitive Science Quarterly*, *2*(3-4), 428-447.

Chomsky, N. (1957). Studies on semantics in generative grammar Mouton. *EUA: The Hague*.

Conte, R., & Castelfranchi, C. (1996). Simulating multi-agent interdependencies. A two-way approach to the micro-macro link. In *Social science microsimulation* (pp. 394-415). Springer Berlin Heidelberg.

Dignum, V., Dignum, F., & Meyer, J. J. (2004). An agent-mediated approach to the support of knowledge sharing in organizations. *The Knowledge Engineering Review*, *19*(2), 147-174.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American psychologist*, *49*(8), 709.

Esteva, M., De La Cruz, D., & Sierra, C. (2002, July). ISLANDER: an electronic institutions editor. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3* (pp. 1045-1052).

Esteva, M., Rosell, B., Rodriguez-Aguilar, J. A., & Arcos, J. L. (2004, July). AMELI: An agent-based middleware for electronic institutions. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 236-243).

Filipe, J., & Liu, K. (2000, June). The EDA model: An organizational semiotics perspective to norm-based agent design. In *Proc Agents' 2000 Workshop on Norms and Institutions in Multi-Agent Systems* (Vol. 200, No. 0).

Freire, E. S. S., Cortés, M. I., Júnior, R. M. D. R., Gonçalves, Ê. J. T., & De Lima, G. A. C. (2019). NorMAS-ML: Supporting the Modeling of Normative Multi-agent Systems. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *8*(4), 49-81.

Gauthier, D. (1986). Morals by Agreement. *Oxford: Clarendon Press.*

Gauthier, D. (1993). Constituting democracy. *The Idea of Democracy*, 314-34.

Gauthier, D. (1994). Assure and Threaten. *Ethics*, *104*(4), 690–721. http://www.jstor.org/stable/2382214
Gauthier, D. P. (1969). *The logic of Leviathan: the moral and political theory of Thomas Hobbes*. Oxford University Press.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

Grazzini, J., Richiardi, M. G., & Tsionas, M. (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, *77*, 26-47.

Honarvar, A. R., & Ghasem-Aghaee, N. (2009, December). An artificial neural network approach for creating an ethical artificial agent. In *2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation-(CIRA)* (pp. 290-295). IEEE.

Kavka, G. S. (1983). Hobbes's war of all against all. *Ethics*, *93*(2), 291-310.

Kavka, G. S. (1986). *Hobbesian moral and political theory* (Vol. 16). Princeton University Press.

LI, J. and WILENSKY, U. (2009), NetLogo Sugarscape 2 Constant Growback model, Northwestern University, Evanston, IL: Center for Connected Learning and Computer-Based Modeling.

Locke, J. (1696). Further Considerations Concerning Raising the Value of Money. London: Awnsham and John Churchill, Print.

Long, W. (2019). Escaping the State of Nature: A Hobbesian Approach to Cooperation in Multi-agent Reinforcement Learning. *arXiv preprint arXiv:1906.09874*.

Mahmoud, M. A., Ahmad, M. S., Mohd Yusoff, M. Z., & Mustapha, A. (2014). A review of norms and normative multiagent systems. *The Scientific World Journal*, *2014*.

McLean, I. (1981). The social contract in Leviathan and the Prisoner's Dilemma supergame. *Political Studies*, *29*(3), 339-351.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in cognitive sciences*, *11*(4), 143-152.

Moehler, M. (2009). Why Hobbes' state of nature is best modeled by an assurance game. *Utilitas*, *21*(3), 297-326.

Moloney, P. (2011). Hobbes, Savagery, and International Anarchy. *American Political Science Review*, *105*(1), 189-204. doi:10.1017/S0003055410000511

Rawls, J. (1971). A theory o f justice. *Cambridge: Harvard University Press.*

Rawls, J. (1972). A Theory of Justice (Cambridge, Mass., 1971). *RawlsA Theory of Justice1971*.

Riad, M., & Golpayegani, F. (2022, October). Run-time norms synthesis in multi-objective multi-agent systems. In *Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIV: International Workshop, COINE 2021, London, UK, May 3, 2021, Revised Selected Papers* (pp. 78-93). Cham: Springer International Publishing.

Sánchez-López, Y., & Cerezo, E. (2019). Designing emotional BDI agents: good practices and open questions. *The Knowledge Engineering Review*, *34*, e26.

Shaver, R. (1990). Leviathan, king of the proud. *Hobbes Studies*, *3*(1), 54-74.

Skyrms, B. (1990). *The dynamics of rational deliberation*. Harvard University Press.

Skyrms, B. (2014). *Evolution of the social contract*. Cambridge University Press.

Skyrms, B.(1996). *Evolution of the Social Contract*, Cambridge: Cambridge University Press.

Vernon-Bido, D., & Collins, A. J. (2020). Finding core members of cooperative games using agent-based modeling. *arXiv preprint arXiv:2009.00519*.

Vinitsky, E., Köster, R., Agapiou, J. P., Duéñez-Guzmán, E., Vezhnevets, A. S., & Leibo, J. Z. (2021). A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. *arXiv preprint arXiv:2106.09012*.