University of Trento

Doctoral School of Social Sciences

Doctoral Program in Economics and Management

Three Economic Extensions of John Rawls's Social Contract Theory: European Fiscal Union, Tax Compliance and Climate Change

a dissertation submitted in fulfilment of the requirements for the Doctoral Degree (Ph.D.) of the Doctoral Programme in Economics and Management

3rd June 2019

Ph.D. student:

Klaudijo Klaser

Supervisor:

Prof. Lorenzo Sacconi

Doctoral committee:

Dott. Stefano Castriota

Prof. Giacomo Degli Antoni

Prof. Roberto Tamborini

Table of Contents

1. General Introduction

"Every economic action takes place in the framework of a moral or ethics"

Jean-Jaques Laffont

"...economics needs to take account of the alternatives to utilitarianism that have been advanced in the past half century, such as the theory of justice of Rawls..."

Anthony Barnes Atkinson

In my thesis I apply the ethical model developed by John Rawls (1999) to three systems which have an economic dimension: European Union, tax compliance and environmental sustainability.

With this task my purpose is to answer to the following overarching research question: is an impartial and non-binding agreement, conceived in a Rawlsian frame, sufficient to generate fair and stable redistributive institutions? This general research question is then addressed and inflected according to the specific economic domains mentioned above and considered in detail through the next Chapters.

Given the diversification of the analysed topics, also the adopted research methodology turns out to be differentiated: the European distributive institutions are examined in depth by means of a theoretical-deductive approach which catches on Rawls's international social contract theory; the distributive issues linked to tax evasion and environmental sustainability are approached with tools belonging to behavioural and experimental economics.

The reminder of the Introduction is dedicated to outline the distributive issue within John Rawls's social contract theory, with the attention focused on international and experimental distributive justice. The intent is to move gradually from a perspective concerning the general theoretical frame to the specific issues constituting the research core of the thesis, that is the economic analysis in a Rawlsian perspective of the European integration (Chapter 2), of tax compliance (Chapter 3) and of pro-environmental decisions (Chapter 4).

1.1 The role of ethics in the economic theory: in Adam Smith's perspective

The attempt to undertake an economic analysis starting from premises belonging to an ethical theory should not be considered exceptional at all if we remind that more than 250 years ago the person who is usually designated as the founding father of the economic science, Adam Smith, followed exactly the same kind of approach. Indeed, not only the economic discipline, when formally born, was in essence a branch of moral philosophy (Sen 1987); even more relevantly, Smith's *Theory of Moral Sentiments* (representing his ethical doctrine, Smith 1994) is to be considered as the moral-philosophical foundation on which later the Scottish author built his more famous *Inquiry on the Wealth of Nations* (economic analysis, Smith 1976).

With the aim to offer a cogent justification concerning the general methodological approach underlying my doctoral writing, in the final Appendix I illustrate in detail the structured relationship between ethics and economics such as conceived by Adam Smith. However it has to be clear that the general Appendix is not to be interpreted as a part of the answer to the research question, but rather as sort of premise to the research question itself and to its following economic enquire.

Thus in the Appendix I show how the interaction between Smith's two mentioned works and sphere does not give rise to a conflict (the so called Adam Smith problem), but to a broader and coherent system where the moral regulatory mechanism of the impartial observer is assumed at the basis of the correct functioning of the market institution (Smith 1976, pp. 82-83). Thus it is possible to claim that the formulation of the *WN* relies on the thesis developed within the *TMS*, so that the latter essentially constitutes the (moral) foundation of the former. This means that within the Adam Smith's overall project the ethical background plays a fundamental role in the construction of his economic theory based on the market institution, that is his ethical theory constitutes an indispensable prerequisite for a correct functioning and interpretation of his economic system.

Therefore, after having sketched both, the economic system relying on self-interested impulses and on the division of labour proposed in the *WN* and the moral apparatus based on sympathy and on the figure of the impartial spectator described in the *TMS*, I provide a careful description of the precise relationship which links the two books and therefore the two spheres of ethics and economics in Adam Smith: he did not conceive the former as merely accessory to the latter, but he rather considered premises of ethical and moral character as necessary for the development of a structured and reliable economic theory. From this analysis it will be possible to conclude that, according to the Smithian approach, a valid economic analysis has to be grounded on pillars of ethical nature: this is exactly the general idea at the basis of my thesis.

Indeed, what is important for this dissertation is not the specific framework developed by Adam Smith, but rather the extension of the "Smithian method" and its exegesis. In particular, the precise hierarchy between ethics and economics assigned by the father of economic science can be broadened and interpreted on a more general level as the necessity of an ethical contextualization before proceeding with an economic analysis (Sen 1987). Again, this is the overarching approach underlying my whole thesis, which is instead based on the ethical system developed by John Rawls (1999).

Before describing in detail the Rawlsian ethical frame and in order to avoid misunderstanding it may be helpful to highlight some relevant analogies and distinctions between Smith's approach and John Rawls's theory. In particular, in developing their moral frames, both the authors more or less explicitly deal with institutions, meant here simply as the "rules of the game" played by the economic agents. However, while according to Adam Smith we essentially derive moral institutions through empirical experience (bottom-up perspective), Rawls aims at providing a normative theory of just institutions (top-down vision). Thus, although their approaches and their analytical tools diverge, both the authors address relevant issues concerning the shape of economic institutions.

What has to be clear I that I introduced Adam Smith and analysed his books only in order to provide a preamble to the general relationship between ethics and economics I believe it is important for works like mine. However in my thesis I am exclusively concerned with the "economic implications, especially when it comes to economic institutions and their consequences" of Rawls's social contract theory (Little 2014, p. 521).

1.2 John Rawls's social contract theory: an ethical system based on the veil of ignorance

As mentioned in the previous paragraph, the specific moral ground adopted at the basis of the economic research carried out in this doctoral dissertation is the ethical model designed by John Rawls (1999). The following Sections of the Introduction aim thus at providing the theoretical contextualization wherein the analysis of the next three Chapters develops.

Rawls contextualizes his theory within the social contract tradition[1] (Boucher et al. 2003, Darwall 2003 and Skyrms 1996), that is that set of theories which lay the legitimacy and the

1 Rawls explicitly makes clear how his moral theory aimed at representing a systematic alternative to the utilitarian tradition (Harsanyi 1976) which had been dominating the philosophical and economic debate at his time. Rawls pursued this goal because, in his opinion, when the utilitarianism deals with specific distributive issues it admits some morally unacceptable compromises.

foundation of political institutions and moral norms on an agreement. Then, at the basis of his social contract theory John Rawls places the following observation:

> "although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts"; however, the author goes on, "there is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share" (Rawls 1999, p. 4).

Thus, in order to deal with those "conflicting interests" that arise about the distribution of benefits, Rawls proposes an agreement between the involved parties. In other words, a contract becomes the formal tool to derive the norms aimed at governing the distribution of the benefits achieved by means of the social and economic cooperation between individuals.

As other theories inspired by the concept of social contract, also Rawls's scheme develops in two distinct phases (Sacconi 1991, pp. 68-69). In the first (constitutional) phase, the one concerning the contract in the state of nature, people formally free and focused on pursuing their own interests agree on the main principles necessary to regulate all the following relations that will take place in the second phase, the one where individuals "enter" the society (of law). Very importantly, the first phase of the Rawlsian social contract, termed by Rawls himself as "original position", follows a Kantian setting, that is the agreement phase "is not [...] thought of as an actual historical state of affairs" but it is rather "understood as a purely hypothetical situation" (Rawls 1999, p. 11, Rawls 1977, p. 161). In other words, Rawls's original position is equivalent to the adoption of a particular perspective, so the agreement and its principles are conceived as the result of a simple mental experiment[2].

A further feature of the general setting concerns then the specific aim of the Rawlsian social contract. In particular, Rawls circumscribes the goal of his hypothetical contract to the identification

2 From this assumption it follows a second analogy with Kant's social contract theory regarding the role played by the hypothetical agreement. More precisely, since it is not concrete the social contract assumes a prescriptive and comparative stand, such that "[o]ur social situation is just if it is such that by this sequence of hypothetical agreements we would have contracted into the general system of rules which defines it" now (Rawls 1999, p. 12). Said otherwise, we have to assume the perspective of the original position to evaluate the adherence of the existing state of affairs (institutions) to the systems of principles arisen from the (hypothetical) agreement. In this way "one conception of justice is more reasonable than another, or justifiable with respect to it, if rational persons in the initial situation would choose its principles over those of the other for the role of justice (Rawls 1999, pp. 15-16).

of those principles which can shape the society's "basic structure" (Rawls 1999, pp. 6-10 and pp. 47-52), viz. "the way in which the major social institutions fit together into one system, and how they assign fundamental rights and duties and shape the division of advantages that arises through social cooperation" (Rawls 1977, p. 159)[3].

Finally it is important to specify two qualities of the subjects who take part in the original agreement. The definition of the parties involved in the agreement procedure is among the most important features of each contractarian theory, because the attributions in terms of knowledge, beliefs, capacities, limitation etc. inevitably condition the outcome (principles). Rawls himself is clear on this issue: "depending upon how the contracting parties are conceived, upon what their beliefs and interests are said to be, upon which alternatives are available to them, and so on [...] there are many different contract theories" (Rawls 1999, p. 105).

Together with the instrumental rationality, which "must be interpreted as far as possible in the narrow sense, standard in economic theory, of taking the most effective means to given ends" (Rawls 1999, p.12), two main features distinguish the contractual parties of the Rawlsian setting. One element concerns their meta-personal interests: "the parties in the initial situation [are] mutually disinterested" or said otherwise, they "are conceived as not taking an interest in one another's interests" (Rawls 1999, p. 12)[4]. Besides, the parties are constrained with regards to the particular information they can access. More precisely, subjects in original position are excluded from any particular information which could twist the substance of the norms they are called to decide about. Thus, in original position,

> "no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence

3 Since for Rawls the subject of the original agreement is exclusively the "basic structure", Amartya Sen (2009) reproached Rawls of transcendental institutionalism (or, as I prefer, for institutional fundamentalism). Sen moves his criticism about Rawls's theory over two complementary lines. First of all Sen insists on how Rawls's approach is excessively focused on institutions, not paying enough attention to the actual life of people. Indeed in Sen's opinion a theory of justice cannot be neither "confined to the choice of institutions, nor to the identification of ideal social arrangements. The need for an accomplishment-based understanding of justice is linked with the argument that justice cannot be indifferent to the lives that people can actually live" (Sen 2009, p. 18). Secondly, Sen highlights how Rawls's theory is limited because it identifies, by means of two principles of justice (which might not be unique), an ideally perfect solution, while providing no hints on how to reduce actual and real inequalities in a comparative perspective: "in general the identification of a transcendental alternative does not offer a solution to the problem of comparisons between any two non-transcendental alternatives" (Sen 2009, p. 17, and pp. 96-98).

4 Translated in a more familiar economic language, Rawls assumes that subjects in original position are not characterized by any kind of other-regarding preference (like altruism, benevolence, inequity aversion or reciprocity etc.).

> and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism". Eventually, "[t]he persons in the original position have no information as to which generation they belong" (Rawls 1999, p 118).

"In this manner the veil of ignorance is arrived at in a natural way", since it is excluded "the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices" (Rawls 1999, p. 17). Basically, behind the veil of ignorance no one can design principles (norms or institutions) which might favour their particular situation, that is no one can take advantage of their personal contingencies in defining the principles[5].

Thus, since for Rawls "no one should be advantaged or disadvantaged by natural fortune or social circumstances in the choice of principles" (Rawls 1999, pp.16), only behind the veil of ignorance the principles for the basic structure of a society are the outcome of a fair agreement: given that the veil of ignorance restricts the particular information available to the parties everybody is equally represented since everybody has to choose in the same situation of perfect (mis)informational symmetry. On the contrary, "[a]greements reached when people know their present place in an ongoing society would be influenced by disparate social and natural contingencies" (Rawls 1977, p. 161).

1.2.1 The two principles of justice

The next step requires to focus on the decision-making process, viz. the actual reasoning adopted by individuals behind the veil of ignorance to derive those norms which are meant to regulate the institutional network of a society. In Rawls's opinion, given the setting of the original position, it is possible "to think of the [norms] principles as the maximin solution to the problem of social justice" (Rawls 1999, p. 132). Said otherwise, the heuristic device of the maximin, through which the alternatives are ranked according to their best worst-off outcome, is the most appropriate decision-making criterion to adopt given the (un)available information to the parties behind the veil of ignorance (Rawls 1999, pp. 130-139).

5 Given the veil of ignorance on particular information the parties have to reach an agreement only on the basis of impartial and general considerations. Indeed Rawls (1999, pp. 118-123) specifies how the knowledge of laws and general theories of human society remains available to the parties.

Rawls specifies then how his setting does not contemplate any assumption concerning individuals' risk aversion as well as regarding the probabilistic distribution of the outcomes which might characterize a society during its second phase. Not only the adoption of probabilistic or risk aversion assumptions would make the maximin criterion less suitable for his social contract model, but these hypotheses are also implicitly excluded by the (thick) veil of ignorance in original position: with the veil of ignorance not only the involved parties have no access to the probabilities concerning the different outcomes of the second phase, but they are also excluded from knowing all the possible social states (Rawls 1999, p. 134).

Within this general framework it is possible to describe in detail the objects of the agreement, which constitute the central nucleus of the Rawlsian ethical theory. However, before specifying their formulation and their meaning in depth it is essential to define some general features[6] of the principles themselves. According to Rawls they have to be: general, public, definitive and they have to impose a clear ordering about conflicting claims (Rawls 1999, pp. 112-118). This means that any redistributive principle without one of these features would be automatically rejected. Moreover, for Rawls the principles are to be ranked in a lexicographic ordering. This kind of ranking "requires us to satisfy the first principle in the ordering before we can move on to the second, the second before we consider the third, and so on", that is "[a] principle does not come into play until those previous to it are either fully met or do not apply" (Rawls 1999, p. 38).

Given all the features described so far, and above all given the restrictions on the particular information, according to Rawls people behind the veil of ignorance would unanimously agree[7] on two well defined principles (Rawls 1999, pp. 52-78), which are here stated through one of their numerous formulations:

> "the first requires equality in the assignment of basic rights and duties, while the second holds that social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society" (Rawls 1999, p. 13).

6 These features are endogenously defined within the original position itself, before deciding about the principles.

7 Unanimity follows by construction: since everybody is subject to the same restrictions about the particular information, everybody applies and everybody is "persuaded" by the same reasoning. This setting implies that there is no real bargaining on the principles (since all the parties are equal, they have formally and substantially no bargaining powers).

The Rawlsian ethics is condensed in these two principles."The first principle simply requires that certain sorts of rules, those defining basic liberties, apply to everyone equally and that they allow the most extensive liberty compatible with a like liberty for all". Being this principle located at the top of the lexicographic ranking it follows that "[t]he only reason for circumscribing basic liberties and making them less extensive is that otherwise they would interfere with one another" (Rawls 1999, p. 38). From a more practical perspective that means that for no reasons it is allowed to barter higher material welfare of somebody with a smaller set of liberties of anybody else. Said otherwise, it is not allowed to compensate a loss of freedom with any kind of economic benefit. For Rawls there is only one exception for not fully achieving the requirements of the first principle of justice: it occurs when two or more liberties are conflicting (Rawls 1999, pp. 171-228),

The second principle, termed by Rawls "difference principle", governs economic and social inequalities and it requires that "each person [and in particular the least advantaged, has to] "benefit from permissible inequalities in the basic structure"[8] (Rawls 1999, p. 56). While the first principle describes a sort of general background for the society "to start", the difference principle, which is "strictly speaking, a maximizing principle" (Rawls 1999, p.68), embodies in a more explicit way the redistributive rule which aims at regulating the division of those material benefits which are the result of the socio-economic cooperation between people.

One immediate issue related to the difference principle concerns the way to identify the least advantaged subjects within a certain society, and more in general the way to measure benefits and inequalities. Rawls tackles the problem adopting the concept of "social primary goods" (Rawls 1999, pp. 78-81). As defined by Rawls, these are objective (instrumental) goods which every rational individual is supposed to wish in a quantity as large as possible, independently from their specific ex-post preferences. Indeed for Rawls, more social primary goods enhance the possibility to achieve our own aims (plans of life) whatever they are. Thus, regarding the social primary goods every person has the same kind of preference [9]: the more the better.

Getting back to the essence of the difference principle, this basically states how an ex-post unequal distribution of social primary goods is to be considered fair (from the original position point of view) only as long as the uneven distribution maximizes the expectations of the least advantaged subjects. However, in order to explain why the difference principle is considered by

8 The complete version of the second principle, which includes a further specification, is provided by Rawls a few pages later, but that formulation does not add anything to the present analysis: "[s]ocial and economic inequalities are to be arranged so that they are both (a) to the greatest expected benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity" (Rawls 1999, p. 72).

9 This concept allows also to build up an objective and publicly shared index to measure benefits and inequalities. Thus the comparability challenge which usually hinders the utilitarian conception is simplified too.

Rawls himself as the best rule among all the possible alternative distributive schemes, it is necessary to look also at its philosophical justification.

Rawls explains the moral reason for the difference principle as it follows: "[t]he natural distribution [of talents and resources] is neither just nor unjust; nor is it unjust that persons are born into society at some particular position. These are simply natural facts". However "[w]hat is just and unjust is the way that institutions deal with these facts" (Rawls 1999, p. 87). That means that, since "[n]o one deserves his greater natural capacity nor merits a more favorable starting place in society" (Rawls 1999, p 87), the spontaneous distribution of those benefits generated through the social and economic cooperation cannot be considered morally legitimate. Therefore the construction of the difference principle in the original position is based on the idea to implement those institutions which can turn some natural facts into a morally just distribution, exactly redistributing primary social goods towards the most disadvantaged by the natural lottery.

Thus, from a pure philosophical perspective (which acts in parallel and not in substitution of with the economic maximin reasoning), the second principle of justice aims at defining an equitable way to deal with the arbitrariness of the case concerning the initial distribution of resources and talents, so that in a good approximation these are to be considered also as a common asset and not only as individual instruments achieved by chance. Said otherwise, since we have to take the natural distribution simply as given, the only way to morally legitimate the arbitrary distribution of talents and wealth is to act on the ex-post outcome of the social and economic cooperation based on those capabilities and resources.

In synthesis, through the redistribution of social primary goods operated by the difference principle it is ideally (practically) possible to nullify (diminish) the differences which depend on mere chance and which might, through the economic cooperation, benefit the most (arbitrary) advantaged subjects in the natural distribution, without at the same time taking into account the (naturally) most disadvantaged parties (Sandel 2009, pp. 150-166). Thus, inequalities are justified only if they improve the conditions of the worst-off.

Because the Rawlsian decision-making model constitutes the ethical frame at the basis of the whole following economic analysis, a final remark seems to be necessary: through the next Chapters the focus will be exclusively on the difference principle and its redistributive prescriptions. In other words, the first principle of justice, that is the realization of the broadest scheme of liberties for anyone, will be implicitly assumed as realized. Indeed, according to their lexicographic order, only considering as completely satisfied the prescriptions of the first principle the second principle can be taken into consideration.

1.2.2 The problem of the agreement stability and the sense of justice

After individuating the principles that are supposed to shape the basic structure, John Rawls dedicates a consistent part of his work (Rawls 1963, Rawls 1999 and Rawls 2001) to analyse the stability of the original agreement underlying the principles of justice. His main goal is to explain how the chosen principles, representing a distributive preference, agreed ex-ante behind a veil of ignorance, can become stable ex-post in the real word, after the veil is dropped.

This kind of analysis is not trivial at all if we highlight that, according to Rawls, the agreement in the original position, despite being fair, is not conceived as binding. In other words the distributive preferences (ethical norms) expressed behind the veil of ignorance are neither automatically implemented nor they are enforced by external mechanisms: in the second phase anyone could not follow the agreed rules, that is, after the veil is droppeseveryone could decide to deviate from the unanimously chosen distributive principles because these do not coincide with their ex-post individual real interests. Said in a more familiar economic language, in the real world everybody is in the condition to free-ride, gaining higher benefits by taking advantage of the collective efforts.

Thus, even if taking part in the agreement the parties unanimously recognize the mutually beneficial perspective embodied in a particular distributive preference (difference principle), the game where the involved parties independently decide whether or not to implement those preferences keeps being non-cooperative. For this reason, according to Rawls, in the second phase the parties are supposed to adhere to the prescriptions of the first moment of the agreement. Thus,

> "[j]ust arrangements may not be in equilibrium then because acting fairly is not in general each man's best reply to the just conduct of his associates. To insure stability men must have a sense of justice or a concern for those who would be disadvantaged by their defection, preferably both. When these sentiments are sufficiently strong to overrule the temptations to violate the rules, just schemes are stable. *Meeting one's duties and obligations is now regarded by each person as the correct answer to the actions of others*" (Rawls 1999, p. 435, *italics added*).

Therefore, according to Rawls, every subject entering the original position (agreement) should develop a sense of justice, that is a strong and effective desire to act ex-post consistently with the ex-ante chosen principles without the intervention of any external enforcement is necessary. Said

otherwise, a distributive preference (or principle) agreed ex-ante in the original position must generate by itself an endogenous motive (sense of justice) to comply with it, without the necessity to appeal to any exogenous system of enforcement. This must be true even when it is apparently against agents' strict self-interest, that is when the game in the real word is non-cooperative.

Further details about the sense of justice will be analysed in depth through the next Sections. However, a preliminary specification seems to be necessary in order to make clear this concept. In particular, according to Rawls the sense of justice is based on a the tendency to answer in kind (Rawls 1963 and Rawls 1999, pp. 429-440), that is a person decides to adhere to the prescriptions of the principles of justice on the condition she thinks that the other individuals who took part in the agreement will do the same.

1.3 International and empirical distributive justice from a Rawlsian perspective

Rawls's ethical system is much broader and richer than what sketched so far. Besides his theory had a remarkable influence beyond the philosophical debate: his work generated a sizeable secondary literature in the political and economic spheres as well. In order to understand how much his theory has been at the centre of a wide research involving different fields it is sufficient to recall that in the last decades John Rawls's writings have had on average about 10,000 quotes per year, with a total of about 182,000 quotes at the time this thesis was completed.

This enormous amount of secondary literature means that it would require another doctoral thesis to provide a complete and systematic analysis of all the debate arisen around Rawls's social contract theory. Instead, the purpose of this Introduction is more modest and limited. In particular, the next few pages aim at providing a circumscribed introductory description concerning the two macro areas of research taken into consideration in the reminder of the thesis: the distribution of resources between states, with a specific attention to the European framework, and the experimental works about Rawls's social contract theory.

1.3.1 John Rawls on international distribution of resources

According to Rawls, the theoretical frame and its conclusions presented in the Section 1.2 apply exclusively within closed systems like nation-states (Rawls 1999, p. 7). However, in the §58 of his main book (Rawls 1999, pp. 331-335) he did not neglect to sketch the way in which the international context should be approached from the perspective of his social contract theory. Instead, he provided a complete framework which extends in a systematic way his theory to relations between states only in two following works, both of them titled The Law of Peoples (Rawls 1993, 2001),

In particular, Rawls deals with the international setting by means of a second level original position[10] (Rawls 2001, p. 10), where the moral subjects of the social contract are not single individuals anymore, but rather (representatives of) peoples[11]. In the second level original position

> "the parties are subject to a veil of ignorance properly adjusted for the case at hand: they do not know, for example, the size of the territory, or the population, or the relative strength of the people whose fundamental interests they represent [...] they do not know the extent of their natural resources, or the level of their economic development, or other such information" (Rawls 2001, pp. 32-33).

Considering these analogies with his domestic theory and since in one passage Rawls underlines how "the procedure of construction, and the various steps gone through, are much the same in both cases" (Rawls 1993, p. 37), it might be intuitive to think about a simple and linear extension of the two principles of domestic justice (Rawls 1999) valid this time across nations. However in another passage Rawls specifies how "there is no reason to think that the principles that apply to domestic justice are also appropriate for regulating inequalities in a society of peoples" (Rawls 1993, p. 63). Therefore the domestic principles cannot be considered as fully general norms.

Thus, even if the standard procedure is applied at international level, the Rawlsian international theory is not immediate in its subsequent conclusions. Indeed Rawls "decomposes" the

10 "[A]fter the principles of justice have been adopted for domestic justice, the idea of the original position is used again at the next higher level" (Rawls 1993, p. 41).

11 With the term "peoples" Rawls basically means citizens united by a common sympathy (Rawls 2001, p. 23, note 17), and according to him "what distinguishes peoples from states [...] is that just peoples are fully prepared to grant the very same proper respect and recognition to other peoples as equals" (Rawls 2001, p. 35).

hypothetical second order (international) original position in three distinct moments[12] (Rawls 2001, p. 70), with the intention to take into account the possible and reasonable pluralism among peoples (Rawls 2001, p. 40). The first (Rawls 2001, pp. 11-58) and the second part (Rawls 2001, pp. 59-88) deal with what Rawls calls the "ideal theory". In the "ideal theory", liberal and decent (or hierarchical)[13] peoples, behind the veil of ignorance, agree on the following 8 principles for regulating their international relations (Rawls 2001, p. 37):

1. Peoples are free and independent, and their freedom and independence are to be respected by other peoples
2. Peoples are to observe treaties and undertakings
3. Peoples are equal and are parties to the agreements that bind them
4. Peoples are to observe a duty of non-intervention
5. Peoples have the right of self-defense but no right to instigate war for reasons other than self-defense
6. Peoples are to honor human rights
7. Peoples are to observe certain specified restrictions in the conduct of war
8. Peoples have a duty to assist other peoples living under unfavorable conditions that prevent their having a just or decent political and social regime[14]

The third moment of The Law of Peoples (Rawls 2001, pp. 89-120) is then split by John Rawls into two sub-types of "non ideal theory":

> "one kind deals with conditions of noncompliance, that is, with conditions in which certain regimes refuse to comply with a reasonable Law of Peoples" (outlaw states); "the other kind of non ideal theory deals with unfavorable conditions, that is, with the conditions of societies whose historical, social, and economic circumstances make their achieving a well-ordered regime, whether liberal or decent, difficult if not impossible" (burdened societies) (Rawls 2001, p. 90).

12 As clarified by Paden, "the delegates are divided into [different] groups that negotiate [three] separate contracts" (Paden 1997, p. 222).

13 Decent peoples, though they are not liberal, they are characterized by the two following features: they do not have aggressive aims towards other peoples and they respect human rights (Rawls 2001, pp. 64-67).

14 The last principle (8) was not included in the first version of The Law of Peoples (Rawls 1993, p. 43). Moreover, one question seems legitimate with regards to the one-way duty of assistance: why should liberal and decent peoples, the only ones who take part in the second order original agreement, assume a redistributive and altruistic commitment towards those peoples who do not take part in the agreement? (Pettit 2006, p. 54).

According to this framework outlaw states and burdened societies are not represented in the second order original position because they would not agree on the principles of international justice mentioned above. Indeed for Rawls it would be senseless to ask to those parties who hold unreasonable views or conditions to comply with reasonable principles of international justice (Beitz 2000, p. 676). In conclusion for Rawls it is not possible to imagine an agreement (a society) between peoples who are too different[15].

1.3.1.1 The relationship between Rawls's domestic and international theory

The one presented in the previous Section was a brief reconstruction of Rawls's social contract theory when dealing with relationships between states. According to the overall theoretical framework, John Rawls basically developed two distinct theories, one focused on a national level and the other one with an international horizon. This articulation makes then clear that for Rawls: the principles of domestic justice cannot suit the international context (and vice versa); the international principles themselves are not universally valid, that is they are not shared by all peoples (Nagel 2005, p. 127), because some of them do not hold the minimal political or economic conditions to reach a fair and durable agreement.

This differentiation given by Rawls to his two theories is even more evident if we focus on the distributive issue. More precisely, although within the international framework Rawls conceives a duty of assistance valid between peoples (Rawls 2000, p. 37), according to the structure given to his theories, distributive justice in the strict sense seems to hold within peoples but not between peoples (Barcelos et al. 2008, p. 3, Nagel 2005 p. 114, Pogge 1994, p. 195 and Wenar 2006, p. 99)[16]. In

15 The structure given by Rawls to his international theory triggers one important issue. In the domestic theory Rawls (1999, p. 118, 2001, p. 34) explicitly assumes that the veil of ignorance prevents the parties involved in the agreement from knowing their particular conception of the good. However in developing his international theory Rawls does not seem to be consistent with himself. More precisely, in The Law of Peoples (Rawls 2001) the parties entering the international original position are required to know if they are liberal, hierarchical, burdened societies or outlaw states (Rawls 2001, p. 71, note 10), that is they are required to be aware about their own conception of the good. Paden expresses such criticism in a clear way: "the veil of ignorance for the international original position differs significantly from that for the domestic original position in that, although the delegates are to be denied knowledge of the particular society they represent [...], they are to know that they represent [or not] liberal societies" (Paden 1997, p. 219). Said otherwise, the parties are supposed to be conscious about their specific conception of the good (Buchanan 2000, pp. 704-707, Caney 2002, pp. 99-114, Kuper 2000, pp. 648-650, Pogge 1994, pp. 206-207 and Pogge 2001 p. 247).

16 Thus, "though it is a universal principle that is to apply severally, or within every society, the difference principle is not global in reach" (Freeman 2006, p. 29), that is the difference principle is locally universal (Blake et al. 2015).

other words Rawls establishes a marked "distinction between the strong solidarity which must govern a generous redistribution between the members of the national community they claim to represent and the much weaker solidarity which must govern a more parsimonious and conditional assistance from the richer national communities to the poorer ones" (Van Parijs 2012b, p. 643). Thus

> "it may make a great deal of difference on Rawls's theory where the boundary of [a] society is drawn" (Scanlon 1973, p. 1066), because "it does not really matter whether one is born in Kansas or in Iowa" while "it matters a great deal whether one is born a Mexican or a U. S. citizen". Therefore we should "justify to a Mexican why [Americans] should be entitled to life prospects that are so much superior to hers merely because [they] were born on the other side of some line – a difference that, on the face of it, is no less morally arbitrary than differences in sex, in skin color, or in the affluence of one's parents" (Pogge 1994, p. 198).

These restrictions given by Rawls himself to the international redistributive horizon have been strongly criticized by the secondary literature[17], mainly because his international theory substantially ignores the paramount pillar of Rawlsian domestic justice: to provide a fair conception "that prevents the use of the accidents of natural endowment and the contingencies of social circumstance as counters in a quest for political and economic advantage" (Rawls 1999, p. 14), while it is quite evident how "the parties in the international original position would view the natural distribution of resources as morally arbitrary" as the distribution of talents or of social positions (Beitz 1999, p. 138, Nagel 2005, p. 119 and p. 124, Pogge 1988, p. 238).

Being aware about this kind of possible objection, Rawls tried to prevent it providing a justification for not considering an analogous of the difference principle valid at international level, but for conceiving only a weaker duty of solidarity:

> "I believe that the causes of the wealth of a people and the forms it takes lie in their political culture and in the religious, philosophical, and moral traditions [...], as well as in the industriousness and cooperative talents of its members [...] I would conjecture that there is no society anywhere in the world – except for the marginal cases – with resources so scarce that it could not, were it reasonably and rationally organized and governed, become well-ordered'' (Rawls 2001, p. 108).

17 One ecxeption to this stream is Reidy (2005, pp. 197-201) who finds it attractive to speculate on some corollaries which would be implicit in the duty if assistance.

The quoted passage contains a double claim: it does not exist any people so dramatically poor to be unable to accomplish with the goal of a just society, so a redistributive principle between peoples is not strictly necessary; even if there are strong and evident differences between peoples in terms of resources and political and economic development, those cannot be attributed to pure arbitrary contingencies, but rather to specific and conscious choices made by peoples themselves (Rawls 2001, pp. 117-118). In other words the latter consideration simply means that in Rawls's opinion "the causes of international inequality [are to be considered] purely domestic" (Pogge 2001, p. 252) and if a people is poor, even in terms of institutions, it is because that people up a certain extent decided to be poor (Rawls 2001, pp.117-118). Therefore, according to Rawls, no redistributive reasoning or principle between nations can be invoked: it would be unjust to transfer resources from one people to another one since there is a direct causal relation between the current political and economic development of a people and its previous decisions.

In synthesis, the Rawlsian international distributive justice is substantially characterized by different elements and triggers distinct issues compared to the domestic redistributive case (Pettit 2006, p. 52 and Wenar 2006, pp. 102-104). Rawls himself explicitly specifies that "how peoples treat each other and how they treat their own members are, it is important to recognize, two different things" (Rawls 2001, p. 83), highlighting how to two different issues of redistributive justice might correspond two alternative approaches and then two distinct solutions, that is two different sets of principles (Nagel 2005, pp. 122-124 and p. 127). In conclusion, even though some authors accused Rawls of creating a "structural disanalogy" between his domestic redistributive theory and his international justice (Pogge 2001, p. 249 and Pogge 2003, pp. 1745-1746), the differences should not be interpreted as between-theories inconsistencies, but more simply as different approaches to different circumstances.

1.3.1.2 The European Union within the Rawlsian theoretical frame

Given the specific purposes of this thesis, it is now important to understand which of the two approaches, the domestic or the international frame, is more suitable to interpret the European institutional framework. This task is not as straightforward as it might appear because of two parallel reasons: first of all Rawls did not clearly include the European Union as a formal object of any of his theories; at the same time the European Union "is neither a nation or a state, nor mankind as a whole" (Van Parijs 2012a). Therefore "none of the values defended [in Rawls's] works

provides alone a definitive axiological model that might elucidate the character of the European Union" (Barcelos et al. 2008, p. 6). This implies that there is not a precise overlap between any Rawlsian system singularly taken and the European Union. Instead, the last one represents a hybrid institutional framework not structurally contextualized within the two theories. However, this articulated theoretical framework does not have only a negative connotation, but also positive aspects because it means also that both theories become potentially suitable to interpret the European context.

Since the European Union is a set of nation-states, in a first moment it might be intuitive to approach it with the interpretative categories belonging to the Rawlsian international theory. What can The Law of People say about the European integration? According to the theoretical frame described so far we can claim that the European Union has to be considered a just international institution, not only because its member states (peoples) behind the veil of ignorance would formally agree on the principles of international justice designed for the "ideal theory", but also and overall because for Rawls a just Europe (world) is basically represented by a Europe (world) of just states (Barcelos et al. 2008, pp. 4-5, Nagel 2005, p. 115 and Pogge 1988, p. 235). In other words, the European Union has to be positively considered in the light of Rawls's international social contract theory since its single member states (peoples) are by themselves just in Rawlsian terms.

This is true in the light of Article 49 of the Treaty of Lisbon which identifies the minimum political requirements for the formal admission of a state[18] to the European Union. Article 49 of the Treaty of Lisbon basically provides the formal basis for any state to join the European Union, identifying the minimum qualifications that a candidate (future member) state must satisfy to enter the Union. These minimum requirements are also recalled in Article 2, in the Article 6 and more generally through the criteria of Copenhagen. They go from stability of institutions to democracy, from respect of human rights to equality. All of them essentially coincide with the features which make a people just according to Rawls (1999 and 2001). Thus, that the European Union is a just Society of Peoples (a just international arrangement in the Rawlsian vocabulary) is the simplest and the most coherent interpretation of the European integration we can provide in the perspective of the Rawlsian international theory.

But why does this linear and on one side satisfying conclusion does not seem to be fully compelling? Because we have to recognize how the European Union, given the current level of interdependence between its member states, is much more than a simple set of endogenously just systems. Therefore the Rawlsian theory of international justice represents a correct but limited

18 Geographically belonging to the European continent.

perspective to look at the European Union and to appreciate all its structural elements. In other words, the European Union triggers some issues which the Rawlsian international justice, such as it was conceived, does not deal with. In particular, as it was showed above, the Rawlsian international theory does not contemplate a redistributive scheme between peoples (as strong as the one present in the domestic theory).

Therefore, in order to analyze the distributive issue within the European Union I suggest using the Rawlsian domestic theory too. Adopting Rawls's domestic social contract theory in parallel to his international justice can provide a more complete and complementary interpretation of the European Union institutional arrangement, in particular regarding the distribution of resources between its member states.

It is within this theoretical context that Chapter 2 of the thesis deals with redistribution of resources at European level from the perspective of Rawls's domestic theory. In the second Chapter the general research question is perfected in the following formulation: by means of a Rawlsian agreement of domestic nature is it possible to justify a redistributive policy valid between the European member states? In order to answer to this question the adopted methodology is purely theoretical, and the core result is that we are essentially allowed to extend the range of applicability of the difference principle from the domestic dimension to the European one to the extent that the European context satisfies two basic requirements of Rawls's domestic theory.

In particular, in order to apply Rawls' domestic model and to draw the related conclusions, at European level the existence of two specific conditions should be proved: *a)* a mutually advantageous cooperative scheme among the involved parties and *b)* a set of formal institutions which defines a common basic structure. Chapter 2 shows that the European Union satisfies, from an empirical and substantial point of view, the conditions *a)* and *b)*, therefore it is possible to derive an actual European difference principle. The last one, in particular, prescribes to maximize the expectations of the European least advantaged, regardless of national borders. This preliminary result implies that the related research question is positively validated.

In conclusion to Chapter 2 I also claim that in order to give an actual realization to the normative prescription of the European difference principle, the European Union should put itself in the perspective of creating a sizeable European budget. In particular, by means of a European Fiscal Union it might be possible to channel in a European budget those resources which are the result of the economic cooperation taking place at European level and which are supposed to serve the European difference principle proposes. Thus the analysis of the European Union in a Rawlsian perspective leads to an important corollary: the European member states should take into

consideration to implement Fiscal Union not only for purposes of economic coordination, financial macro-stability and provision of European public goods. Instead the European countries are expected to move towards a Fiscal Union also for an important reason of moral-normative nature, which is in general embodied in the constitution of every fiscal system: the reallocation of resources with the intent to reduce inequalities. This might move the European Union from being a mere economic arrangement, to an integration which has some social traits.

1.3.2 Empirical and experimental distributive justice on John Rawls

The impulse to inquiry distributive justice from an empirical perspective can be traced back to two simple games that lead to a (r)evolution in the economic discipline: the dictator game (Engel 2011 and Kahneman et al. 1986) and the ultimatum game (Forsythe et al., 1994 and Güth et al. 1982). In the former a player has to choose if and how to divide an earned or a windfall endowment with a dummy player who has no voice on the division: the unilateral decision of the dictator constitutes the final payoff distribution between the two players involved in the game. The latter represents an extension of the dictator game, where to the receiver it is provided a veto power on the distribution proposed by the dominant player. In synthesis, in the ultimatum game the receiver can decide whether to accept or whether to refuse the dictator's proposal. If the receiver rejects the division proposed by the sender, both of them get nothing.

The participants' behaviours displayed and observed in these simple games gave rise to new research queries, because in none of the games the empirical evidence, that is the actual average payoff distribution, coincided with the standard theoretical predictions. In particular, while the standard economic theory expected that the participant playing in the dominant role would have shared on average (approximately) zero (in the ultimatum game) in the dictator game, the empirical results showed that this prediction was inconsistent with the actual behaviour, since the dictators share with the receiver on average (40%) 30% of their endowment. Such an inconsistency could not be explained by the standard consequentialist model of the purely self-interested *homo oeconomoicus*, who is supposed to take into account only the own material status. Therefore this gap between theory and empirical evidence gave a new stimulus to the economic research to analyse in depth the agents' distributive concerns.

In particular, the economic discipline, in the attempt to enrich the theoretical description of the economic agent by including in the last one the results of the empirical evidence, took two

different ways. On the one hand the research focused on better understanding the cognitive limits of actual economic agents, designing them as rationally bounded (Simon 1997) and therefore endowed with a set of cheap and intuitive psychological mechanisms (heuristics) which operate as fallible shortcuts in the self-interested maximization process (Kahneman et al. 2011). On the other hand the economic discipline extended its horizons taking into account social norms and preferences. In this second case the economic research developed the idea that, sometimes, self-interested behaviour is accompanied by other regarding concerns such as fairness, willingnes to comply, envy, equity, inequity aversion, spite, altruism, positive or negative reciprocity and the like. Nowadays there are many theoretical approaches which aim at modelling the economic behaviour with different mechanism of social norms or preferences (Bicchieri 2005, Bolton et al. 2000, Charness et al. 2002, Fehr et al. 1999 Fehr et al. 2004, Gintis et al. 2005, Kinbrough 2014, Konow 2000, Krupka et al. 2013, Levine 1988 and Rabin 1993).

Complementary to this development occurred from the theoretical side and which constituted a remarkable progresses in designing a *social homo oeconomicus*, the economic science started recognizing the importance of empirical research as an additional tool to the conceptual and analytical approach. Thus, through the years we have witnessed thousands of surveys, field researches and laboratory experiments whose goal was to amend and to provide a sort of empirical guidance to the theoretical speculation. It is within this second sphere that we can identify the study of modern empirical distributive justice. Starting from the two simple experiments described above, the experimental practice developed many alternative designs and approaches in order to describe as precisely as possible the different types of norms and social preferences which frame an economic decision.

This enlargement of the economic discipline in order to include an empirical perspective on distributive preferences has taken place through two main approaches. On the one hand subjects may be required to take fictitious distributive decisions (Gaertner et al. 2012). In this case the decision is enquired through a survey or a questionnaire and usually there is not a direct relationship between the choices and the decision-maker's material status. On the other hand the decisions taken by the individuals are meant to produce specific material consequences on the wealth of the involved subjects. Both these methodologies have been adopted also to test Rawls's theory of justice and his maximin principle (see for instance Bond et al. 1991 and Michelbach et al. 2003 respectively). In general, however, the behavioural and experimental research about the Rawlsian decision-making model focused on testing Rawls's principles against alternative theories which deal with distribution of resources (Aguiar et al. 2013, Engelmann et al. 2004 and Frohlich et al. 1987).

Frohlich et al. (1987) simulated in a laboratory environment Rawls's veil of ignorance procedure. In particular they tested his predictions on the unanimous consensus on a distributive rule that maximizes the welfare of the worst-off against other distributive principles (like maximizing the average wealth). The understanding of the available principles, measured by means of a test, was a precondition for the participants for continuing with the experiment. After this preliminary phase the participants were asked to decide together between different distributions of income representing the mentioned principles. In particular, they were asked to discuss about the given distributive schemes not knowing their position in any scheme. Indeed, they were assigned to a position after they agreed on a principle, that is on a distributive scheme. The results of this experiment showed that the subjects always reached a unanimous agreement on the distributive principle but none of the 29 groups who took part in the sessions chose the distributive scheme compatible with Rawls's difference principle[19].

Engelmann et al. (2004) presented a simple one-shot distribution experiment where they compared the relative importance of maximin preferences against other concerns like efficiency or inequality aversion. They provided to the experimental subjects a decision sheet containing three different allocations of money between three individuals. Each subject subject had to choose one scheme among the three proposed. Before taking any decision they were also informed that after their choice they would have been randomly assigned to the three roles. In this way they face role uncertainty like behind Rawls's veil of ignorance. Lastly, only the choice of the participant selected as person 2 mattered, that is it was considered for the actual payoff distribution. Furthermore, a control treatment assigned fixed roles in advance. The final results of they study showed that the maximin preferences (together with other components) have a substantial impact on the distributive choices of the experimental subject. Nevertheless the control treatment did not provide any indication that their results were driven by the introduction of the uncertainty stage.

Aguiar et al. (2013) designed an experiment in order to investigate three different mechanisms to achieve impartiality in distributive justice. In particular they considered a first-person procedure, inspired by the Rawlsian veil of ignorance, and two third-party procedures, that is an involved spectator and a detached observer. In the laboratory, groups of four people with different endowment levels were constituted. In particular, every group was formed of three veiled stakeholders and one third-party observer (either a detached observer or an involved spectator). The task of each of the four participants was to choose how to distribute a surplus among the three

19 Of the remaining groups 25 groups chose to maximize the average with a floor constrain and five to simply maximize the average.

stakeholders. The authors found that the three methods induced a fair amount of redistribution between the subjects. However, the levels of redistribution showed to be significantly different across the three mechanisms of impartiality. In particular, the detached observers behaved in a more egalitarian way (in particular 68% of the decisions were perfectly egalitarian), followed by the veiled stakeholders (57%) and then by the involved spectators (54%)[20].

In synthesis, this empirical literature produced no strong and compelling evidence in favour of Rawls's distributive model, even tough it was ascertained that preferences of Rawlsian type do not depend only on risk attitudes (Schildberg-Hörisch 2010).

Notwithstanding these non-encouraging empirical outcomes on the distributive preferences proposed by John Rawls, the literature did not exhausted its research interest concerning the empirical enquiry of his theory of justice. Indeed, it is necessary to underline how the difference principle is more sophisticated than it might seem to be. More precisely, the implementation of a distributive preference of Rawlsian type is not an innate predisposition, but rather a rational product of two distinct moments, associated with both of the two stages of the social contract.

More precisely, according to Rawls the difference principle is the result of an ex-ante phase, where the subjects express their distributive preferences from a particular perspective (the original position) and of an ex-post phase, where the individuals display those preferences through their actual behaviour in the real world. Therefore, within the Rawlsian theory both phases become essential to observe an economic agent behaving according to the difference principle. However, the previous empirical literature not always took into sufficient account this fundamental interdependence between the two mentioned momenta.

In particular, it is important to highlight that given their different structural configuration the two phases might differ with regards to the chosen distributions, that is the same subject might opt for the difference principle in the first stage and then in the second one, when he realizes he is in an advantaged position, he might display an actual behaviour which is not concerned at all with the status of the worst-off. Despite this theoretical possibility, it should be considered senseless to conceive two phases, if only the second is supposed to have real and effective consequences in the society. It would be senseless unless the first moment was not conceived with the aim to have a precise influence on the second one.

As reported in the Section 1.2.2, within Rawls's theory, the distributive preference displayed behind the veil of ignorance cannot be unbound from the pragmatic implementation of the

20 In particular 68% of the decisions of the participants who played in as the detached observers were perfectly egalitarian; this percentage reduced to 57% for the veiled stakeholders and to 54% for the involved spectators.

difference principle in the real word. In other words, the ex-post behaviours cannot overlook the mental experiment of the original position, that is in Rawls system any ex-ante distributive preference, if not reasonably confirmed by means of an ex-post actual choice, would be an empty result. Within John Rawls's social contract theory this is the problem of the stability of the agreement through the sense of justice. This issue was thus considered also by somne reced behavioural and experimental economic literature.

1.3.2.1 The sense of justice in the laboratory: the social conformist preference model

An innovative field of literature on distributive justice explored in depth the Rawlsian egalitarian conception and in particular his idea of the sense of justice (see the Section 1.2.2), which is the substantial glue that bonds the two phases of the social contract. Modelling Rawls's moral psychology on a belief-dependent disposition, since for Rawls "the capacity for a sense of justice [is] built up by responses in kind" (Rawls, 1999, p. 433), the idea at the basis of the proposed behavioural model is that an individual should implement the distributive preference showed behind the veil of ignorance only at the condition that he or she believes that the other subjects involved in the agreement would or will do the same. In particular,

> "[t]he reason that explains a particular decision in the *ex post* game is knowledge of what the players will effectively do. Moreover, this knowledge about the other players' decisions must be consistent with their being symmetrically able to predict the others' behavior and to choose their best response to those predictions. Therefore, it is not the impartial selection of a desirable *ex ante* solution, but the knowledge of other players' *de facto* behaviors, that provides the proper reason for acting in the *ex post* context (Sacconi et al. 2011, p. 281)".

The behavioural model of contractarian conformist preferences (Faillo and Sacconi 2007, Sacconi and Faillo 2010, Grimalda and Sacconi 2005 and Sacconi and Grimalda 2007), following Rawls's setting described so far, introduces in the utility function, together with the standard material payoff, a psychological payoff assigned by the agent to the compliance expected from the other players involved in the agreement. In this way "psychological equilibria based on conformist preferences – with which we formally represent the `sense of justice' – provide an endogenous

explanation of social contract compliance" (Sacconi et al. 2011, p. 286) and more in general a solid justification of distributive preferences of Rawlsian type in the real world.

Consistently with the formal structure given by Rawls, at the basis of the conformist preferences model there is the assumption that the individuals play a non-cooperative game where the Nash equilibria are not mutually beneficial. Besides, the model assumes that before this main game is played, the subjects take part in a preliminary stage where they can agree on a distributive rule under an impartial perspective (behind a veil of ignorance). Finally, the rule chosen in the agreement phase is not automatically implemented, i.e. it is not conceived as binding, therefore the agreed norm should not prevent players to reach, in the second phase, the standard Nash equilibrium where they maximize only their own material payoff according to the best respond to the others' freeriding.

In particular, the theory of conformist preference explains how the impartial agreement becomes a tool for the selection of an alternative (psychological) equilibrium where players comply with an ex-ante counter-maximizing distributive preference counterbalancing the potential material loss with a psychological gain. The resulting utility function (1) is thus made of two main components: a standard consequentialist part based on the material payoff; a conformist preference part which provides psychological utility under the condition of (expected) reciprocal compliance.

$$(1) \quad V_i(\sigma) = U_i(\sigma) + \lambda_i F[T(\sigma)]$$

In particular:

- U_i represents the canonical utility gained from the material payoff achieved in the state σ (that represents a combination of the players' strategies). The remaining part of the function embodies instead the psychological utility;

- T is the collective distributive preference unanimously agreed during the agreement phase. In other words, T represents the chosen social welfare function aimed at ordering all the possible states σ of the ex-post world. Moreover, the closer the ex-post distribution of material payoffs to the norm T, the higher is the value of T;

- F is an index of agent i's conditional and reciprocal conformity with principle T. More precisely this parameter measures the contribution, through her choice, of the player i to the maximization of T, conditioned on the expected actions (which in turn are conditioned on the player i's expected actions) of the other players in the state σ. The index can range from 1 (full

conformity) to 0 (no conformity at all) and operates as a weight on λ, determining how much the last one can affect the player's psychological utility;

- λ is an exogenous value meant to represent Rawls's sense of justice. It embodies the agent's motivational force (psychological disposition) to act on the motive of reciprocal conformity with the agreed norm.

The presented model was tested in different versions of the so called exclusion game and it demonstrated to have a very robust predictive power (Degli Antoni et. al 2016, Faillo et al. 2008, Faillo et al. 2014, Sacconi and Faillo 2005, Sacconi and Faillo 2010 and Tammi 2011). Basically, the exclusion game is a resource allocation experiment, in particular a multiple dictator game, with a preliminary stage where the participants, in an anonymity condition, have the possibility to reach a unanimous agreement about the norm (distributive preference) to follow during the second stage, the actual exclusion game.

Three are the most important features of the game that replicate Rawls's theory and that at the same time allow to implement the social conformist preference model: the choice of the distributive rule is taken behind a veil of ignorance, that is the players agree on a distribution not knowing their (future) role in the second stage of the game. When assigned, the players' roles are differentiated with regard to their endowments or to their decision-making powers. Excluding some players from active roles is meant to reproduce the arbitrary distribution of social and economic conditions. Finally, during the actual exclusion game the agreement concerning the chosen distributive norm is neither formally nor substantially binding, that is the players endowed with full decision-making powers are not bounded, so they can adopt any available alternative strategy.

Therefore, given the game design, during the actual exclusion game every player in the dictator role should choose the strategy which maximizes his or her own payoff independently from the rule agreed during the ex-ante vote. Nevertheless, the provided experimental evidence showed in a compelling way how the unconstrained ex-post compliance to the ex-ante chosen distributive preference is unexpectedly high even in those cases where the groups unanimously agreed about a counter-maximizing rule, and in particular about a distributive scheme compatible with Rawls's difference principle. These are the words of the authors summarizing their main results about the exclusion game (Faillo et al. 2014, p. 242):

> "in the agreement treatment we observed that all groups reached an agreement, that the large majority of groups agreed on the equal division rule and that a high

> percentage of subjects chose to comply with the rule believing that other members of their group would do the same. In addition, on considering the relation among agreement, expectations and actual choices, we can conclude that the agreement 'under the veil' induced the convergence of subjects' beliefs of reciprocal compliance and consequently activated a preference to act in accordance with fairly agreed principles conditionally on reciprocal compliance beliefs"

Thus, the general behaviour observed in the exclusion game was considered consistent with the Rawlsian concept of the sense of justice and with its formal representation through the social conformist preferences model. The last one was assumed at the basis of the development of the research questions in the Chapters 3 and 4, which inquire respectively tax compliance and international agreements concerning the reduction of greenhouse gas emissions.

1.3.2.2 The sense of justice and tax compliance

Developed in partnership with Professor Luigi Mittone of the University of Trento, the Chapter 3 approaches tax compliance in a Rawlsian perspective. In this Chapter the general research question is inflected as it follows one: can a (laboratory) Rawlsian veil of ignorance generate more equitable (redistributive) tax regimes and guarantee long-lasting voluntary tax compliance?

In order to answer to this question the paper moves along three converging research tracks. First of all, the study is part of the "slippery slope" framework. This theoretical tool clarifies how individual tax compliance is not the result of a mere mathematical weighting of risks and benefits connected to the decision to pay or to evade taxes. Instead, according to the "slippery slope" theory, tax compliance depends as much on deterministic elements like audits and fines as on a broader set of psychological and environmental variables which drive voluntary cooperation.

Second, the research specifically focus on the Rawlsian concept of the sense of justice and its behaviouralist interpretation mentioned above. Summarized in a few words, the sense of justice embodies the idea that an agreement of Rawlsian nature should generate by itself an endogenous and increasing over time willingness to comply with the chosen redistributive scheme. Shifted in the tax context the sketched framework can be approximately illustrated in the following way: after an ex-ante voting phase where the individuals choose a tax regime behind a veil of ignorance,

external enforcement mechanisms like audits and sanctions should not be indispensable in order to guarantee ex-post tax compliance

Third, the answer to the research question is entrusted to a laboratory experiment which relies on the exclusion game design.

In the literature on tax evasion many experimental and field studies proved a behavioural regularity regarding a positive causation effect between individuals' participation in the definition of tax rules (like rates, audits, fines etc.) and the following level of tax compliance. According to this literature stream, compared to a situation where rules are exogenously assigned, norms and institutions legitimized in a direct way enhance the cooperative attitude. In brief, letting individuals to have a voice on specific tax issues increases voluntary tax compliance. However, our study does not focus on the so-called "participation effect". Instead, with our experiment we aimed at testing the effects of two different voting conditions, one of them inspired by the Rawlsian social contract theory.

In particular, before accessing the actual tax compliance game, in a preliminary voting stage groups of three players each are asked to reach a unanimous agreement concerning a tax regime (set of tax rates) to apply to three possible levels of income. The four tax schemes are designed to generate the same ex-post average wealth. However, they are more or less progressive, that is they are differentiated with regards to their redistributive effects.

In the baseline experiment we ask participants to reach an agreement on a tax scheme after they are randomly assigned their personal level of income. In the veil treatment instead, during the voting stage, a Rawlsian (laboratory) veil of ignorance is introduced, that is we ask players to unanimously agree on a tax regime before assigning them a level of income.

According to Rawls's social contract theory and some of its experimental evidence the veil of ignorance should produce some effects on the distribution of votes itself in the voting stage as well as on compliance level in the actual game. In particular, with a veil of ignorance the tax scheme which maximizes the expectations of the lowest level of income (representing the worst-off in the game) should be chiefly chosen and compliance in the veil condition should be not lower and constant across rounds compared to the no-veil treatment.

The empirical results showed in Chapter 3 only partially confirmed our theoretical predictions. This mean that the research question did not find a complete empirical validation. In particular, in the veil condition the distribution of votes shifted towards a more (but not the most) progressive, that is redistributive, tax regime. Besides, average tax compliance was not lower than

in the baseline treatment. However, compliance in the veil treatment was not constant across rounds, viz. the impartial agreement was not stable as predicted.

Nevertheless, taken together these are considered a positive result in favour of the Rawlsian veil of ignorance and its social conformist preferences model. Indeed, even though in the veil treatment groups agreed on more progressive tax schemes, which are more demanding for richer people by construction, the average level of compliance was not negatively affected, and this behavioural trajectory was driven by the expectations concerning the level of compliance of the other players in the tax game.

1.3.2.3 The distribution of resources between generations in a laboratory experiment

Chapter 4, developed in collaboration with Professor Lorenzo Sacconi of the University of Milan and Professor Faillo of the University of Trento, deals with economics and ethics of climate change in a Rawlsian framework. More in general the research focuses on distribution of resources between generations and in particular on modern international agreements which aim at reducing global greenhouse gas emissions.

Our study moves from the observation that notwithstanding more than thirty years of international negotiations on climate actions, mankind has not been able to reduce noxious gas emissions which, through the global warming, might seriously harm future generations. Thus, not limiting consumption of natural resources today, the present generation profits of its privileged position on the timeline and enjoys a higher level of wealth at the expenses of future generations, who will have to bear the costs. This stalemate situation is the result of the intersection of two well identified circumstances occurring at international level.

First of all, the global reduction of global greenhouse gas emissions, and therefore of the total present consumption, is systematically conditioned upon the distribution of costs between nations. More precisely, no nation is really available to pay more than the others reducing more than others its own current consumption of natural resources in order to reach the common goal.

Second, given that within the current geopolitical frame there are not international institutions which can enforce, by means of audit or sanctioning mechanisms, compliance to any agreement, even formal contracts (like the Kyoto protocol) are intrinsically fragile. In other words, whatever the ideal outcome of an agreement on climate actions might be, given the absence of any institution

which can change the economic incentives' structure, the Nash equilibrium always prescribes to defect not reducing of the current consumption of natural resources.

Within this broad frame we inflect the general research question of the thesis as it follows: can an agreement of Rawlsian type concerning the management of common natural resources help to reduce present greenhouse gas emissions and more in general to produce a fair path of consumption between generations?

In his main works Rawls only marginally dealt with distribution of resources between generation. Nevertheless his decision-making model inspired to the idea of a social contract is considered suitable to inquire the management of natural resources and the related agreements on reduction of greenhouse gas emissions. Indeed, in moving from the (intragenerational) kernel of his social contract theory to the intergenerational application of it, Rawls specifies along many passages that the impartial agreement which is supposed to generate principles for the distribution of resources between generations is intragenerational: since the contemporaries are deprived of the information on the generation of the history they belong, they are constrained to be impartial also on the intergenerational norms. Within this framework Rawls derives the just saving principle, which basically requires to each generation to contribute in a fair way to those coming later.

In order to answer the research question we designed a laboratory experiment shaped on Rawls's intergenerational social contract theory. In particular, we group participants in sets of three people and we place the groups along chains of different lengths. Each group of players is then meant to represent a set of contemporaries subjected to the decision of all previous groups in the chain and decides about the destiny of the following ones.

In the baseline treatment, within the active groups players individually have to decide how much money to withdraw from a common fund. If the total withdraw is lower or equal to the fixed threshold, the chain can continue and the group in the next position become active. Otherwise the chain shortens starting from the last generation of the chain itself. The groups cut out are forced to leave the game and they cannot take any decision.

In the Rawlsian treatment, the sequential dictator group game is preceded by a preliminary voting phase, where we ask groups to agree, among their inner members, on a rule concerning the management of the common fund. However, they are asked to choose an intergenerational rule before they are told which generation (position in the chain) they belong. Members of every group can then agree whether to limit their individual withdrawal from the fund, letting the chain to continue intact, or whether to appropriate an amount of resources higher than the threshold, shortening in this way the chain.

Consistently with the Rawlsian just saving principle and its behavioural model we expected groups to agree mainly on the self-containing rule which allows chains to continue intact, whereas according to the sense of justice theory we expected participants to abide by the impartially chosen rule even though in the actual game the agreed rule is not designed as binding.

The experimental evidence strongly confirmed the first hypothesis. However, notwithstanding an average compliance rate of 80%, given the specific structure of the game, that was not sufficient to guarantee continuation of the chains compared to the baseline treatment, where groups directly enter the sequential game without taking part in any agreement stage.

Thus, our answer to the research question is consistent with the general framework about climate change agreements: everybody is ready to agree that a reduction of greenhouse gases, and therefore of current consumptions is necessary. However, compliance keeps being an open issue.

References

Aguiar, F., Becker, A., & Miller, L. (2013). Whose impartiality? An experimental study of veiled stakeholders, involved spectators and detached observers. Economics & Philosophy, 29(2), 155-174.

Atkinson, A. B. (2009). Economics as a moral science. Economica, 76, 791-804.

Barcelos, P. & Queiroz, R. (2008). Which Values for the European Union? Europe between Principles of Domestic and International Justice. Published on http://www.ifilnova.pt.

Beckerman, W. (2017). Economics as applied ethics: value judgements in welfare economics. Palgrave Macmillan.

Beitz, C. R. (1999). Political theory and international relations, with afterward. Princeton University Press.

Beitz, C. R. (2000). Rawls's Law of Peoples*. Ethics, 110(4), 669-696.

Bicchieri, C. (2005). The grammar of society: The nature and dynamics of social norms. Cambridge University Press.

Blake, M. & Smith, P. T., "International Distributive Justice", The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/spr2015/entries/international-justice/>

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. American economic review, 90(1), 166-193.

Bond, D. & Park, J. C. (1991). An empirical test of Rawls's theory of justice: A second approach, in Korea and the United States. Simulation & Gaming, 22(4), 443-462.

Boucher, D., & Kelly, P. (2003). The social contract from Hobbes to Rawls. Routledge.

Buchanan, A. (2000). Rawls's law of peoples: Rules for a vanished Westphalian world. Ethics, 110(4), 697-721.

Caney, S. (2002). Cosmopolitanism and the Law of Peoples. Journal of Political Philosophy, 10(1), 95-123.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. The Quarterly Journal of Economics, 117(3), 817-869.

Darwall, S. L. (2003). Contractarianism, contractualism. Malden, MA, Blackwell Pub., 2003.

Engel, C. (2011). Dictator games: A meta study. Experimental Economics, 14(4), 583-610.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. American economic review, 94(4), 857-869.

Faillo M., Ottone S. & Sacconi L. (2014). The Social Contract in the Laboratory: An Experimental Analysis of Self-Enforcing Impartial Agreements. Public Choice, 163(3-4), 225-246.

Faillo, M., Sacconi, L. (2007). Norm Compliance: The Contribution of Behavioral Economics Theories, in: Innocenti, A./P. Sbriglia (eds.), Games, Rationality and Behaviour, Essays in Behavioural Game Theory and Experiments, London, 101-133.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. The quarterly journal of economics, 114(3), 817-868.

Fehr, E. & Fischbacher, U. (2004). Social norms and human cooperation. Trends in cognitive sciences, 8(4), 185-190.

Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. (1994). Fairness in simple bargaining experiments. Games and Economic behavior, 6(3), 347-369.

Freeman, S. (2006). The law of peoples, social cooperation, human rights, and distributive justice. Social Philosophy and Policy, 23(1), 29-68.

Frohlich, N., Oppenheimer, J. A. & Eavey, C. L. (1987). Laboratory results on Rawls's distributive justice. British Journal of Political Science, 17(1), 1-21.

Gaertner, W. & Schokkaert, E. (2012). Empirical social choice: questionnaire-experimental studies on distributive justice. Cambridge University Press.

Gintis, H., Bowles, S., Boyd, R. T. & Fehr, E. (2005). Moral sentiments and material interests: The foundations of cooperation in economic life (Vol. 6). MIT press.

Grimalda, G. & Sacconi, L. (2005). The constitution of the not-for-profit organisation: reciprocal conformity to morality. Constitutional Political Economy, 16(3), 249-276.

Güth, W., Schmittberger, R. & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. Journal of economic behavior & organization, 3(4), 367-388.

Harsanyi, J. C. (1978). Bayesian decision theory and utilitarian ethics. The American Economic Review, 68(2), 223-228.

Kahneman, D. & Egan, P. (2011). Thinking, fast and slow (Vol. 1). New York: Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1986). Fairness and the assumptions of economics. Journal of business, S285-S300.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. American economic review, 90(4), 1072-1091.

Krupka, E. L. & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. Journal of the European Economic Association, 11(3), 495-524.

Kimbrough, E. O., Miller, J. B. & Vostroknutov, A. (2014). Norms, frames and prosocial behavior in games. Working paper, Simon Fraser University, Bocconi University, Maastricht University.

Kuper, A. (2000). Rawlsian global justice: beyond the law of peoples to a cosmopolitan law of persons. Political theory, 28(5), 640-674.

Laffont, J. J. (1975). Macroeconomic constraints, economic efficiency and ethics: An introduction to Kantian economics. Economica, 42(168), 430-437.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. Review of economic dynamic.

Little, D. (2014). Rawls and economics. A Companion to Rawls, First Edition. Edited by Jon Mandle and David A. Reidy, published by John Wiley & Sons, Inc, 504-525

Michelbach, P. A., Scott, J. T., Matland, R. E. & Bornstein, B. H. (2003). Doing Rawls justice: An experimental study of income distribution norms. American Journal of Political Science, 47(3), 523-539.

Nagel, T. (2005). The problem of global justice. Philosophy & public affairs, 33(2), 113-147.

Paden, R. (1997). Reconstructing Rawls's law of peoples. Ethics & International Affairs, 11, 215-232.

Pettit P. (2006). Rawls' peoples, in Martin R. & Reidy D. (ed.), Rawls's Law of Peoples: A Realistic Utopia? (pp. 38-55), Oxford, Blackwell.

Pogge, T. W. (1988). Rawls and global justice. Canadian Journal of Philosophy, 18(2), 227-256.

Pogge, T. W. (1994). An egalitarian law of peoples. Philosophy & Public Affairs, 23(3), 195-224.

Pogge T. W. (2001). Rawls on international justice, The Philosophical Quarterly, 51(203), 246-253.

Pogge, T. W. (2003). The incoherence between Rawls's theories of justice. Fordham L. Rev., 72, 1739.

Rabin, M. (1993). Incorporating fairness into game theory and economics. The American economic review, 1281-1302.

Rawls J. (1977). The basic structure as subject. American Philosophical Quarterly, 14(2), 159-165.

Rawls J. (1987). The idea of an overlapping consensus, Oxford J. Legal Stud., 7, 1.

Rawls J. (1993), The Law of Peoples, Critical Inquiry, 20(1), 36-68.

Rawls J. (1999). A theory of justice, revised edition. Harvard University Press. Cambridge.

Rawls J. (2001). The law of peoples: with the idea of public reason revisited, Harvard University Press.

Reidy, D. A. (2007). A just global economy: In defense of Rawls. The Journal of Ethics, 11(2), 193-236.

Sacconi, L. (1990). L'etica degli affari. Milano, Il Saggiatore.

Sacconi, L. & Faillo, M. (2010). Conformity, reciprocity and the sense of justice. How social contract-based preferences and beliefs explain norm compliance: the experimental evidence. Constitutional Political Economy, 21(2), 171-201.

Sacconi, L. & Grimalda, G. (2007). Ideals, Conformism and Reciprocity: A Model of Individual Choice with Conformist Motivations, and an Application to the Not-for-Profit Case, in: Bruni, L./P. L. Porta (eds.), Handbook of Happiness in Economics, London.

Sacconi, L., Faillo, M. & Ottone, S. (2011). Contractarian Compliance and the 'Sense of Justice': A Behavioral Conformity Model and Its Experimental Support. Analyse & Kritik, 33(1), 273-310.

Sandel, M. J. (2009). Justice: What's the right thing to do?. Penguin. London

Scanlon T. M. (1973). Rawls' theory of justice. University of Pennsylvania Law Review, 121(5), 1020-1069.

Schildberg-Hörisch, H. (2010). Is the veil of ignorance only a concept about risk? An experiment. Journal of Public Economics, 94(11-12), 1062-1066.

Sen, A. (1987). On ethics and economics. Blackwell Publishing. Oxford

Smith, A. (1976). The theory of moral sentiments. Edited D. D. Raphael and A. L Macfie. Claredon Press. Oxford

Smith, A. (1994). An inquiry into the nature and causes of the wealth of nations. Edited by Edwin Cannan. The Modern Library. New York.

Simon, H. A. (1997). Models of bounded rationality: Empirically grounded economic reason (Vol. 3). MIT press.

Skyrms, B. (1996). Evolution of the social contract. Cambridge: Cambridge University Press

Van Parijs P. (2012a). No Eurozone without Euro-dividend, Unpublished Manuscript. URL: http://www.fondationpaulricoeur.fr/uploads/medias/articles_pr/no-eurozone-without-eurodividend.pdf.

Van Parijs P. (2012b). International distributive justice, in Goodin R. E., Pettit P., & Pogge T. W. (ed.), A Companion to Contemporary Political Philosophy (Vol. 2, pp. 638-652), John Wiley & Sons.

Wenar L. (2006). Why Rawls is not a cosmopolitan egalitarian, in Martin R. & Reidy D. (ed.), Rawls's Law of Peoples: A Realistic Utopia? (pp. 95-113). Oxford. Blackwell.

2. The European Social Contract: a Rawlsian Approach in Favour of Fiscal Union[a]

Abstract. Why might the European member states seek for Fiscal Union? General fiscal coordination, macro-stability purposes and provision of (European) public goods are certainly goals of paramount importance for the implementation of Fiscal Union at European level. However, there is an equally important reason of moral-normative nature embodied in the constitution of any fiscal system: reallocation of resources. The core of the paper is the idea that Rawls's social contract theory can provide some insights about the implementation of European Fiscal Union in the re-allocative perspective. The reasoning put forward in the paper shows how the current European framework can be essentially considered an appropriate object of Rawls's domestic theory since the European Union holds those two descriptive elements which are necessary and sufficient to raise redistributive issues, to apply Rawls's pure procedural justice and then to derive a difference principle at European level: *a)* the mutually advantageous cooperation among its members and *b)* a set of formal institutions which constitute a basic structure. The European difference principle prescribes to redistribute resources in order to maximize the expectations of the most disadvantaged European citizen(s). A corollary which follows from this conclusion is that the actual redistribution according to similar scheme is achievable by means of Fiscal Union at European level.

JEL Code: D30, E62, F55

Keywords: Difference Principle, European Integration, European Union, Fiscal Union, John Rawls

a Fondazione Cassa Rurale di Trento provided a considerable grant for the development of the paper. Different versions of the paper were presented at various conferences and workshops (Cesifo, INET, Fondazione Istituto Cattaneo and SISP). Published as working paper: *Klaser, K. (2018). The European Social Welfare Function Shaped on a Difference Principle: A Normative Rawlsian Approach in Favour of Fiscal Union (No. 7186). CESifo Group Munich*

"What would a just Europe look like? What does justice mean when applied to that weird political entity now called the European Union, which is neither a nation or a state, nor mankind as a whole?"

Philippe Van Parijs

Introduction

Indeed, without any of these substantial conditions it is formally improper to insist to formulate a difference principle, as much at domestic level as at international level.

The European Union is a compound of nation-states characterized by a particular institutional asymmetry: as pointed out with different emphasis by Ferrera (2009), Martinsen (2013) and Scharpf (2002) while some crucial economic issues are directly or indirectly addressed at community level, social welfare policies remain an exclusive prerogative of the single member states. This implies that each European member country has its own optimal social policy to respond (with potential negative spillover effects, see Andreozzi et al. 2017) to the same common economic framework outlined at European level[1].

This situation of institutional asymmetry contributes to generate a fragmented social structure. Indeed, because of a deficient community social frame, significant inequalities and heterogeneous re-distributive effects emerge between and within the European countries (Avram et al. 2014, Beckfield 2006, Immervoll et al. 2006, Fredriksen 2012). The financial and economic crisis of the last decade contributed then to accentuate the impact, in terms of inequalities, of this structural asymmetry: the last one prevented a unified and effective response from the European institutions to the social needs of its citizens during the crisis. Ferrera (2014) and Martinsen et al. (2014) highlight this kind of hindrance and the consequent lack of a joint reaction during the crises.

Given this current two-levels design there are three possible options for the future development of the European Union. The baseline alternative is to keep the status quo: for the European Union it is certainly possible to continue to be an institutional chimera, that is a condition where many national welfare systems coexist within a uniform European economic framework. The

1 A significant practical example is the interest rate: 19 structurally different European countries face the same interest rate settled by the European Central Bank. Stability and growth pact is another example which involves every European country.

second option relies on a never old-fashioned paradigm, that is to re-establish the symmetry between the economic and social sphere moving back to the original state of affairs. In this last perspective the European countries are supposed to take back those economic and political decision-making powers they have gradually ceased in favour of the European Union: in the light of the recent Brexit case this option is not an abstract case study for professional thinkers anymore. The third and the last hypothesis is suggested by Maduro (2000), Sangiovanni (2013) and Vandenbroucke (2013). It looks at the opposite direction of the previous one and embraces the idea to fill up the institutional gap shifting in a certain degree the European Union from being a mere economic infrastructure towards a reality more careful to the social dimension of its citizens. A natural consequence of this perspective is then the necessity to endow the European Union with some concrete social welfare tools and with specific dedicated resources[2].

The reasoning suggested in the paper shows how the constitutive elements which currently characterize the European Union substantially imply the third option. The conclusion that the European Union, given its current configuration, should move towards a stronger social integration, reducing its inner institutional asymmetry and then the underlying inequalities, is reached through the adoption of a peculiar perspective[3]: the Rawlsian social contract theory and its impartial mechanism of the veil of ignorance (Rawls 1999). The reasoning which follows can be essentially split into two main blocks:

- the first step consists in identifying, within the European Union, those elements which are sufficient and necessary to apply Rawls's social contract theory. In particular, I will show the existence, from an empirical point of view, of two fundamental elements at European level: *a)* a mutually advantageous cooperation among its member states, *b)* a set of formal institutions which constitute a basic structure.

- it is then necessary to linger the attention on the direct consequences deriving from the application of Rawls's social contract theory to the European Union, that is the European difference principle. Furthermore I will introduce a corollary of the major conclusion: if European citizens

2 The European Union does not totally lack of any social view. The European Union currently has some (thin) social traits (see Buchanan 1996, p 253, Dluhosch 1996, pp. 338-339, Kölling 2015, p. 86, Streit et al. 1995, p. 319 and p. 338, Vaubel 1996, p. 317). However its redistributive policy so far has been modest and mostly driven by reasons of pure economic compensation. Thus the adopted criteria for the limited redistribution of the limited resources mostly depended on bargaining powers of the single countries rather than on some explicit social purposes (Maduro 2000).

3 Different perspectives and methodologies might be adopted to derive the principles for the European social justice (Dunaiski 2013, Rawls et al. 2003, Sangiovanni 2013, Scharpf 2002, Van Parijs 2012a, Viehoff 2016).

want their institutions to reflect the restributive social policy embodied in the difference principle scheme, they might implement Fiscal Union at European level[4].

At this stage of the reasoning the locution “Fiscal Union” might be misleading. Depending on the context in which is adopted it can reference to different meanings, suggest multiple aims, usually derived from the broader field of public finance (Musgrave et al. 1989), and more generally it can be implemented in different degrees and can be characterized by different specific elements (Fuest et al. 2012). For instance (European) Fiscal Union might be realized with the unique aim to provide some specific (European) public goods, like a common military defence. Fiscal Union can suggest also a concept of shared and binding rules concerning the tax policy within a set of defined economic and political entities. Another possible interpretation implies a shared pool of resources aimed to face some systemic risks which can be managed better together than from an isolated point of view.

All these representations are certainly relevant when they are associated to the European context. However, to avoid misunderstandings, it is important to specify from the beginning which exact meaning is given to the expression “Fiscal Union” throughout the next pages: hereafter with the words “Fiscal Union” it will be meant a European system which can pool together into a common budget the resources necessary to pursue and to implement a Rawlsian redistributive social policy within the European Union. Neither economic factors (specific tax policies, exact amount of the common budget, etc.) nor political equilibria (legitimization, bargaining rules, decision-making powers, etc.) will be taken in consideration if not in an accessory way.

The next Sections are then organized as follows. Section 1 briefly introduces the methodological issue of approaching the European Union integration from the perspective of John Rawls's domestic social contract theory: The Appendix further examines in depth this methodological issue. Section 2 is descriptive and focuses the attention on those European empirical elements which allow the adoption of Rawls's social contract theory. Section 3 provides the main outcome of applying Rawls’s social contract theory to the European Union, that is the European difference principle. Section 4 lingers on a corollary of the European difference principle, that is the European Fiscal Union. The Conclusions summarize the main ideas, provide some final remarks and address the future research.

4 Within the context of the present paper Fiscal Union at European level is not to be interpreted as a direct outcome of the analysis of the European Union in a Rawlsian perspective, but simply as a corollary. The straightforward outcome of the analysis is the European difference principle, which embodies the purpose of social integration; Fiscal Union is only one of the possible means to achieve that purpose.

2.1 John Rawls's social contract theory and the European Union

John Rawls (1999) conceived an impartial procedure to identify those principles (norms) which can guarantee a fair institutional arrangement at national level and at international level (Rawls 2001). In developing his theories Rawls proposing a contractual procedure to establish the main principles which are supposed to lead human society and its institutions. Within the Rawlsian theory the impartiality in the choice of the norms is achieved through the veil of ignorance, a tool which excludes the access to any particular information to the parties involved in the agreement. Furthermore, although the whole decision-making process and the agreement are conceived by Rawls as purely hypothetical, their ideal derivation has precise concrete effects. Indeed, even tough the veil of ignorance is supposed to be only a mental experiment, the agreed principles of justice (norms to shape institutions) have prescriptive effects in the real world after the veil (again, hypothetically) is dropped.

In particular, at national level (Rawls 1999) the first principle establishes the implementation for single individuals of a scheme of liberties as broad as possible and compatible with the liberties of everybody else; the second principle of justice, relabelled by Rawls himself as "difference principle", requires to redistribute those resources achieved by means of social and economic cooperation in order to maximize the expectations of the (group of) individual(s) most disadvantaged. As far as the international framework is concerned Rawls (2001) lists eight distinct principles (Rawls 2011, p.37) which are supposed to regulate the relationships between countries in a fair frame. Thus, according to Rawls's social contract theory, the main institutions of a modern society must be arranged to fulfil as much as possible the prescriptions of the principles of justice impartially chosen behind the veil of ignorance.

Given this theoretical framework, when we try to approach the European Union in the light of Rawls's international theory we bump into some difficulties: not only the Rawlsian international setting leads to some ordinary conclusions with regards to the European Union[5]; in addition, the task is made even more arduous by the fact that John Rawls appears to be eurosceptic, and this should prevent any attempt to analyse further in depth the issue. Indeed, it seems to be natural to interpret some of his explicit references to the European Union as symptoms of a more or less marked euroscepticism: "one question the Europeans should ask themselves, if I may hazard a suggestion, is how far-reaching they want their union to be" (Rawls et al. 2003 p. 9); "the large open

5 See the Section 1.3.1.2.

market including all of Europe is aim of the large banks and the capitalist business class whose main goal is simply larger profit" (Rawls et al. 2003 p. 9).

Thus, basing their analysis on the quoted claims and on the idea that Rawls was skeptical toward transnational economic practices and institutions that weaken the independence and the democratic content of single domestic societies, the previous (rare) attempts to interpret the European Union in the light of Rawls's thought directly or indirectly provided an unfavourable exegesis of the European project.

Kamminga (2014) insists on rejecting the idea to interpret the European Union from any Rawlsian perspective, domestic or international, because according to the author in both cases the European Union lacks of some fundamental elements to apply Rawls's theories. In particular, in Kamminga's opinion, the European Union does not have the structure of a Rawlsian domestic society because at the European level the mutually advantageous cooperative relations occur between states rather than between individuals and because the Union lacks of a clear single political constitution. At the same time, the European Union does not meet the requirements of the Rawlsian international theory because the this accepts liberal as well as decent people in the Society of People, while the Union, accepting as members only liberal society, is too selective and demanding. In this perspective, the European Union is too intolerant according to Rawls's international standards.

Barcelos et al. (2008) assume a different perspective. In their opinion the European Union is an unidentified political object, characterized by a mix of national and international elements. Therefore we cannot approach the European Union with the standard Rawlsian theories, because neither Rawls's international theory nor his domestic justice squarely fit a hybrid and continuously evolving institution like the European Union. They conclude that "given this hybrid nature of the Union, the description of its values by analogy with the domestic society [...] is, therefore, unacceptable […] This same hybridism, in the same way, excludes the possibility of conforming the EU to the [...] model defended in *The Law of* Peoples" (Barcelos et al. 2008, p. 9).

Using a numerical example Morgan (2008) claims the implicit contradictions of a European (between states) redistributive principle with the national (within states) redistributive policies. That consideration apparently prevented Rawls from endorsing the European project, because he was aware of those possible contradictions.

Therefore, relying on the eurosceptic interpretations we are lead to infer that the European countries should avoid adopting any common fiscal system for redistributive purposes, since, apparently also according to Rawls himself, when we deal with any remarkable project of European

integration cons prevail on pros[6]. However, the conclusion about John Rawls's euroscepticism has to be considered too hasty for one important reason: in his international theory Rawls only marginally took into consideration the European framework and more in general the European Union is not structurally contextualized within his works. In other words, Rawls neither conceived a specific European social contract theory nor he explicitly included the European Union as a formal object of any of his writings. This theoretical emptiness, more than his few explicit (non positive) references concerning the European integration, paved the way to the eurosceptic interpretations.

In this paper I sustain the idea that it is possible to infer some positive and innovative results applying the Rawlsian social contract theory to the European Union. In particular, adopting Rawls's domestic (national) theory (Rawls 1999) it is possible to achieve some Euro-optimist outcomes, more precisely concerning the redistribution of resources at European level.

According to the structure given by John Rawls to his domestic theory (Rawls 1999) two ingredients are sufficient and necessary in order to trigger a redistributive issue within a domestic framework:

a) a strong social and economic cooperation among the involved parties, which generates benefits and conflicts. Indeed Rawls opens his main work with the following words:

> "although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed" (Rawls 1999, p. 4).

b) a solid basic structure (Rawls 1977 and 1999), meant as "society's main political, social and economic institutions, and how they fit together into one unified scheme of social cooperation" (Rawls 1987, p. 3). Said otherwise, the basic structure refers to "the way in which the major social institutions fit together into one system, and how they assign fundamental rights and duties and shape the division of advantages that arises through social cooperation" (Rawls 1977, p. 159).

6 This conclusion appears to be even more true if we consider that Rawls himself explicitly refused to derive any kind of international redistributive principle (Rawls 1993, 2001).

Without these two components it is not possible to apply the domestic theory. Therefore, now it is also easier to understand why Rawls refused to conceive a redistributive principle valid at international level. Indeed, within Rawls's international theory the two mentioned elements are completely missing or more simply ignored: there is not (at least for Rawls) a cooperation between peoples qualitatively similar to the one between individuals belonging to a closed system (Beitz 1999, pp. 132-143, Freeman 2006 p. 39 and Rawls 2001, pp.117-118[7]); there is not (at least for Rawls) any clear and specific international basic structure like the one required for redistribution at domestic level (Buchanan 2000, pp. 700-701, Freeman 2006, p. 39 and Pogge 2003, p. 1741). Therefore the derivation of a redistributive principle is neither required nor formally conceivable since there are not those minimal structural conditions which trigger redistributive concerns.

Freeman is quite straightforward in making this kind of interpretation explicit: "the idea of social cooperation [...] is central to Rawls's account of social justice. It underlies his distinction between "domestic justice" and the Law of Peoples. Moreover, the idea of social cooperation informs Rawls's account of the difference principle. What makes social cooperation possible for Rawls are the basic institutions that constitute 'the basic structure'" (Freeman 2006, p. 38). Therefore without the presence of both those elements (cooperation and institutions) at international level, moral redistributive dilemmas simply do not emerge and applying the standard version of the Rawlsian pure procedural justice (Rawls 1999) essentially becomes superfluous, if not wrong (Freeman 2006, p. 61).

Here I assume that together with the two mentioned elements all the other Rawlsian "circumstances of justice" (Rawls 1999, pp. 109-111), that is "the normal conditions under which human cooperation is both possible and necessary", are completely fulfilled as well. However, I deliberately avoid entering the debate on which further elements in the literature of political theory are considered essential to make redistributive concerns emerge and then to justify principles of redistributive justice and redistributive institutions, like for example a certain degree of coercion (Nagel 2005 or Blake 2001). Two are the reasons for this precautionary choice.

First of all the topic is quite recent and the debate on coercion is very animated and still open (Blake 2016, Sangiovanni 2016 and Valentini 2011). Trying to follow it would uselessly complicate the theoretical framework necessary for the analysis of the European Union in a Rawlsian

7 In the examples provided by Rawls in The Law of People (Rawls 2001) there is not any sort of social or economic cooperation between the two societies taken as reference. The two peoples are presented as economically isolated and independent one from each other, that is they are basically designed as autarkies. See also (Martin 2015, p. 748).

perspective. Therefore here it is preferable to stick as much as possible to the basic Rawlsian framework.

Second, and even more importantly, Rawls never makes explicit that the role of the principles of justice is to protect individuals from coercive institutions that could harm somebody. Indeed, within his main work Rawls (1999) uses the words "coercion" and "coercive" only 20 times, the first in §32 at page 177. So, coercion does not seem to be a fundamental element of the Rawlsian theory. Thus, even though Rawls seems to admit that "it is reasonable to assume that even in a well-ordered society the coercive powers of [institutions] are to some degree necessary for the stability of social cooperation", however "the establishment of a coercive agency is rational only if these disadvantages are less than the loss of liberty from instability" (Rawls 1999, p. 211). Nevertheless, the idea that shines through Rawls's words is that social institutions should be structured to avoid coercion as much as possible, and in case coercion is introduced for some reasonable motivation, the degree of coercion has to be limited and clearly defined.

The next Section aims at showing that the two main elements mentioned above, that is cooperation and a system of institutions, are present within the European framework, therefore it is possible to apply the interpretative categories of Rawls's domestic theory and to draw the related conclusions.

2.2 The European Union: economic cooperation and basic structure

Is was showed that in order to apply Rawls's domestic theory to the European Union it is necessary to verify, from an empirical point of view, whether or not the European Union holds the two requirements mentioned above, that is the economic cooperation scheme and an institutional basic structure. However, before moving in that direction it is useful to highlight how it is beyond the aims of the present analysis to enter the debate on what exactly the European Union is (a federation, a confederation, an association of compound states, see Buchanan 1996 and Blankart 2007), how the powers within the Union are or should be balanced (Vaubel 1996 and Vaubel 1997), or how its institutions are or should be legitimized. The existence of certain structural elements is independent from how we technically prefer defining the European Union. The main intent of the

next paragraphs is simply descriptive, that is they aim to provide the empirical evidence of those elements[8] which allow to apply Rawlsian domestic justice and its categories at European level.

As for the economic and social interaction meant as a cooperative venture for the mutual advantage (which generates benefits and conflicts), it is not difficult to acknowledge similar scheme of cooperation within the European Union. Following Beitz's insight about the effects of globalization (Beitz 1999, pp. 143-153) it can be immediately noticed how the "international economic interdependence constitutes a scheme of social cooperation" (Beitz 1999, p. 154) as exactly as meant by Rawls for a simple national (closed) system. As regards the specific case of the European Union and from the point of view of constitutional economics[9] Beitz's insight is even more convincing: the economic integration process which had begun with the Treaty of Rome (1957, Title I and Title III) gave birth to a formal scheme of mutual cooperation which can be easily interpreted in the Rawlsian sense. The European Economic Community (Single Market) with its free circulation of goods, persons, services and capitals constitutes a clear example of social and economic scheme of mutually advantageous interdependence. Of course, as suggested by Beitz, similar kind of cooperation exists even at a broader international level, but that is not relevant with regards to the current analysis.

As far as the benefits generated by the European economic cooperation are concerned, although the economic theory does not agree about the permanent or temporary effects of a market integration (Badinger 2005), the positive economic outcomes of a market enlargement are broadly recognized since Adam Smith's Wealth of Nations (Smith 19994), where he grasped the positive implications of the size of a market on the division of labour, and then on the productivity through specialization. The literature is not unanimous about the precise quantitative effects derived from the European economic integration (Badinger et al. 2011) and its determinants are not always completely clear (Campos et al. 2014 and König et al. 2012). Nevertheless many studies agree on how the European countries have benefited from the institution of the common market institution (Badinger et al. 2011, even though the authors highlight how most of the studies are more ex-ante predictive analysis rather than ex-post quantitative investigations, p. 308).

Despite different methodologies and a "quantitative disagreement", the following remarks are sufficient to highlight some of the benefits gained from the European common market (economic

8 Those two (Rawlsian) elements are empirical assumptions it is possible to disagree about, as clearly explained by Blake (2012, p. 122-126).

9 According to Buchanan the "constitutional economics examines the choice of constraints as opposed to the choice within constraints" (1991, pp. 134-135). Furthermore it is important to remark how the same author (Buchanan 1991, p. 141) explicitly claims how the Rawlsian distributive problem is an object of study of the constitutional economics.

cooperation). Over the period 1950-2000 the "European integration has significantly contributed to the post-war growth performance of the current EU member states" such that "GDP per capita of the EU would be approximately one-fifth lower today if no integration had taken place" (Badinger 2005, pp. 73-74). "EU membership has had a positive and asymmetric effect on long-term economic growth" on the EU-15 member states (Crespo Cuaresma et al. 2008, p. 652). In addition "there seems to be strong evidence on positive pay-offs from EU membership, despite considerable heterogeneity across countries" and in a prudent counterfactual evaluation "incomes would have been around 12 per cent lower today if European Integration had not happened" (Campos et al. 2014, p. 25 and p. 21). What should be clear from these empirical instances and from the framework explained so far is that the overlap between the Rawlsian concept of a venture for the mutual advantage and the European Economic Community (Single Market) is straightforward: the Treaty on the common market formally defined the mutually advantageous venture[10] meant in a Rawlsian sense. It is then an empirical task to measure the exact economic surplus gained from the free European market integration.

As for the second element considered essential in order to apply Rawls's theory of domestic justice, it is not difficult to identify within the European Union a set of common institutions and agencies which, in the logic of the present writing, can be interpreted as a European basic structure. The Treaty of Lisbon (Article 13) formally lists seven common institutions[11] whose tasks are to provide political direction, to manage the Union and to integrate the conflicting interests (Peterson et al. 2012): elective European Parliament, European Council, Council (of Ministers), European Commission, Court of Justice of the European Union, European Central Bank and Court of Auditors. Together, they exert the legislative, the executive and the judicial powers with the aim to define European policies.

The seven institutions are then surrounded by hundreds of agencies and organizations[12] (Mathieu 2016) which, performing sometimes at the limits of their formal powers (Chamon 2016), operate in accordance with the guidelines of the main institutions mentioned above. These agencies, together with the main institutions, affect individuals' prospects of live in different spheres, ranging from ensuring an area of freedom, security and justice (for example Frontex, the European Border

10 Of course it is not necessary a formal treaty for the existence of mutually advantageous economic and social relationships. Nevertheless a formal treaty is an additional element which strengthens the Rawlsian domestic interpretation of the European Union.

11 It is not among the aims of the present analysis to provide a detailed description of the main European institutions, nor, as stated in advance, enter in a debate regarding the equilibria between them or their legitimization.

12 For a complete map of the agencies see the official website https://euagencies.eu/.

and Cosat Guard Agency) to supervising financial systems (for instance EBA, European Banking Authority), from providing defense (EDA, European Defence Agency) to supporting EU business and innovation in the digital, energy, innovation and transport sectors, from directly fostering citizens' well-being, like for example through the European Centre for Disease Prevention and Control (ECDC) to helping the developing countries to exploit the potential of their human capital through the European Training Foundation (ETF).

In short, the main European institutions and all the agencies which surround them constitute a dense institutional network which gives rise to a European basic structure. However, some remarks are due about the set of European institutions interpreted as a Rawlsian basic structure. First of all in a Rawlsian perspective the quantity of institutions (on the European territory) is not relevant. Instead, what matters is the substantive quality of those institutions: they are supposed to be capable of affecting, either by themselves or in conjunction, the distribution of duties and rights, that is to affect people's prospects of life.

Considered in this perspective the main European institutions, together with their derivatives, can effectively and concretely "distribute fundamental rights and duties and determine the division of advantages from social cooperation" (Rawls 1999, p. 6). They can deeply affect Europeans' plans of life. This conclusion can be reinforced with some concrete examples: the European Parliament is elected by all the European citizens, so it constitutes a direct link between the European institutions and the people living on the European territory, and a ban on pesticides voted by the European Parliament effectively redistributes duties and rights between European citizens. A sentence of the Court of Justice of the European Union can directly and radically affect the prospect of life of any (group of) European citizen(s) in case the national laws conflict with the European ones. Again, the European Central Bank, setting the interest rate, through the financial and credit institutes can concretely and effectively redistribute the benefits of the European economic cooperation.

However, beyond these concrete examples, it is of fundamental importance to remark three points. First of all a basic structure is not a binomial (zero or one) outcome, which either does not exist at all or which exists through its full configuration. Instead, a basic structure is an arrangement which spans on a continuous spectrum and which, according to Rawls, should tend to one, up to the point that "even in a [perfectly] well-ordered society, adjustments in the basic structure are always necessary" (Rawls 1977, p. 164) in order to maximize the expectations of the worst-off.

In other words the current European institutional framework represents a configuration which should be considered "just throughout, but not the best just arrangement" (Rawls 1999, p. 68). This means the European basic structure, despite being evidently incomplete compared to the national

ones, has to be constantly improved to tend as close as possible to its full representation. The idea of moving towards a stronger social basic structure, which can affect the prospects of life of European citizens by redistributing the benefits of the economic cooperation in a more incisive way than the actual European institutional arrangement, is the exact attitude of the present work.

Second, it is relevant to underline how the European basic structure is not constituted by mere second-side institutions which have the only aim of fostering the national ones (Blake 2013, pp. 108-132). The European institutions act also with their own tools, goals and values, often independent from the interests of the single nations. They go from promoting the peace within its territory to defending its external borders, from reinforcing the economic and social cohesion to fostering the sustainable development, from safeguarding the cultural diversity to promoting the welfare of all its citizens. This specific feature of autonomy makes the European institutional framework closer to the one designed by Rawls for his domestic theory.

Third and last specification, it is important to highlight how another international basic structure as qualitatively complete as the European one cannot be identified beyond the European boundaries nor at any other international level. Therefore the uniqueness of the European institutional arrangement plays a fundamental role in the possibility of interpreting it from the perspective of the Rawlsian domestic theory.

In conclusion, the European economic cooperation and the European basic structure represent the essential prerequisites to apply Rawls's domestic social contract theory to the European Union. The next Section focuses on the outcomes of the European social contract in a Rawlsian perspective following a reasoning by analogy: the derivation of the European difference principle as normative redistributive rule at European level and its corollary, that is European Fiscal Union.

2.3 The European difference principle

Given that the European Union holds the essential empirical elements for the application of the Rawlsian domestic theory it is possible to adopt its formal categories of interpretation. Said otherwise, it is now possible to make explicit which principles are conceivable for the European Union as a whole in the perspective of the Rawlsian social contract. For a reasoning by analogy,

European individuals behind the veil of ignorance[13] would substantially agree on the standard (national) principles (Rawls 1999, pp. 52-56)[14] to shape the European institutions.

Despite being derived through a domestic original position, the principles this time substantially become valid across the European member states, regardless of the national borders. Indeed, since the Europeans are deprived by the veil of ignorance of any particular information, overall of that specific information concerning the territory where they (might) live in terms of resources, size of the population, boundaries, economic development and so on and so forth (Rawls 1999 pp. 32-33) they would choose exactly the same two principles conceived by Rawls for the standard domestic case (1999, pp. 52-65, pp.130-139 and pp. 153-160) and they would rationally decide to apply them across the European Union considered as a whole.

Indeed the European original position is structurally the same as for the domestic procedure, and in the specific case of pure procedural justice, the same decision-making structure means the same outcome, that is the same sort of principles. As mentioned in the section 1.3.1.1, it is true that Rawls did not come to the same conclusions when formulating his international social contract theory. At international level he proposed different principles from those domestic. However, it should be remarked how he derived different principles because the structural conditions he considered valid for the two cases were different. Instead, I have shown that at the European level that there is a strong equivalence with the Rawlsian domestic framework.

Thus, the content of the principles chosen behind the European veil of ignorance remains exactly the same as for the typical national “closed system”, but the actual range of their application is identified according to the considered basic structure (extension of cooperation), in this case the European Union. The outcome is rigid with regards to the essence of the principles (same conditions, same outcome), while the range for their application is tailored on the European territory. Again, the cut-off point (Martin 2006, pp. 227-234 and Martin 2015, p. 749) for the application of the (domestic) principles in a European perspective has to coincide with the European Union as a whole. Indeed, behind the (European) veil there are no rational reasons to decide to apply the principles within the boundaries of the single member countries, and this is particularly true for the difference principle which affects the redistribution of resources. Behind the veil it is irrational to apply the principles within specific countries: in a European perspective any ex-ante

13 See (Rawls 1999, pp. 118-123).

14 "The first principle simply requires that certain sorts of rules, those defining basic liberties, apply to everyone equally and that they allow the most extensive liberty compatible with a like liberty for all" (Rawls 1999, p. 56), whereas the second principle, the so called difference principle requires to "maximize the expectations of the least favored position" (Rawls 1999, p. 69).

(behind the veil) decision which identifies a specific internal boundary or a territorial limit for the application of the principles has to be considered arbitrary and against the maximin reasoning. Thus, the range of validity of the European principles must coincide with the extension of the institutions (cooperation) considered behind the veil.

Focusing now the attention on the redistributive issue, it is possible to claim how behind the veil of ignorance European citizens would agree to arrange the social and economic inequalities to the greatest expected benefit of the least advantaged European individuals[15], regardless the country or the nationality: that is equivalent to enunciate a European difference principle which operates at individual level[16] across and beyond the boundaries of the single European member states. A normative analysis of the European redistributive issue in a Rawlsian (domestic) perspective leads to shape the European institutions so that the resources generated by the social and economic cooperation within the Union must be redistributed to favour the least advantaged (in terms of social primary goods[17]) Europeans: the difference principle is not "statist" in its assumptions anymore (Kuper 2000, pp. 653-654) and it basically becomes transnational (even if derived from the application of a national theory).

To reinforce the reasoning presented so far it is possible to provide an introductory example concerning the working mechanism of the European difference principle. To start, imagine three European countries (Histogram 1) which act in isolation and which arrange their inner inequalities in such a way that the least advantaged[18] in country A reaches an index of social primary goods of 2 points, the worst-off in B reaches an index of 7 points and the worst-off in country C an index of 5 points. This is the typical situation conceived by John Rawls, where every country substantially

15 See (Rawls 1999, pp. 69-72).

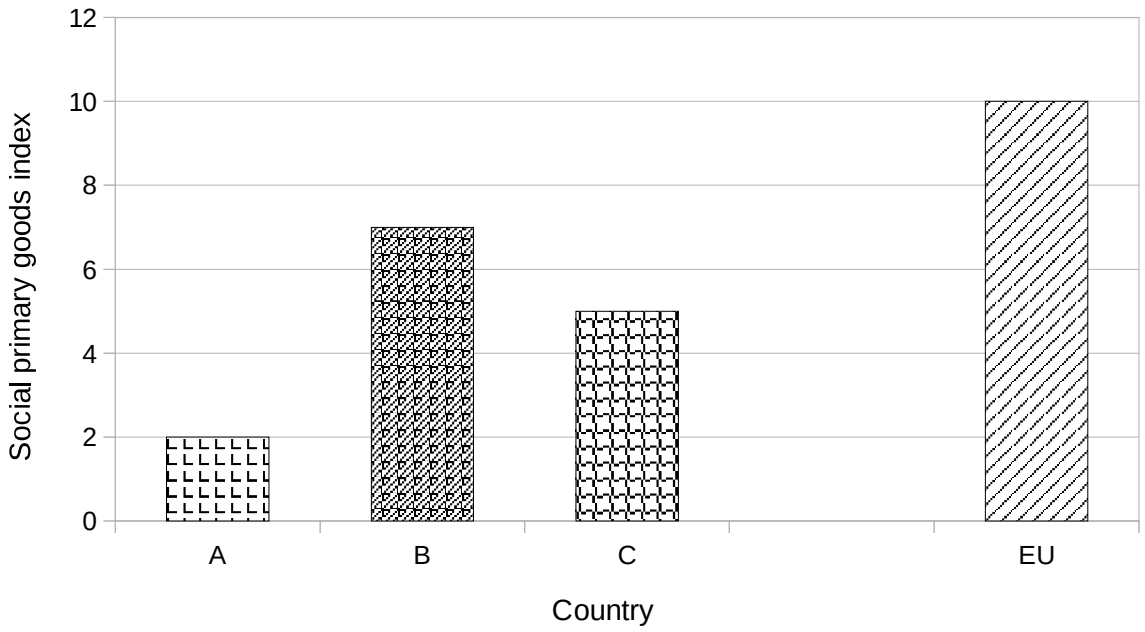
16 This specification seems to be necessary because we might be tempted to think that, since the actors at the European level are basically nation-states, the rationale should be to redistribute toward the worst-off member state instead of the least advantaged (group of) individuals between states. However, this hypothesis violates the assumptions of the Rawlsian domestic social contract theory (which takes individuals as actors) and generates a clear paradox: if we assume redistribution between member states, in the perspective of a difference principle at European level we should redistribute resources from Greece (higher GDP) to Luxembourg (lower GDP), or from Poland (higher GDP) to Norway (lower GDP), although Luxembourg and Norway are two countries among those with the highest GDP per capita.

17 The social primary goods "are things which it is supposed a rational man wants whatever else he wants" (Rawls 1999 p. 79), and the social primary goods index avoids problems of comparability (Rawls 1999, pp. 78-81). Furthermore, the existence of a redistributive principle between states may lead to the problem of redefining the Rawlsian bundle of social primary goods (Rawls 1999, pp. 78-81), as very cleverly grasped by Paden (1997, pp. 226-227).

18 Of course also those groups of individuals who are better-off should be taken into consideration. However, the reppresentation of those groups in the provided examples would not add anything to the general concept.

worries about the distribution of its own resources and countries do not consider endorsing any agreement to share or to redistribute their own social primary goods beyond their boundaries. At the same time Histogram 1 adds a further hypothesis: if the three countries operate not only as autarkies, but they cooperate together (that is their citizens constitute an international venture for the mutual advantage), they can generate a common surplus (EU) of 10 points, since they would profit from their (international) comparative advantages.

Histogram 1 – Worst-off per country and EU surplus

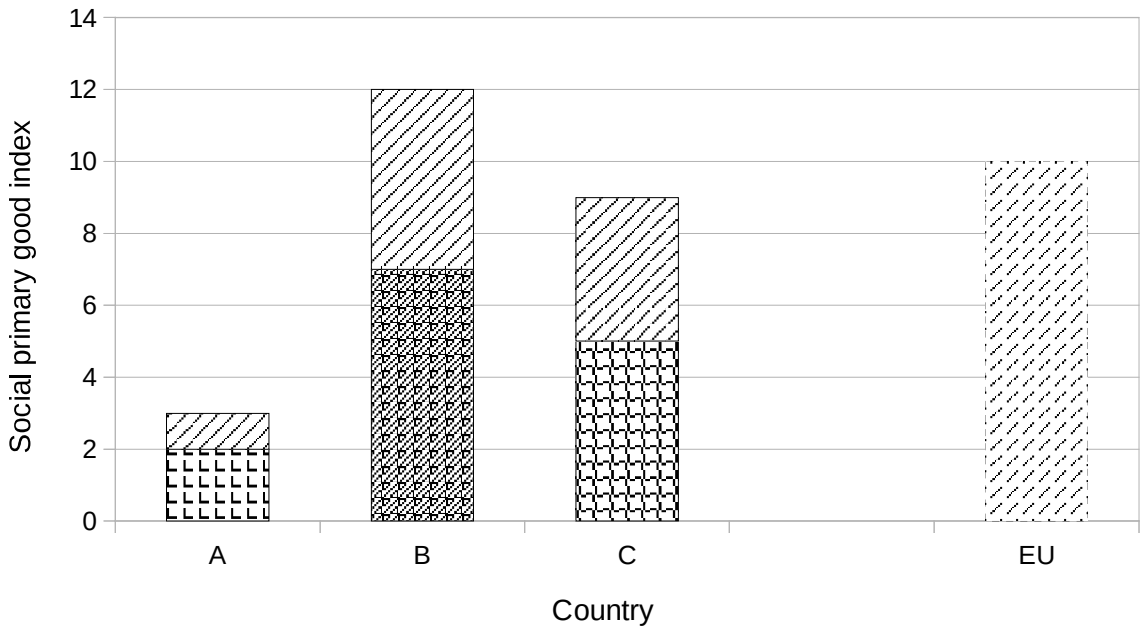


HISTOGRAM 1: Acting in isolation (and according to their domestic difference principle) the three European countries A, B and C can maximize the expectation of the least advantaged of their own nations as shown by the respective coloured histograms. Should the individuals of the three countries cooperate together they could generate a common surplus, EU.

Now imagine a second situation. Individuals in the three European countries engage in a mutually advantageous cooperation and, exploiting their comparative advantages, they generate a surplus at European level. However there is no any clear redistributive scheme, or said otherwise, European citizens do not agree on any redistributive principle which prescribes how to share the surplus EU. It is then plausible to assume how the surplus is arbitrary divided by market forces or by the bargaining powers of the single nations (Histogram 2). In this situation, where there is not a formal redistributive principle or where there is not a set of institutions (basic structure) which can concretely redistirbute the common surplus EU, the social primary goods index of the worst-off

among the least advantaged (located in country A) is assumed to improve by only 1 point. In this situation the worst-off in country A reaches an index of social primary goods of 3. The total index increases up to 12 for the least advantaged in country B and up to 9 in country C. However, "the distribution resulting from voluntary market transactions (even should all the ideal conditions for competitive efficiency obtain) is not, in general, fair". And this outcome is possible, Rawls specifies, "even though nobody acts unfairly".(Rawls 1977, p. 160).

Histogram 2 – Market division of the EU surplus



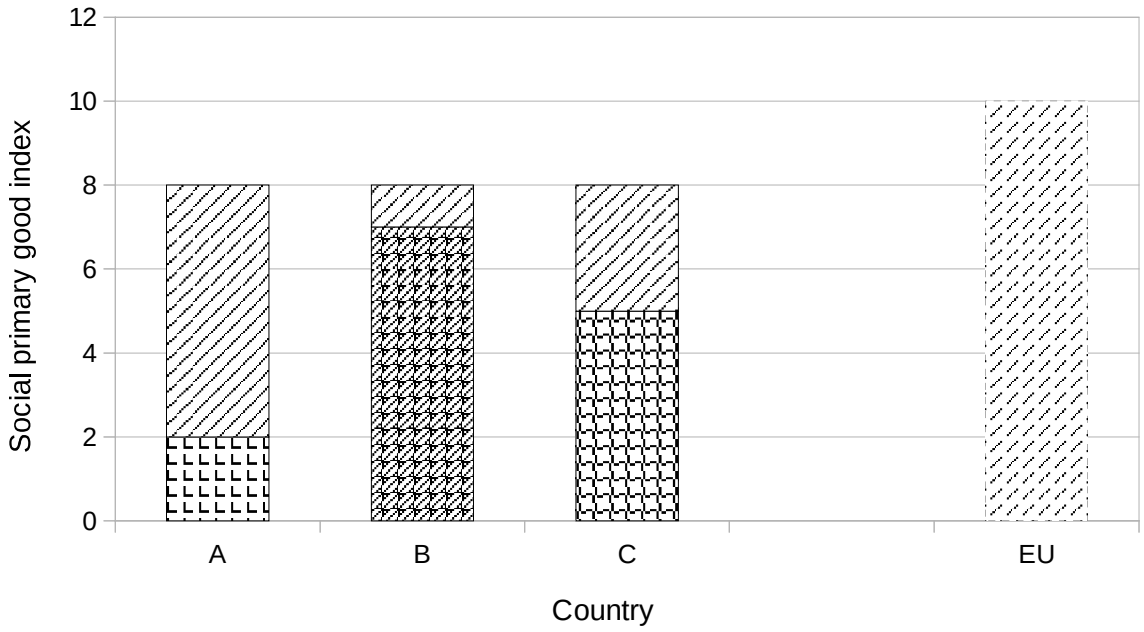
HISTOGRAM 2: Cooperating at European level without reaching an agreement on how to distribute the surplus EU or without having a basic structure leaves the redistribution at the mercy of other forces.

Instead, how is the European difference principle supposed to act? The initial situation is the same as in the second example, with the three countries that engage in mutually cooperative relationships. This time however, the European surplus EU is supposed to be redistributed in order to benefit as much as possible the worst-off between the three countries (considered as a unique set). Histogram 3 shows an egalitarian distribution consistent with the European difference principle: all groups get a total of 8 points in terms of social primary goods: the index of the least advantaged is thus maximized. In short, the redistributive scheme proposed in Histogram 3 is to be considered as the natural outcome of the European difference principle.

As for the Histogram 3 is concerned a further clarification is due. The egalitarian distribution proposed in the example is the ideal outcome of the maximin rule – it is "a perfectly just scheme" where "[n]o changes [...] can improve the situation of those worst off" (Rawls' 1999, p. 68) - but it is not the only distribution coherent with the difference principle prescriptions: unequal distributive schemes are possible and allowed as long as they lead to the maximum advantage of the worst-off (Rawls' 1999, pp. 65-73).

In other words, the pure egalitarian solution is just one of the possible configurations admitted by the difference principle. Paradoxically, also the market division of the Histogram 2 might be consistent with the difference principle if the other available distributive options could not increase the expectations of the least advantaged group of European citizens. Indeed, in the Rawlsian social welfare function weight one is assigned to the most disadvantaged and weight zero to everyone else, so inequalities are not a primary concern. This means that the index of social primary goods of the least advantaged is the only variable to take into account in the maximization process. From this it follows that the egalitarian solution is just a possibility of this one and not a pre-settled goal.

Histogram 3 – Division of the EU surplus according to the difference principle



HISTOGRAM 3: Egalitarian distribution of the EU surplus according to the European difference principle, which maximizes the expectations of the worst-off between those who are the least advantaged within the single countries.

2.4 European Fiscal Union in a Rawlsian perspective

Beyond the theoretical analysis provided so far, it remains unclear how the surplus EU should be quantified and how the resources generated by means of the European cooperation should be physically pooled together in order to be then redistributed according to the European difference principle. As for the first issue (quantification) the answer seems to be quite straightforward, because the estimate of the European surplus is matter for an empirical investigation. Following the studies presented in Section 2 (Badinger 2005, Badinger et al. 2011, Campos et al. 2014 and Crespo Cuaresma et al. 2008) the European surplus might be calculated as the counterfactual difference between the current level of the European economic activity and the (aggregated) hypothetical economic level should the European countries act as pure autarkies, that is with no transnational cooperation with other European member states.

Even if the empirical estimate of the European surplus is not a theme regarding the present theoretical paper, one final remark about this measure is considered essential: the European difference principle, such as conceived, does not have to substitute the national redistributive mechanisms, because it is not supposed act on all the available resources within the European Union (both national and common). Instead, it prescribes to redistribute exclusively that surplus generated by the social and economic cooperation which takes place at European level, that is the redistribution has to be realized only on those resources which are the product of the European cooperative scheme. Conversely, the resources generated by the cooperation which takes place exclusively within the single countries, that is without any European interdependence, remain immune from the communitarian redistributive policy. Thus the European difference principle is not meant as a cosmopolitan distributive principle (Beitz 2000) for the European Union, because the national surpluses are to be redistributed according to the domestic redistributive schemes.

In general it is not so immediate to think about a compelling rationale for redistributing across Europe resources generated by the social and economic cooperation which takes place exclusively within the boundaries of the single member states. Indeed, paradoxically, applying a unique European redistributive policy on the total surplus (national ones and European considered together) might erode the social primary goods index of some individuals who are better-off within the single countries with no common redistributive scheme.

Morgan (2008, p. 9) provides a clear numerical example which highlights such a possibility: implementing a unique difference principle between nations might lead to a contradicting situation where the worst-off individual of some particular state is made more disadvantaged compared to the

situation where a difference principle is applied within every single country with no redistribution between states. Thus, with a unique over-arching distributive mechanism there is the risk that a "society, or group of such societies, [are] worse off under a [European] basic structure than in perfect isolation" (Pogge 1988, p. 249).

However, the situation designed in this paper through the European difference principle is different from Morgan's (2008) example. Indeed, in the present analysis the redistribution of resources can be implemented only according to the specific level of cooperation: within-country (domestic) redistributive rule on those resources generated by the national cooperation and between-countries (European) difference principle on the surplus generated by means of the European cooperation. This is because the European basic structure, even though it is interpreted from the perspective of the domestic theory, does not substitute the national ones, but it constitutes, together with the system of cooperation, a second level framework where redistribution of resources can take place. This becomes even more evident if we consider that not all the social primary goods on the European territory are the outcome of the European cooperation, therefore a two-level institutional framework is the situation considered more representative of the European case.

In conclusion there is not any constrain to redistribute resources only once at the highest level, like stated by Freeman (2006, p. 63)[19]. Redistribution according to the difference principle can take place twice or more times, for any level of cooperation and where a basic structure is available. Thus, national and European redistributive schemes are not exclusive or conflicting, but rather compatible and in some degree complementary.

As far as the exact European redistributive mechanism is concerned, I assert how the redistribution of the European surplus according to the European difference principle might be obtained implementing Fiscal Union at European level. As anticipated in the Introduction, in the present context Fiscal Union is to be meant merely as a system which can pool together into a common budget the resources generated by the European economic cooperation. To comply with the European difference principle a common European budget, reflecting the European surplus, becomes essential. Otherwise, without Fiscal Union, which pools the resources together, the European difference principle remains an elegant normative and theoretical outcome with no concrete perspectives.

It is not then an aim of the present paper to enter the debate concerning the practical (technical or political) implementation of European Fiscal Union. Nevertheless it seems to be legitimate to

19 In this perspective the following consideration becomes then false: "[t]he difference principle can apply only once to structure economic and property institutions, either globally or domestically. It cannot apply to both." (Freeman 2006, p. 63).

promote some general considerations. In particular, there are two possible ways to interpret European Fiscal Union in the light of the European difference principle. On the one hand it is possible to think about a system which constantly collects and transfers the European surplus in order to maximize the expectations of the European worst-off. On the other hand it is also imaginable a sort of mechanism of insurance (Bénassy-Quéré et al. 2016 and Thirion 2017) which pools together the resources and which acts against systematic risks. In the latter perspective the European surplus might be collected in order to protect the weakest European parties in case of specific unfavourable conditions[20], or the budget might get into action when a set of European subjects goes below a certain minimum threshold.

About the way to implement concretely Fiscal Union at European level, Rawls himself provides some hints in §43 (Rawls 1999, pp. 242-251) when he describes the distributive branch. He states how one element of the "distribution branch is a scheme of taxation to raise the revenues", and which "make[s] the transfer payments necessary to satisfy the difference principle" (Rawls 1999, p. 246). Thus, a system of taxation at European level should be considered an essential part of European Fiscal Union in a Rawlsian perspective. Rawls then goes even further suggesting some specific taxes which might be adopted to generate the resources required for the redistribution (insurance mechanism) according to the difference principle: inheritance and gift taxes; proportional expenditure tax, that is tax on consumption; a proportional tax on annual consumption; an income tax is considered as well. It is even possible to speculate about a tax on those activities which make business across national borders (but within the Union), and so on and so forth.

Clearly these ideas deserve a further deepening, since it is also necessary to consider that an improper tax scheme might generate frictions which nullify the benefits of the common market for some countries. However, those specific lucubrations are considered beyond the research question of the paper.

Final remarks and conclusions

The main conclusions of the analysis concerning the European Union in the perspective of the Rawlsian social contract theory can be summarized as follows: the current constitutional elements which characterize the European Union framework imply a precise redistributive scheme embodied

20 Rawls himself, between the background institutions for distributive justice, mentions a "stabilization branch" (Rawls 1999, p. 244).

in the European difference principle which is supposed to act only on the European surplus. A corollary that follows that conclusion concerns European Fiscal Union, which represents a possible way to implement such a redistributive scheme and therefore to complete the European basic structure.

In reaching these conclusions, we have to take into account the difficulties to interpret Rawls's thought regarding the European Union, as he hardly ever lingered on the topic. However, we can be confident in the formulation of the European difference principle for two specific reasons. On the one hand we have to understand the fact that the European Union experienced by Rawls was very different from the today's Union. Although John Rawls was probably right in showing some eurosceptic traits in imagining a European Union based only on mere (socially empty) economic evaluations, the current Union is constituted by specific institutional elements which allow us to move beyond the mere functional economic structure. It is not longer possible to have a European market, many European supranational institutions, a European Parliament, a European common currency but not a European system of welfare redistribution: such an institutional asymmetry unavoidably creates unjustified inequalities because of a "redistributive bias on national policy choices" (Scharpf 1998, p. 6).

On the other hand we should consider that Rawls himself left the possibility for a difference principle at European level open. He clearly stated how there is "room for various forms of cooperative associations and federations among peoples" (Rawls 2001, p. 36), making explicit then the following hypothesis: "what does the Law of Peoples say about the following situation? Suppose that two or more of the liberal democratic societies of Europe, say Belgium and the Netherlands, or these two together with France and Germany, decide they want to join and form a single society, or a single federal union [...] A voter [behind the veil of ignorance] might vote for the difference principle (the most egalitarian liberal conception)" between the two states (Rawls 2001, p. 43, note 53).

To conclude, the normative analysis provided throughout the paper applies Rawls's theory of domestic justice to the European Union. The European Union concretely holds those elements, the scheme of mutually advantageous cooperation and the parallel basic structure thanks to which we can conceive a European difference principle that requires to maximize the expectations of the European least advantaged. The concrete implementation of similar social welfare function requires to pool together those resources which are generated by means of the cooperation taking place at European level: this goal might be achieved by means of European Fiscal Union. In this way the Rawlsian redistributive scheme and Fiscal Union at European level can contribute to reduce the

institutional gap between the European economic integration and the European social integration. Instead, the evaluation of the exact amount of the European surplus and the specific redistributive mechanism are matter for a further research on the topic.

Appendix: elucidations on the application of Rawls's domestic theory to the European Union

The reasoning to apply Rawls's domestic theory to the European Union deserves some further specifications. Indeed, although the deduction of the European principles is essentially consistent the Rawlsian domestic justice, at the same time the analysis implicitly "borrows" a couple of ideas usually belonging to cosmopolitan justice[21], that is:

- in order to regulate the redistributive issue between societies we should not reason from the perspective of collective entities like peoples or states, but from the point of view of single individuals;

- the principles of justice, and in particular the redistributive principles, should be applied as broadly as possible, and potentially at global level.

As far as the first the cosmopolitan idea is concerned, it is worth recalling how for Rawls (2001) the parties who are required to choose the international principles are representatives of peoples and not single individuals. However this approach to international justice adopted by Rawls raised many critical reactions, including some concerns regarding the "dramatic adaptation of the [second order] original position" (Pogge 1988, p. 235 and Pogge 2006, p. 206). For instance, Buchanan (2000, pp. 698) claims how the device of having the representatives of peoples is inadequate because "to say that the parties represent peoples is, in effect, to ensure that the fundamental principles of international law that will be chosen reflect the interest of those who support the dominant or official conception of the good or of justice in [a domestic] society" (Buchanan 2000, pp. 698). Said otherwise, considering peoples as the only relevant moral subjects in order to derive international principles implies that the voice of minorities or in general of those who might disagree is not taken adequately into account[22].

A further critical aspect is emphasized by Beitz, who sustains that to define peoples as the only relevant moral entities at international level implies that we still do not provide any indication on how to regulate justice between single individuals who belong to different peoples (Beitz 1999, p. 132): Rawls's setting "obscures the fact that the interests of persons and peoples do not necessarily coincide" (Kuper 2000, p. 246), leading in this way to "potentially sub-optimal results

21 "Cosmopolitan principles tend thus to have two parts: a defense of a particular metric of fair distribution, and an argument as to why this metric should not be limited to any particular subset of humanity" (Blake 2012, p. 127).

22 From Buchanan's consideration it follows that Rawlsian international justice basically lacks of that pluralism which Rawls (2001, p. 40) tried to embody in his theory at the cost of a byzantine structure.

for persons" (Kuper 2000, p. 247) because "the interests of individual persons [across the different peoples] are taken into account only indirectly" (Beitz 2000, pp. 673-674 and Nagel 2005, p. 134).

Another critique in the same direction is provided by Pogge, and it is related to the ambiguity in the definition of the notion of peoples. Pogge believes that the conceptual role assigned by Rawls to the term "people" is inadequate, because "in many parts of the globe, official borders do not correlate with the main characteristics that are normally held to identify a people [and] whether some group does or does not constitute a people would seem, in important ways, to be a matter of more-or-less rather than either-or" (Pogge 1994, p. 197). In other words, there is not a clear-cut distinction between the notion of peoples and other ways of grouping individuals. Therefore using peoples' representatives instead of other categories should be considered as an arbitrary choice. The same kind of perplexity is shared by Van Parijs who states that "a key issue that immediately arises is who the peoples are" because all over the world "there are over 3000 living languages and only 212 sovereign countries to accommodate them" (Rawls et al. 2003, p. 10). Thus, it is problematic to identify in the real word precise sets of individuals who constitute the corresponding peoples.

Taken all together these criticisms can be summarized in the following statements:

- at international level we have interests even as single persons and not only as individuals belonging to specific groups;
- grouping individuals might exclude somebody's claims or interests; it is not completely clear which exact criteria should be adopted to identify specific groups (peoples);
- there is a concrete possibility to end up with different principles of international justice given different ways of grouping individuals under the notion of peoples;
- lastly, it might be problematic to derive a "whole family of original positions, each corresponding to one dimension of our [moral] identities" (Van Parijs 2012, p. 643 and Nagel 2005, pp. 141-142).

In applying Rawls's domestic theory to the European Union all these perplexities are avoided, because behind the European veil of ignorance it is assumed the perspective (the direct interests) of single European individuals[23], even if the geopolitical context is international. Together with the mutually advantageous cooperation and the European basic structure, the individual perspective is a further element which concurs yo the derivation of a European distributive principle of domestic nature.

As for the second (cosmopolitan) idea, anyone familiar with the topic, in the claim that the principles should be applied as broadly as possible, and potentially at global level, can substantially

23 This is also coherent with the requirements of the Rawlsian domestic theory.

recognize Beitz's theory of global justice (1999, pp. 127-176), whose "intuitive idea is that it is wrong to limit the application of contractarian principles of social justice [only within] the nation-state; instead, these principles ought to apply globally" (Beitz 1999, p. 128). In particular, Beitz's main achievement built on the Rawlsian (domestic) social contract theory is a global difference principle[24]: in Beitz's opinion "the difference between citizens and foreigners is not morally significant [...] The fact of global interdependence combined with Rawls's *Theory of Justice* leads Beitz to enunciate a principle of international redistribution of wealth so as to maximize the position of the least advantaged person in the world" (Arnopoulos 1981, p. 193).

However Beitz's reasoning, which blindly extends the principles of the Rawlslian domestic justice at global level, has to be take into account some fair limitations. His reasoning, despite faultless in theory, clashes in practice, because a normative (redistributive) principle not feasible in the real world is to be considered useless. In particular it is not possible to agree on the global range of Beitz's difference principle because there is not a global basic structure qualitatively similar to the basic structure conceived by Rawls for a closed domestic system. Although some authors claim the opposite (Buchanan 2000, pp. 705-706), at global level there is not a set of international institutions which can effectively redistribute fundamental rights and duties and the benefits generated by the global economic cooperation. In this perspective "the limits of what is empirically plausible are worth understanding as part of political [and moral] philosophy" (Blake 2016, p. 319). "[I]f there is [or not] possibility of doing justice with the [international] institutions we have" (Blake 2016, p. 318) has to be taken into account by the theory. Thus, Beitz (1999, p. 128) and other cosmopolitan authors' conclusions are incautious insofar they do not endorse the existence of global conditions (institutions) for the application of redistributive principles at a global level (see Blake 2012, pp. 127-128 and Pettit 2006, p. 107).

As showed in the paper, instead, within the European Union there are the structural conditions for the application of an analog of the difference principle valid between the European member states. Thus, for the European Union it is possible to conceive a common redistributive scheme which applies beyond the national ones, which takes into consideration the interests of the single European individuals and which at the same time remains feasible from a practical point of view. That is basically the idea to realize a cosmopolitan view for a circumscribed union of states:

24 A similar conclusion is reached also by Paden: "the delegates [in the] international original position would choose two principles of justice that are similar in form to the principles chosen by the parties to the domestic original position" (Paden 1997, p. 226).

"cosmopolitan justice could be realized in a federal system, in which the members of individual nation-states had special responsibilities toward one another that they did not have for everyone in the world" (Nagel 2005, p. 120 and Pogge 2001 p. 248).

References

Andreozzi, L. & Tamborini, R. (2017). Why is Europe engaged in an inter-dependence war, and how can it be stopped? DEM Working Paper N. 2017/06. Available at SSRN: https://ssrn.com/abstract=2965766.

Arnopoulos, P. (1981). Political Theory and International Relations Charles R. Beitz Princeton: Princeton University Press, 1979, pp. 200, Canadian Journal of Political Science, 14(01), 192-193.

Avram S., Levy H. & Sutherland H. (2014). Income redistribution in the European Union. IZA Journal of European Labor Studies, 3(1), 22.

Badinger, H. (2005). Growth effects of economic integration: evidence from the EU member states. Review of World Economics, 141(1), 50-78.

Badinger, H., & Breuss, F. (2011). 14 The quantitative effects of European post-war economic integration. International handbook on the economics of integration: factor mobility, agriculture, environment and quantitative studies, 3, 285.

Beckfield J. (2006). European integration and income inequality, American Sociological Review, 71(6), 964-985.

Barcelos, P. & Queiroz, R. (2008). Which Values for the European Union? Europe between Principles of Domestic and International Justice. Published on http://www.ifilnova.pt.

Beitz, C. R. (1999). Political theory and international relations, with afterward. Princeton University Press.

Beitz, C. R. (2000). Rawls's law of peoples. Ethics, 110(4), 669-696.

Bénassy-Quéré, A., Ragot, X. & Wolff, G. B. (2016). Which fiscal union for the euro area? Bruegel Policy Contribution ISSUE 2016/05, 1-17.

Blake, M. (2001). Distributive justice, state coercion, and autonomy. Philosophy & public affairs, 30(3), 257-296.

Blake, M. (2012). Global distributive justice: Why political philosophy needs political science. Annual review of political science, 15, 121-136.

Blake, M. (2013). Justice and foreign policy, Oxford University Press.

Blake, M. (2016). Agency, coercion, and global justice: a reply to my critics. Law and Philosophy, 35(3), 313-335.

Blankart, C. B. (2007). The European Union: confederation, federation or association of compound states? Constitutional Political Economy, 18(2), 99-106.

Buchanan, A. (2000). Rawls's law of peoples: Rules for a vanished Westphalian world. Ethics, 110(4), 697-721.

Buchanan, J. M. (1996). Europe as social reality. Constitutional political economy, 7(4), 253-256.

Campos, N. F., Coricelli, F., & Moretti, L. (2014). Economic growth and political integration: estimating the benefits from membership in the European Union using the synthetic counterfactuals method.

Chamon, M. (2016). EU agencies: legal and political limits to the transformation of the EU administration, Oxford University Press.

Crespo Cuaresma, J., Ritzberger-Grünwald, D. & Silgoner M. A. (2008). Growth, Convergence and EU Membership, Applied Economics, 40(5), 643-656.

Dluhosch, B. (1997). Convergence of income distributions: another measurement problem. Constitutional Political Economy, 8(4), 337-352.

Dunaiski, M. (2013). Principles of distributive justice within the EU. Published on http://www.e-ir.info/, available at: http://www.e-ir.info/2013/04/05/principles-of-distributive-justice-within-the-eu/

Ferrera, M. (2009). The JCMS annual lecture: National welfare states and European integration: In search of a 'virtuous nesting'. JCMS: Journal of Common Market Studies, 47(2), 219-233.

Ferrera, M. (2014). Social Europe and its components in the midst of the crisis: a conclusion. West European Politics, 37(4), 825-843.

Freeman, S. (2006). The law of peoples, social cooperation, human rights, and distributive justice. Social Philosophy and Policy, 23(1), 29-68.

Fredriksen, K. B. (2012). Income inequality in the European Union. OECD Economic Department Working Papers, No. 952.

Fuest, C. & Peichl, A. (2012). European Fiscal Union: What Is It? Does It Work? And Are There Really'No Alternatives'?. In CESifo Forum (Vol. 13, No. 1, pp. 3-9). München: ifo Institut–Leibniz-Institut für Wirtschaftsforschung an der Universität München.

Immervoll, H., Levy, H., Lietz, C., Mantovani, D., O'Donoghue, C., Sutherland, H. & Verbist, G. (2006). Household incomes and redistribution in the European Union: quantifying the

equalizing properties of taxes and benefits. In The distributional effects of government spending and taxation (pp. 135-165). Palgrave Macmillan, London.

Kamminga, R. M. (2014). Rawls and the European Union. Published on philica.com. Article number 425

Kölling, M. (2015). How much solidarity is in the EU budget?. Perspectives on Federalism, 7(3), 77-97.

König, J. & Ohr, R. (2012). The European Union–A heterogeneous community? Implications of an index measuring European integration. Universität Göttingen, mimeo.

Kuper, A. (2000). Rawlsian global justice: Beyond the law of peoples to a cosmopolitan law of persons. Political theory, 28(5), 640-674.

Maduro, M. P. (2000). Europe's social self: 'The sickness unto death', in Shaw J. (ed). Social Law and Policy in an Evolving European Union (pp. 325-349), Hart Publishing, Oxford.

Martin, R. (2006). Rawls on international distributive economic justice: taking a closer look, in Martin R. & Reidy D. (ed.), Rawls's Law of Peoples: A Realistic Utopia? (pp. 226-242), Oxford, Blackwell.

Martin, R. (2015). Rawls on international economic justice in the law of peoples. Journal of Business Ethics, 127(4), 743-759.

Martinsen D. S. (2013), Welfare states and social Europe, in U. Neergaard, E. Szyszczak, J. W. van de Gronden, & M. Krajewski (ed.), Social Services of General Interest in the EU (pp. 53-73), chapter 3, The Hague: TMC Asser Press..

Martinsen, D. S. & Vollaard, H. (2014). Implementing social Europe in times of crises: Re-established boundaries of welfare?. West European Politics, 37(4), 677-692.

Mathieu, E. (2016). Regulatory delegation in the European Union: networks, committees and agencies. Springer.

Morgan G. (2008), John Rawls: eurosceptic? European integration as a realistic utopia. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1113223.

Musgrave R. A. & Musgrave P. B. (1989), Public finance in theory and practice, McGraw-Hill Internationa Edition.

Nagel, T. (2005). The problem of global justice. Philosophy & public affairs, 33(2), 113-147.

Paden, R. (1997). Reconstructing Rawls's law of peoples. Ethics & International Affairs, 11, 215-232.

Peterson J. & Shackleton M. (2012). The institutions of the European Union, Oxford University Press.

Pettit P. (2006). Rawls' peoples, in Martin R. & Reidy D. (ed.), Rawls's Law of Peoples: A Realistic Utopia? (pp. 38-55), Oxford, Blackwell.

Pogge, T W. (1988). Rawls and global justice. Canadian Journal of Philosophy, 18(2), 227-256.

Pogge, T. W. (1994). An egalitarian law of peoples. Philosophy & Public Affairs, 23(3), 195-224.

Pogge T. W. (2001). Rawls on international justice, The Philosophical Quarterly, 51(203), 246-253.

Pogge, T. W. (2003). The incoherence between Rawls's theories of justice. Fordham L. Rev., 72, 1739.

Pogge T. W. (2006). Do Rawls's two theories of justice fit together?, in Martin R. & Reidy D. (ed.), Rawls's Law of Peoples: A Realistic Utopia? (pp. 206-225), Oxford, Blackwell.

Rawls J. (1977). The basic structure as subject. American Philosophical Quarterly, 14(2), 159-165.

Rawls J. (1987). The idea of an overlapping consensus, Oxford J. Legal Stud., 7, 1.

Rawls, J. (1993). The law of peoples. Critical Inquiry, 20(1), 36-68.

Rawls J. (1999). A theory of justice, revised edition. Harvard University Press. Cambridge.

Rawls J. (2001). The law of peoples: with the idea of public reason revisited, Harvard University Press.

Rawls J. & Van Parijs P. (2003), Three letters on the Law of Peoples and the European Union. Autour de Rawls, 7-20.

Sangiovanni, A. (2013). Solidarity in the European Union. Oxford Journal of Legal Studies, 33(2), 213-241.

Sangiovanni, A. (2016). Is coercion a ground of distributive justice?. Law and Philosophy, 35(3), 271-290.

Streit, M. E., & Mussler, W. (1995). The economic constitution of the European community: from 'Rome'to 'Maastricht'. European Law Journal, 1(1), 5-30.

Scharpf, F. W. (1998). Interdependence and Democratic Legitimation, working paper, Max Planck Institute for the Study of Societies, Cologne.

Scharpf, F. W. (2002). The European social model. JCMS: Journal of Common Market Studies, 40(4), 645-670.

Thirion, G. (2017). European fiscal union: Economic rationale and design challenges. CEPS Working Document No. 2017-01/January 2017, Archive of European Integration.

Vandenbroucke, F. I. G. (2013). A European social union: why we need it, what it means. Rivista Italiana di Politiche Pubbliche, 2, 221-247.

Valentini, L. (2011). Coercion and (global) justice. American Political Science Review, 105(1), 205-220.

Van Parijs P. (2012). No Eurozone without Euro-dividend, Unpublished Manuscript. URL: http://www.fondationpaulricoeur.fr/uploads/medias/articles_pr/no-eurozone-without-eurodividend.pdf.

Vaubel, R. (1996). The constitutional future of the European Union. Constitutional Political Economy, 7(4), 317-324.

Vaubel, R. (1997). The constitutional reform of the European Union. European Economic Review, 41(3-5), 443-450.

Viehoff, J. (2017). Maximum convergence on a just minimum: A pluralist justification for European Social Policy. European Journal of Political Theory, 16(2), 164-187.

3. Neither Punishments nor Rewards: Fostering Tax Compliance Through the Rawlsian Veil of Ignorance in a Laboratory Experiment

Klaudijo Klaser and Luigi Mittone

Abstract: It is well known that different deterministic mechanisms (like formal audits and material punishments) can stem free riding behaviour in social dilemmas. However the behaviouralist literature shows us how several other environmental and psychological variables can influence agents' attitude to cooperate. By means of a repeated tax compliance game run in an experimental laboratory, our study measures the effects of a Rawlsian veil of ignorance on cooperation over time. In particular we found that in our experimental design the (laboratory) veil of ignorance has an effect both on the ex-ante distribution of votes concerning the adoption of a specific tax regime and on the ex-post tax compliance level between treatments, but not on compliance across rounds, which shows to be decreasing.

JEL Code: D02, D63, H26

Keywords: Experimental Economics, Inequality, John Rawls, Tax Compliance, Veil of Ignorance

"Models of tax evasion need to take into account that taxpayers may not only want to maximise their interests, however defined, but also desire to see justice and fairness realised"

Wenzel

Introduction

In the early 70's Allingham and Sandmo (1972) proposed a utility function which aimed to explain individual tax compliance as a risky portfolio choice based on only two exogenous parameters, the probability of being audited and the fine amount in case of ascertained misbehaviour. However, the proposed normative model proved to be insufficient to accurately describe agents' observed tax behaviour, that is their basic theoretical framework could not predict in a satisfying way tax payers' positive choices.

In response to Allingham and Sandmo's limited approach and in order to better understand agents' choice whether or not to abide by the tax law, the following literature on tax evasion focused the attention on other behavioural variables of psychological, procedural and environmental nature (Andreoni et al. 1988, Braithwaite 2017, Jackson et al. 1986, Feld et al. 2007, Kirchler 2008, Pickhardt et al. 2014, Richardson 2006 and Tolgler 2002).

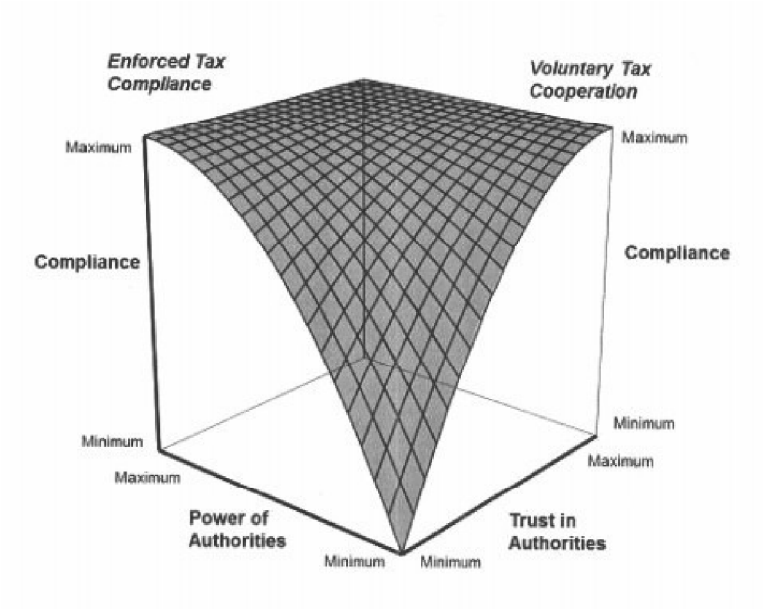
Some of the variables which have been recognized to influence agents' tax behavior are: endogenous participation in the definition of tax rules (Tyran et al. 2001); perceived fairness of the tax system (Becker et al. 1987 and Bordignon 1993); audits' sequences in a repeated framework (Kastlunger et al. 2009 and Mittone 2006); having a voice on the destination of the tax revenue (Casal et al. 2016b and Pommerehne et al. 1996); influence exerted by public opinion (Casal et al. 2016b and Kahan 1997); ethical concerns (Alm et al. 2011 and Feld et al. 2002); social norms (Wenzel 2004 and Wenzel 2005).

A recent theoretical advance tried then to summarize and to conciliate in a unique conceptual framework the standard economic variables like audit probabilities and fines (gathered together under the label "power of authorities") with a broad set of behavioural and environmental elements (labelled as "trust in authorities"[1]) which have been recognized to impact on tax compliance

1 In this paper the locution "trust in authorities" might not precisely coincide with the authors' original meaning. Here "trust in authorities" is meant in its broadest sense, that is as any psychological or environmental element that can enhance voluntary tax compliance.

decisions. The so called "slippery slope" (Figure 1) is "a conceptual tool [which] may serve to understand the importance of determinants of tax behaviour" (Kirchler et al. 2010, p. 214 and Kichler 2008).

Figure 1 – Slippery slope geometrical representation



In particular, the "slippery slope" curve is conceived as a surface within a three-dimensional geometrical space. The degree of tax compliance, the dependent variable, is measured on the vertical axis, while the two sets of forces which are recognized to drive the decision to abide by the tax law are placed on the horizontal plane.

One axis includes those standard coercive tools which can mechanically enforce compliance ("power of authorities"). The other horizontal axis gathers together all those behavioural elements ("trust in authorities") which cannot be controlled in a deterministic way but which can influence voluntary tax compliance (Muehlbacher et al. 2011).

Within the "slippery slope" framework formal and informal forces, that is power and trust, are conceived to be jointly-responsible in determining the degree of tax compliance[2]. Nevertheless, and without questioning the importance of the reciprocal influence of the two mentioned groups of

2 The two sets of forces are recognized to interact dynamically with each other (Filippin et al. 2013, Gangl et al. 2015 and Kirchler et al. 2010). In other words, the two set of variables are not conceived as secluded or independent. Instead they reciprocally influence each other, and they can either enter a positive symbiotic relationship, mutually reinforcing each other and then pushing together the tax payer towards full tax compliance, or coercive powers and (mis)trust can enter a spiral where they have a negative impact on each other, inducing the economic agent to evade more taxes.

forces, it is important to highlight how according to the analytical representation given to the "slippery curve", the possibility to achieve full tax compliance by means of one single set of variables is also admitted.

The study carried out in the paper goes in the direction of the just sketched intuition. It takes into consideration the theoretical possibility to obtain full tax compliance by means of one single set of forces contemplated by the literature in general and by the "slippery slope" framework in particular. Indeed, while it is quite immediate to imagine some cases where full tax compliance might be obtained by means of coercive tools ("power of authorities")[3], it is not obvious to conceive a frame where full tax compliance is achievable as a pure voluntary mechanism, not enforced by external constraints.

In particular we inquire the latter option through an experimental methodology. Indeed, a controlled environment like an experimental laboratory allows to exclude authorities and any coercive tool which usually enforce mechanically tax compliance from the tax game design[4]. In this way, from a game generally structured as taxpayer vs. tax authority, we move to a tax game framed in the form of taxpayer vs. taxpayer (Pickhardt et al. 2014). Thus, without any authorities or any other kind of exogenous coercive mechanism, the tax compliance game basically becomes a voluntary contribution mechanism where the vertical "trust in authorities" assumes the form of a horizontal "trust in other tax p(l)ayers". The specific framework for the experimental design takes into consideration two specific fields of literature mainly based on empirical outcomes.

On the one hand we move from the literature which provides a compelling evidence of a causation effect between individuals' participation in the definition of tax rules (rates, audit probabilities, fines, destination of the tax revenue, etc.) and their following level of tax compliance (Alm et al. 1993, Alm et al. 1999, Casal et al. 2016a, Feld et al. 2002, Feld et al. 2006, Pommerehne et al. 1996 and Wahl et al. 2007). In other words, when people have a concrete voice in tax issues a "participation effect" emerges (compared to a situation where the same variables are exogenously assigned, Bortolami 2009 and Bortolami et al. 2009). By and large, rules and institutions

3 It is sufficient to imagine a case where the audit probability is settled to p=1 (Feld et al. 2002), or a case where to stimulate cooperation (read compliance) strong rewarding or punishing institutions are introduced (Fehr et al. 2002, Gintis 2005, Gürerk et al. 2006 and Sefton et al. 2007). However informal institutions deserve particular attention, because it has to be taken into account that positive effects on compliance are sometimes counterbalanced by negative effects in terms of average returns (Kroll et al. 2007 and Masclet et al. 2003). Indeed, heterogeneity in beliefs can lead to normative conflicts concerning the right behaviour to adopt (Nikiforakis et al. 2008, Nikiforakis et al. 2012 and Sefton et al. 2007), such that swords without words can be worse than words without swords (Ostrom et al. 1992).

4 Artificially excluding one set of variables automatically prevents any possible dynamic interaction between the two set of forces contemplated by the "slippery slope", see the footnote number 2.

legitimized in a direct way enhance the cooperative attitude (Dal Bó et al. 2010), therefore also tax compliance (Feld et al. 2000).

On the other hand we rely on the exclusion game design (Sacconi and Faillo 2005). The branch of literature based on the exclusion game (Degli Antoni et. al 2016, Faillo et al. 2008, Faillo et al. 2014, Sacconi et al. 2005, Sacconi et al. 2010, Sacconi et al. 2011 and Tammi 2011) studies the effects of an impartial and non-binding agreement on a one-shot resource allocation game. The agreement is structured in the perspective of Rawls's social contract theory and its outcomes are interpreted consistently with his idea of the sense of justice (Rawls 1999).

The novelty of the present research is the adoption of some elements belonging to the exclusion game to inquire the effects of an agreement of Rawlsian type, that is reached behind a veil of ignorance, on tax behaviour. In particular the sense of justice might be one of those psychological forces which, inducing compliance to the agreement on a chosen tax regime, meant to represent a distributive scheme, directly generates voluntary tax compliance.

Thus, excluding tax authorities and adopting a laboratory veil of ignorance to choose a tax regime, the research aims to inquire tax compliance behaviour as a pure voluntary mechanism within a Rawlsian framework. The engage of Rawls's social contract theory and the related experimental literature in the tax evasion field has further interesting implications beyond testing the two just sketched hypotheses.

First of all, it provides an empirical test for the "slippery slope" theoretical shape. In particular, if it is not possible to achieve voluntary compliance by means of any combination of environmental and psychological variables, the "slippery slope" surface should be revisited and drawn as an asymmetric curve[5]. Of course we test only one of these "authority free" frames, but our experiment might be considered a further step on analysing the conditions which can foster voluntary tax compliance, and the "slippery slope" theory should take into consideration similar empirical evidence.

Second, the main aim of inquiring tax evasion should be to find out those tools which can help to prevent tax evasion, and the veil of ignorance becomes a potential candidate. Thus, even though the veil mechanism could be hardly implemented in the real word, understanding which structural conditions foster voluntary tax compliance might have relevant policy implications regarding the configuration of tax rules and their compliance through different procedures of legitimation. The experiment we propose engages with this topic.

5 Furthermore, this revision might have a relevant impact on the dynamic between the two forces mentioned above.

Third, fiscal policies and tax laws, as well as the Rawlsian theory have explicit redistributive aims and effects. This overlap might help to calibrate the configuration of the tax regimes given that we want to obtain a specific outcome concerning the redistribution of resources within a particular set. In particular, our experiment focuses the attention on the redistributive scheme which maximizes the expectations of the least advantaged.

Fourth, it might be possible to extend the Rawlsian concept of the sense of justice to the more familiar concept of civic duty (Orviska et al. 2002), according to which citizens can be collaborative even if the system allows non-compliance, so that their behaviours do not have to be regulated by external audits or sanctions ("powers of authorities"), but by their own concern for society and institutions. Thus, voluntary tax compliance would not be linked narrowly only to the Rawlsian theory, but it might be reinterpreted under a more general frame. This is a value added of our experimental framework.

Last but not least, our experimental design allows to extend the one-shot approach of the exclusion game to the context of repeated games. Indeed (paraphrasing Torgler 2002, p.665) a "serious limitation [of the exclusion game] is the nature of [the] experiment, which [is] static" (only one round) while "the decision to [comply] or not is a dynamic rather than a static problem". Therefore (tax) compliance is not an atemporal single decision, but it is rather a set of choices over time, and a game in a repeated frame can shed light on the compliance dynamic after the veil of ignorance is dropped.

The next Sections are organized as follows. Section 1 explains the theoretical framework on which the experiment is designed and it formulates the predictive hypothesis. Section 2 describes in detail the experimental design. In Section 3 data from the experiment are analysed and discussed. Appendix shows the instructions provided to the experimental subjects of the veil treatment.

3.1 Rawls's social contract theory and theoretical predictions

Our tax compliance game takes the well established "participation effect" (Bortolami 2009 and Bortolami et al. 2009) for granted and makes a step aside. In particular the study focuses on the voting procedure itself, measuring the effects of two distinct voting conditions. One of them is a standard voting procedure: all players are assigned an income level and then they are asked to vote on the tax regime (set of tax rates for the different levels of income) they prefer adopting during a second phase, the actual tax compliance game.

The second voting mechanism is instead hinged on the Rawlsian social contract theory and on some of its recent experimental applications. In particular this treatment adopts a laboratory veil of ignorance during the voting phase, that is players are asked to vote for a tax scheme without knowing their personal income level. They become aware of their position in the distribution of wealth only after they have reached an agreement on a tax regime. In order to better understand the differences between the two treatments and the expected specific consequences of the veil of ignorance on compliance it is necessary to recall some further details concerning Rawls's social contract theory (1999).

John Rawls opens his *A Theory of Justice* (Rawls 1999) with the following statement:

> "although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts [while] there is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed" (Rawls 1999, p. 4).

In order to decide on the distribution of benefits generated by socio-economic cooperation, Rawls suggests adopting an impartial perspective termed "veil of ignorance". This mechanism of pure procedural justice guarantees that people unanimously agree on fair principles for the society's main institutions, because the veil of ignorance "excludes the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices" (Rawls 1999, p. 17).

Specifically, according to Rawls, behind the veil of ignorance

> "no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism" (Rawls 1999, p 118).

Therefore, in the ignorance condition, none of the involved parties can design principles which might favour his or her own particular person.

On the contrary, according to Rawls, the impartial reasoning behind the veil of ignorance is supposed to induce the involved parties to assume the perspective of the worst possible scenario and therefore to design distributive principles which aim to "maximize the expectations of the least favored position" (Rawls 1999, p. 69). Thus, with the second treatment of our tax game we simulate the impartial procedure offered by the veil of ignorance.

More in detail, in our tax game we design more or less progressive tax regimes which have different distributive effects on the lowest level of income. Rawls's theory suggests that a veil of ignorance in the voting phase should influence the individual choice of the tax scheme to adopt in the compliance phase - compared to the baseline treatment, where the players can vote according to their interest represented by the position in the wealth distribution.

In the veil treatment we ask the participants to agree on a scheme of tax rates before letting them know the income bracket they will belong to during the compliance task. Basically, in the veil condition, while voting for a specific tax regime the players are deprived of the particular information concerning their place in the distribution of wealth (within the game). Therefore nobody can profit from any specific information concerning her or his own wealth status within the game to propose, that is to vote for a redistributive scheme which mainly benefits their particular person[6].

Instead, according to Rawls, behind the veil of ignorance the players should enter a maximin perspective and vote for the tax regime which maximizes the expectations of the worst-off[7], represented in our game by the player with the lowest income level. Thus, since the alternative tax regimes that the players have to vote about have different material consequences on the wealth of the worst-off, we can formalize the first hypothesis we aim to test with our experiment.

H1: compared to the baseline treatment (where players vote after knowing their level of income in the game), in the veil of ignorance treatment we expect to observe a shift of votes and tax regimes towards the scheme of tax rates which maximises the wealth of the least advantaged position, that is the position occupied by the player with the lowest level of income.

6 That people usually vote for tax rates that advantage their particular position it was demonstrated in other experiments, like (Esarey et al. 2012).

7 So far the evidence is mostly aginst a strong effectivness of the Rawlsian veil of ignorance (Aguiar et al. 2013, Andersson er al. 1999, Bond et al. 1991, Carlsson et al. 2003 and Frohlich et al. 1987). However we have to take into account that simulating empirically a pure and perfect veil of ignorance such as conceived by John Rawls is clearly an impossible task. Rawls's veil of ignorance excludes much more information than what can be hidden in an experimental laboratory. For example in the game the players' personal real wealth and their conception of the good remain perfectly known. Thus the subjects can be made neutral only with regard to their role in the game (perspective on specific concerns), that is the position in the income distribution.

After having introduced the decision-making model based on the maximin reasoning, Rawls (1999) dedicates a considerable part of his theory to analyse the stability of the impartial agreement. His main aim is to explain how some principles, chosen ex-ante behind the veil, can become stable ex-post in the real word, after the veil is dropped.

This kind of analysis is really important because the agreement, despite being fair, is not conceived as automatically enforced: everyone can choose to free ride, that is everyone can decide to deviate from the unanimously chosen distributive rule because this does not coincide with their own ex-post individual interests. Thus, according to Rawls, it becomes necessary to identify a force which can support and restore compliance should any tendency which induces the parties to deviate from the agreement emerge.

In dealing with this issue Rawls does not look for external enforcement mechanisms[8]. Instead, he directly looks at the involved parties and their moral psychology. In particular, in Rawls's opinion, every subject taking part in the agreement behind the veil of ignorance is expected to develop an endogenous, strong and effective desire to act in accordance with the set of the chosen principles.

Said with a Rawlsian terminology, after having reached an agreement behind the veil of ignorance every subject is expected to develop a sense of justice which can counterbalance the individual incentives to deviate from the impartial principles. Thus, by means of the sense of justice, formally based on a system of mutual expectations of compliance, the agreement and its system of principles, even if non-binding, are expected to become self-enforcing.

Through a laboratory experiment a recent field of literature tried to explore the Rawlsian egalitarian conception and in particular his idea of sense of justice. The so called exclusion game (Degli Antoni et. al 2016, Faillo et al. 2008, Faillo et al. 2014, Sacconi and Faillo 2005, Sacconi and Faillo 2010 and Tammi 2011) is a one-shot resource allocation game with a preliminary voting stage simulating an agreement behind a veil of ignorance. In other words, before the role in the game (dictator or dummy player) is revealed to the participants, they have the possibility to reach a unanimous agreement about how to share a common endowment.

Three are the most important features of the exclusion game which reflect Rawls's social contract theory. The choice of the sharing rule is taken behind a veil of ignorance, that is all players are required to unanimously vote for a distributional rule not knowing their (future) role in the actual game. In the second stage, the actual exclusion game, the players' roles are differentiated with

8 That would generate a loop of agreements, because another agreement would be necessary to legitimize those enforcing institutions, which should be enforced by other institutions requiring a third agreement and so on and so forth.

regard to their decision-making powers. Thus some participants enter the dictator role and some of them become dummy players with no voice. Last but not least, the agreement on the distributive norm of the voting phase is not binding in the second phase, that is in the actual exclusion game the players assigned to the dictator role are free to share the common endowment regardless of the agreement reached in the voting stage.

Thus, given the structure of the exclusion game and according to the standard economic theory, at the second stage every rational economic agent in the dictator position should make the choice which maximizes his or her own material payoff regardless of the specific sharing rule unanimously approved in the previous voting phase. However, the provided experimental evidence discloses how the unconstrained ex-post compliance with the ex-ante chosen distributive norms is unexpectedly high even in those cases where groups agreed on an egalitarian (counter-maximizing) distribution.

The observed behaviour was justified and explained through Rawls's concept of the sense of justice. In particular, the adopted model of social conformist preferences (Grimalda et al. 2005) takes into consideration the psychological utility (Attanasi et al. 2006 and Attanasi et al. 2008, Geanakopolos et al. 1989) that is gained by complying with the impartial agreement and that compensates dictator players for their material loss.

Consistently with the mentioned theory and its empirical evidence collected in the exclusion game we can formulate our second hypothesis

H2: in the veil treatment tax compliance will be at least as high as in the baseline treatment

This prediction deserves some further clarifications. First of all we are not making a claim about the absolute level of tax compliance in any of the two treatments, but we are rather comparing the relative value between the baseline and the treatments. Second, ours is a weak hypothesis, because we are assuming that the veil tool will not lower the level of tax compliance, but we cannot make a more precise hypothesis on the positive effect on the veil because we have neither previous empirical evidence on veiled vs. no veiled agreements nor a specific Rawls's conjectures on those two conditions (except that a no veiled agreement would not be reached).

More precisely, according to the conformist preferences model, the compliance level depends on beliefs. However, we have neither theoretical nor empirical elements to predict the effect of the veil of ignorance on players beliefs. We can only be sure it will not have a negative effect on compliance through belief since the veil of ignorance puts everybody in the same perspective,

aligning the individual expectations, but we cannot formulate a prediction on the net effect on compliance.

As for the exclusion game structure and its Rawlsian interpretation, an interpretative limitation emerges. In particular that game is structured as a one-shot compliance task. Instead the sense of justice and the related compliance behaviour, such as it is conceived by Rawls himself, are not one-shot occurrences. They are not occasional achievements, but they are rather conceived as the product of a dynamic process, self-enforcing over time.

Indeed, according to Rawls, in order to be stable "the scheme of social cooperation [...] must be more or less regularly complied" (Rawls 1999, p. 6). Indeed, in Rawls's opinion, "[o]nce a system of co-operation […] is set up and a period of uncertainty survived, the passage of time renders it more stable, given an evident intention on the part of all to do their part" (Rawls 1963, p. 291). In other words a "system in which each person has, and is known by everyone to have, a sense of justice is inherently stable [because] the forces making for its stability increase as time passes (Rawls 1963, p. 293).

This is to say that the goal to justify the dynamic concept of the sense of justice with the result of a one-shot and static game should be considered partially achieved, especially after demonstrating that testing a feedback only once is likely to produce misunderstandings both the outcomes themselves and their interpretation (Hertwig and Ortmann 2001). Thus, conceiving a game design where it is possible to repeat the actual compliance task, like our tax game, easily allows to check the path of compliance, improving in this way the theoretical and empirical interpretation of the sense of justice.

Since after the veil of ignorance is dropped compliance, driven by the sense of justice and its system of mutual beliefs, is described by Rawls as a self-enforcing process and it is expected to be more and more stable over time, and since our tax game is designed to repeat the compliance task across many rounds, it is possible to formulate the third hypothesis we aim to test with our experiment

H3: in the veil treatment the tax compliance path will be at least constant across rounds

According to all the elements described so far and in particular to the mentioned model on conformist reciprocity (Grimalda et al. 2005), the system of reciprocal beliefs, activated by the veil

of ignorance, is fundamental in order to sustain compliance[9] (Rawls 1963, Rawls 1999): the players decide to comply with the agreement, even if that means to renounce a share of their material payoff, on the condition they believes that the other players who took part in the impartial agreement will act or would have acted the same way[10].

That means that in the veil treatment compliance across rounds is expected to be linked to beliefs regarding other players' compliance. In other words,

H4: in the veil treatment the tax compliance path across rounds will be aligned with players' beliefs

The last two predictions are implicitly based on dynamic psychological equilibria, which take into consideration the update of beliefs through time (Attanasi et al. 2006, Battigalli et al. 2005).

3.2 Experimental design

At the beginning of every experimental session the participants are randomly divided into groups of three people. Single groups face then two chronologically ordered phases. In the first phase the players are asked to vote for a tax scheme to be adopted during the second phase, the actual tax compliance task. In particular in the first phase of the game each group has a maximum of 6 rounds to vote for a tax schemes. In order to access the second phase the participants are required to reach a unanimous consensus on a specific tax scheme. Those groups that do not reach a unanimous agreement about the tax scheme by the end of the 6th round cannot enter the second phase of the game and they are paid the show up fee. The unanimous agreement is therefore an essential precondition to enter the actual tax compliance game.

A tax scheme or regime is a set of different tax rates applied to three given levels of income (see Table 1). The two treatments differentiate only with regard to the voting stage: according to the treatment the income levels are assigned before (baseline treatment) or after (veil condition) the

9 Beliefs are widely recognized to have an impact on the cooperative attitude in general (Chang et al. 2004, Chaudhuri 2011, Fischbacher et al. 2010, Frey et al. 2007, Kahan 1997, Keser et al. 2000, Tyran et al. 2001)

10 The dynamic is very different from models of pure conformity, which assume we adapt our own behaviour to match others' expectations on us (Cialdini et al. 2004 and Cialdini et al. 1998).

voting consultation. The compliance phase is then identical for the two treatments under the design profile.

The experiment is designed with Experimental Currency Units (ECU). 4,000 ECU are equivalent to € 1 and the participants are aware of the exchange rate because it is explicitly mentioned in the initial instructions (see Appendix). Within each group the players are randomly assigned[11] (according to the treatment, before or after the vote) one of the following levels of income: 1,500 ECU, 2,000 ECU or 3,000 ECU.

Once assigned the endowment level keeps constant during the experiment, that is individual income does not change across rounds. Moreover, within each group the income levels are exclusive, that is it is not possible for two or three participants of a group to have the same endowment. Given the income exclusivity the initial expected income is 2.167 ECU, while the inequality of the initial distribution, measured by a simple standard deviation, is equal to 764 ECU.

Table 1 describes in detail the tax regimes which the participants are asked to vote on. For example tax scheme A tries to mimic the current tax rates applied by the Italian law on personal incomes: 23% for yearly incomes up to € 15,000, 27% for incomes ranging from € 15,000 to € 28,000 and 38% for incomes between € 28,000 to € 55,000[12]. The inequality of this tax scheme is measured by a standard deviation calculated on the final distribution (full compliance case) of 371 ECU. In other words if the players choose the tax regime A, and they decide to fully comply with it, they can reduce the initial inequality from 764 ECU to 371 ECU.

The other schemes vary in the progressiveness of the tax rates applied to the three levels of income. Tax schemes D and B are more progressive, that is they generate more equal ex-post distributions of wealth than tax scheme A. In particular tax scheme B allows to reach the most equal distribution of wealth, with a standard deviation of 128 ECU.

On the contrary, tax scheme C presents a flat rates structure and it generates the most unequal ex-post distribution, with a standard deviation of 527. That means that tax scheme C, despite reducing the initial inequality, generates an ex-post distribution of wealth which is four times more unequal than the distribution generated by tax scheme B[13].

11 The decision to provide windfall endowments, despite being an extremely controversial issue (Ackert et al. 2006, Antinyan et al. 2015, Cherry et al. 2002, Cherry et al. 2015, Clark 2002, Harrison 2007, Mittone et al. 2012 and Spraggon et al. 2009), is intentionally made to simulate a contingent distribution of assets, fortune and social circumstances. Indeed, the presence of some kind of "undeserved" inequality is a central issue in the Rawlsian social contract theory (Rawls 1999, pp. 10-15).

12 Technically the higher tax rates do not apply linearly to the whole income, but only to the proportion of income that exceeds the lower threshold.

Table 1[14] – Tax regimes (ECU)

initial endowment	1,500	2,000	3,000	tax revenue	rate	compounded tax revenue	ex-post redistribution (full compliance)			expected wealth	dev.st. (inequality)
tax scheme A											
rate	0.23	0.28	0.37								
taxes	345	560	1,110	2,015	2.1	4,232	2,566	2,851	3,301	2,906	371
tax scheme B											
rate	0.09	0.25	0.46								
taxes	135	500	1380	2,015	2.1	4,232	2,776	2,911	3,031	2,906	128
tax scheme C											
rate	0.31	0.31	0.31								
taxes	465	620	930	2,015	2.1	4,232	2,446	2,791	3,481	2,906	527
tax scheme D											
rate	0.19	0.22	0.43								
taxes	285	440	1,290	2,015	2.1	4,232	2,626	2,971	3,121	2,906	254

All the tax schemes are designed to generate exactly the same expected tax revenue (2,015 ECU in case of full compliance). The tax revenue is then multiplied by a capitalization factor of 2.1 and distributed in equal shares (1/3) to the three players of the considered group[15]. Given the structure described so far, also the final expected wealth is constant across the tax regimes (2,906 ECU).

This particular structure has two implications concerning the voting phase: first, any utilitarian reasoning centred on maximizing the expected average material utility (Harsanyi 1978) is formally prevented; second, the choice between tax schemes does not involve any explicit trade-off between efficiency and equality.

Nevertheless, the proposed tax schemes clearly have different distributive effects. Therefore the experimental subjects are supposed to focus and to base their voting decisions only on redistributive concerns.

13 In order to take into account inequality we could have also calculated a Gini index, which is 0.23 for the initial distribution, 0.08 for the regime A, 0.03 for tax scheme B, 0.11 for tax regime C and 0.06 for tax scheme D. However, this does not change the inequality ranking between tax regimes

14 The values in the table are calculated assuming the full compliance case. All the values, except the tax rates, are reported in ECU.

15 In particular, the equal share distribution has three distinct implications which can bee deepened in (Esarey 2012 and Fischbacher et al. 2014): the redistributive structure is conceived as a mechanism of transfers even if it is not a zero-sum game; every tax regime redistributes income from above average earnings to below-average incomes; players holding a different income level and facing different tax rates have different returns on the paid taxes.

The second phase of the experiment concerns the actual tax compliance task: for 10 rounds[16] the players are asked to pay taxes according to the endogenous tax scheme unanimously voted by their group during the first phase and to their own exogenously assigned level of income. After each round is played, together with their own ECU payoff, to the participants only the total amount of the tax revenue is shown. The specific contribution of the other players is neither directly communicated nor precisely evaluable. This is a plausible framework because from the point of view of a single tax payer, in a society it is only possible to observe the total amount of the available taxes or services, while we cannot directly observe the single decisions of other people, that is the contributions or the net incomes of people who surround us.

As mentioned in the Introduction, the compliance phase of our game is characterized by the absence of any external enforcement mechanism which can audit or sanction the players' deceptive behaviour. This means that the compliance phase basically reproduces the structure of a repeated public good game (Chaudhuri 2011 and Ledyard 1995) where the specific public good is represented by the tax revenue. In other words, since the agreement on the tax scheme of the first phase is not binding, in the second phase the players are asked to pay taxes on a voluntary basis in exchange of a monetary public good.

Given the voluntary mechanism on which the tax compliance game relies on, the payoff function for the single individual at every round is

$$(1)\quad \pi_i(t_i, t_{i\neq i}) = E_i - (t_i) + \frac{\beta}{n}\sum_{i}^{n}(t_i)$$

$$\text{with} \quad \frac{\beta}{n} = \frac{2.1}{3} 0.7$$

where E_i represents the assigned endowment (level of income), t_i measures the paid taxes for every individual, β is the capitalization rate and n is the number of players per group.

This payoff function implies that the social optimum is theoretically reached when all players fully contribute to the tax revenue (public good). However, since the ratio between the capitalization factor and the number of players per group is less than one and since there are no external enforcement mechanisms ("authorities") to stem tax evasion, the actual tax game of the second phase mirrors a standard public game, including its theoretical predictions. In other words, in our experimental design the standard Nash equilibrium applies requiring every player to not

16 In the instructions the number of rounds is not communicated.

comply at all (that is to contribute zero) to the chosen tax regime (to the formation of the tax revenue).

However, although the game is "authority free" and the standard equilibrium predicts a pure free-riding behaviour as the best response to others' behaviour, as mentioned in the previous Section, adopting a veil of ignorance in the voting phase is supposed to modify the psychological equilibrium of the game, generating sense of justice to the impartial agreement and therefore tax compliance.

Lastly in the game, contemporary to the compliance decision and through an incentivized structure, the players are asked to predict the level of compliance of the other two players belonging to their own group. In each group the player with the best cumulative predictions earns an extra bonus of €2. Predictions are then used as an indicator of beliefs about others' behaviours.

Except for the show-up fee and the bonus for the predictions, the experimental subjects are cumulatively paid for all the decisions they took across the 10 compliance rounds (Laury 2006). This choice was made with the intention to remark the dynamic process of compliance, which is not supposed to be framed as a series of one-shot decisions independent one from one another.

As mentioned above, the experiment is then run under two distinct treatments concerning the voting phase: the baseline treatment and the veil treatment. In the former treatment the veil is removed and the players vote after they are assigned their endowments. In the latter, which is inspired by the Rawlsian theory and its behaviourist interpretations, during the voting phase the experimental subjects are not informed about their personal level of income.

3.3 Data analysis and discussion

All the experimental sessions took place in the Computable and Experimental Economics Laboratory (CEEL) of the University of Trento. They were run using the open source software for economic experiments oTree (Chen et al. 2016). Each session lasted about 1 hour. The experiment involved a total of 153 participants (69 in the baseline treatment and 84 in the veil treatment), who voluntarily decided to participate after a public call. On average the participants were 22 years old, half of them were female and 48% of them were enrolled in programs related to economic disciplines. The participants were paid by means of bank transfers and on average they earned € 10.50 (show-up fee of €3.00 included).

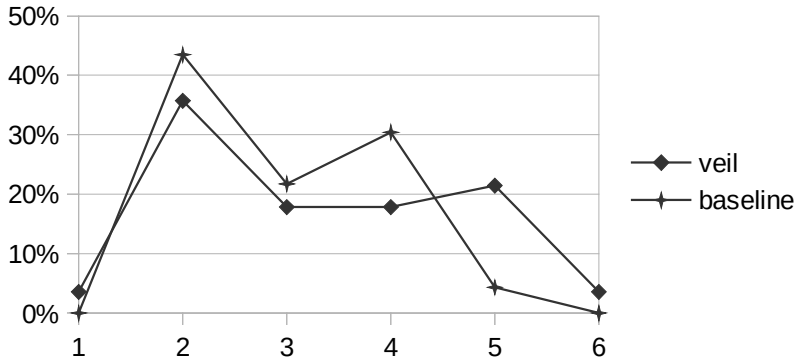
In the experimental laboratory the participants were randomly assigned to a computer terminal. All the stations were isolated by separation walls to avoid communication. The participants read the instructions on the computer screen. The instructions were also read aloud by one of the experimenters in order to ensure common knowledge. Before the actual experiment started six control questions about the structure of the game were asked. At the end of the experiment a non-incentivized questionnaire was provided and 94% of the participants declared that the initial instructions were clear.

Following the instructions, at the beginning of phase 1 of the experiment, regardless of the treatment, the participants were randomly assigned to a group of three people. Therefore a total of 51 groups (23 in the baseline treatment and 28 in the veil condition) took part in the experiment.

In both treatments all the groups accessed phase 2, that is all players reached a unanimous agreement on a specific tax regime. Chart 1 provides the details regarding the round number in which an agreement was reached.

In general almost half of the agreements were reached during the second voting round, showing a quite high propensity of coordination. However, it seems also that the veil of ignorance slowed down the coordination process towards unanimity[17]. Indeed, in the no-veil treatment basically all the groups reached the agreement by the fourth round, while the veil "constrained" 25% of the groups to wait until to the sixth round to find a unanimous consensus on a scheme of tax rates.

Chart 1 – Round of the agreement



During phase 1 in the baseline (veil) treatment a total number of 198 (276)[18] votes were provided. In Charts 2 and 3 it is possible to observe the percent distribution of votes concerning the

17 The opposite was somehow expected since the veil of ignorance is supposed to homogenize players' perspectives concerning the distributive priorities.

different tax regimes. In the two charts tax schemes are ordered from the one which maximizes the wealth of the player with the lowest income level (B) to the one with the most contained effects on the poorest player (C).

Chart 2 – Distribution of votes per tax regime

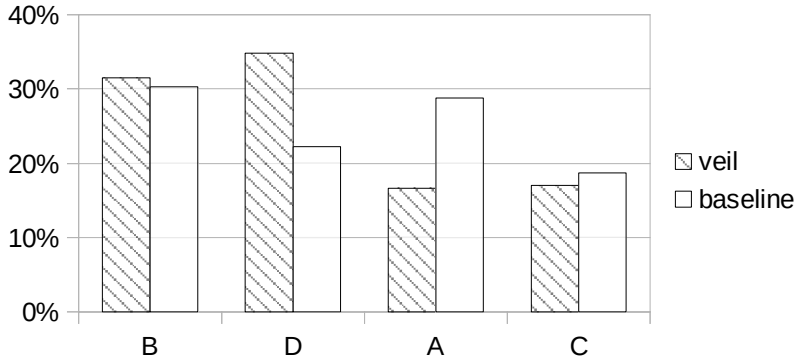
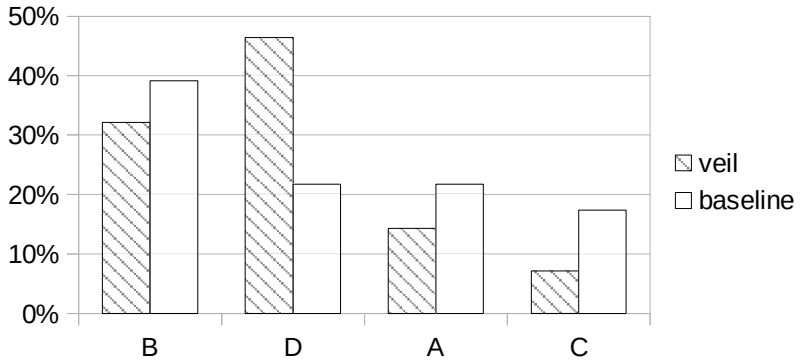


Chart 3 – Distribution of tax regimes



From the charts above we can claim that our hypothesis ***H1*** is, strictly speaking, disproved: the veil of ignorance did not produce any effect on the number of votes provided to the tax regime which maximizes the expectations of the least advantaged player, that is tax scheme B. Nevertheless, the charts show that a veil effect exists, even though not in favour of tax regime B. In particular, the veil of ignorance shifted the votes from tax schemes A and C to tax regime D, which is the second most advantageous for the player with the lowest level of income.

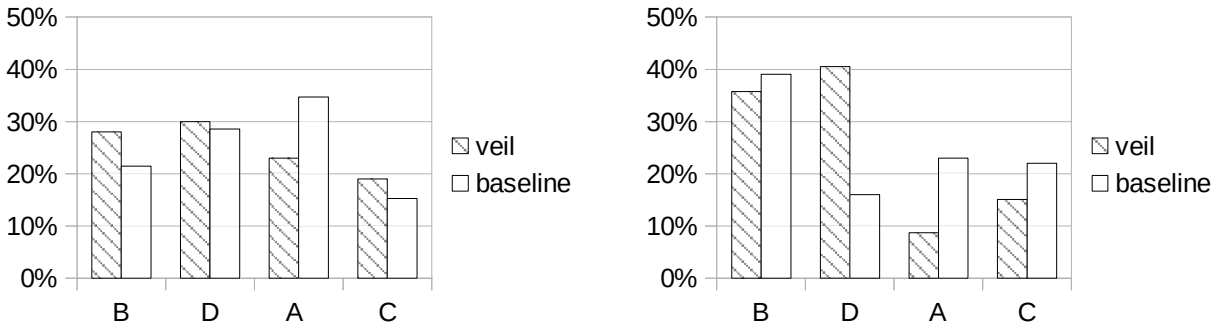
Indeed, in the veil condition tax scheme D was chosen almost half of the times, compared to 22% in the baseline treatment. This is an interesting empirical regularity. It shows us that there is a

18 This is further evidence of the "unanimity slowdown" in the veil treatment compared to the baseline design: in the former case players voted on average 3.3 times, in the latter 2.8.

hard kernel of students thinking that tax regime B is the fairest one regardless of the treatment. On the contrary the veil produces an effect on a share of participants who are not really convinced of the fairness of the two least progressive tax schemes.

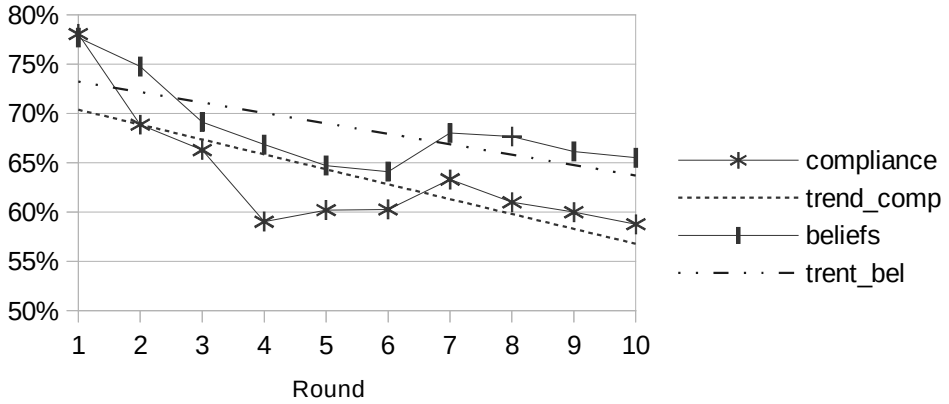
Furthermore, it is interesting to notice how the vote dynamic between the two treatments is mainly driven by male students (Charts 4 and 5).

Chart 4 and 5 – Distribution female (left) and male (right) votes per tax regime



As for the individual compliance, looking at the aggregate data (obtained by pooling together the two treatments), we cannot draw different conclusions from previous results achieved in repeated public good games without punishments (Fehr et al. 2002, Kroll et al. 2007, Ledyard 1995 and Chaudhuri 2011). Substantially the average individual compliance in the first round starts at about 80% and then it steadily declines to less than 60% in the last round[19] (Chart 6).

Chart 6 – Individual tax compliance and beliefs



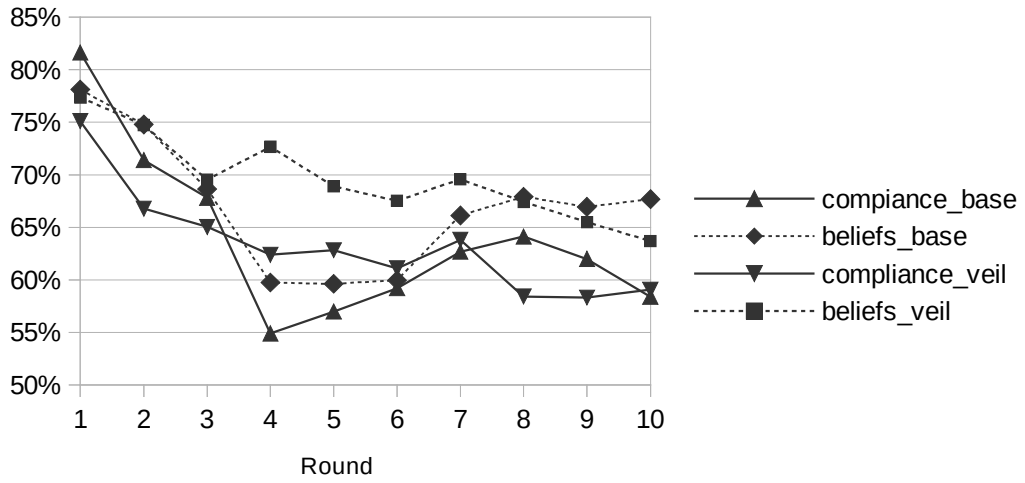
19 Compliance level might be higher than what it is usually find in the related literature, but in interpreting our results we have to take into account the participation effect generated by the voting procedure (see Section 1).

Focusing the attention on the aggregate predictions concerning others' compliance we can notice two interesting empirical facts. First of all, on average beliefs and compliance perfectly match (only) in the first round (78%, with corr. = 0.71). Despite this initial flawless match, from the second round onward a gap (average four, maximum seven percent points) emerges between the two variables, but they keep being highly correlated on average (corr. = 0.66).

Besides, even though the two measures slightly tend to diverge across rounds[20], the compliance rate follows the beliefs path. In general this is consistent with the so called reaction theories (Attanasi 2008, Croson et al. 2004), which claim that individual choices and actions are basically driven by beliefs on the others' behaviour. Thus the participants comply with a tax regime in the (discounted) measure they expect the other players in the group to comply with.

When data is then the separated according to the treatment (Chart 7) we find that compliance in the first rounds is higher in the baseline treatment than in the veil treatment. However, in general, across rounds we do not observe any significant difference in compliance levels. On average compliance in the baseline design is 64%, while it is 63% in the veil treatment. This result is consistent with our hypothesis ***H2***.

Chart 7 – Individual tax compliance and beliefs per treatment

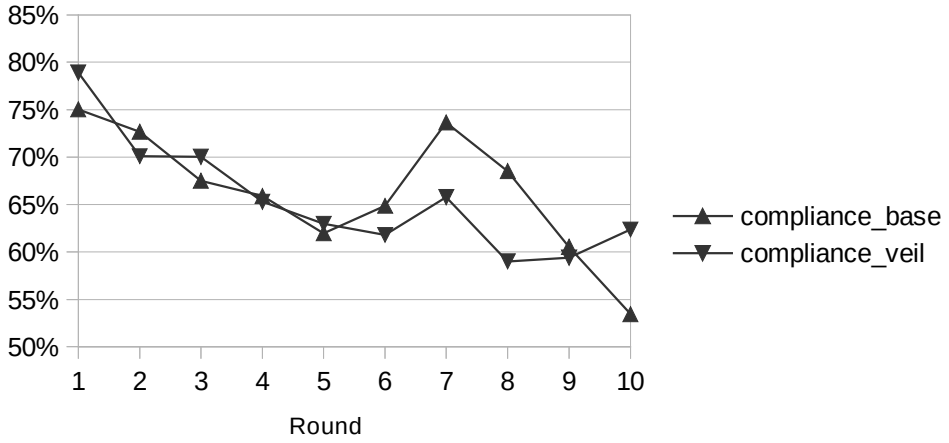


The just mentioned result is not as pleonastic as it might seem to be at first glance. The consideration that about 25% of experimental subjects behind the veil of ignorance "changed their mind" and accepted a fairer tax regime to be adopt in phase 2, joined with ***H2*** constitutes a result in

20 The increasing divergence between predictions and compliance might be due to the experimental structure. Subjects do not receive a feedback on others' individual contributions, but they are only shown the total tax revenue generated in each round. However, the fact that aggregated beliefs and average compliance follow a similar (decreasing) path (Chart 6) indicates that players can clearly adjust their behaviour in response to their beliefs.

favour of the veil of ignorance. Indeed, the fact that people can move to fairer tax schemes, where tax rates are lower for the poorest and higher for the richest, without it affecting the average level of tax compliance (Chart 8) is certainly a merit of the veil of ignorance procedure, which should not be undervalued.

Chart 8 – Compliance tax regime D



Getting back to the general dynamic of tax compliance, we observe (Chart 7) that in the veil treatment the compliance path is constantly decreasing, that is in our game compliance across rounds is neither self-enforcing nor stable as predicted by the theory. Thus we have to reject ***H3***. The rejection of the hypothesis is evident even focusing only on tax regime D (Chart 8), the one that impacted mostly on the deliberative voting in phase 1 and that produced the major reallocation of votes between the two treatments.

H4, which according to Grimalda et al. (2005) essentially predicts an alignment between beliefs and compliance in the veil treatment, is verified (Chart 10), while that is not true in the baseline treatment (Chart 9) because the two measures constantly diverge across rounds. The effect of the veil on beliefs is even more evident when we focus the attention on the behaviour of the participants who chose tax regime D. Looking at the trend lines of compliance and beliefs in the veil treatment (Chart 12) we can clearly see that the two measures, despite showing a constant gap, are almost perfectly aligned. On the contrary, in the baseline condition they do not seem to have a common direction (Chart 11).

This effect of the veil on the beliefs can be also remarked through a simple descriptive statistic: for the tax regime D the correlation between the average level of compliance and the individual predictions at each round is 0.24 in the baseline condition and it raises to 0.81 in the veil treatment. Therefore we can claim that in our repeated game the veil of ignorance, despite not

having a positive effect on compliance, plays a significant role on the empirical expectations (Bicchieri 2008. Bicchieri et al. 2009, Bicchieri et al. 2010) making them more aligned to the actual choice (however, this does not seem to be true in general, where the correlations are 0.92 and 0.89 respectively).

Chart 9 - Baseline compliance all tax regimes

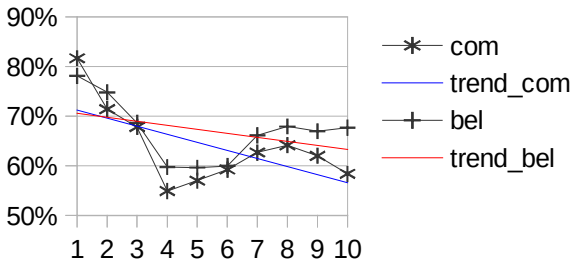


Chart 10 -Veil compliance all tax regimes

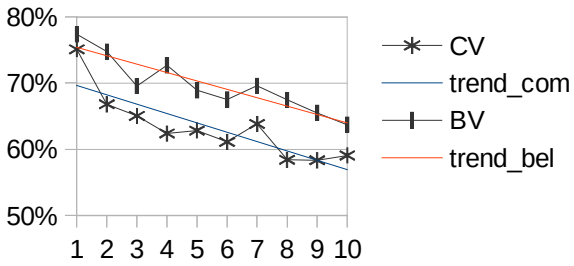


Chart 11 - Baseline compliance tax regime D

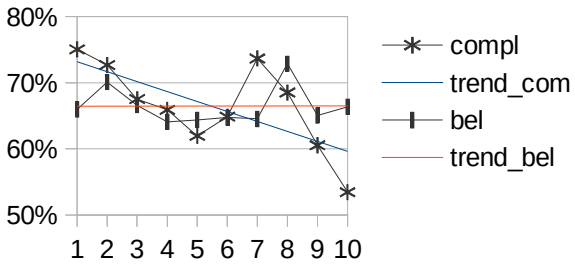
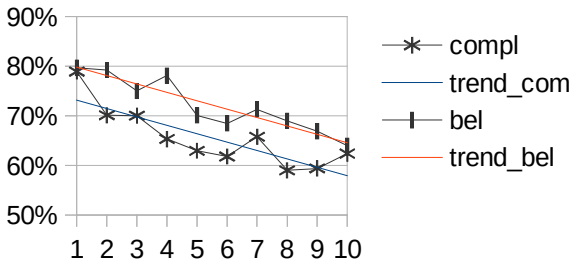


Chart 12 - Veil compliance tax regime D



However, even though behaviours are aligned to predictions, in our specific design choosing a non-binding (tax regime) distributive scheme behind a veil of ignorance did not produce a stable path of (tax) compliance as predicted. Instead, consistently with the standard literature, voluntary compliance keeps being fragile and the impossibility to communicate or to sanction free riders after the agreement represents a strong limit to the duration of cooperation and therefore of compliance itself (Fehr et al. 2002, Kroll 2007 and Ostrom et al. 1992).

The fact that the impartial perspective offered by the veil of ignorance cannot generate, by means of sufficiently stable beliefs, a constant level of compliance across rounds has two immediate implications.

First of all the conclusions related the one-shot exclusion game (Faillo et al. 2008, Faillo et al. 2014, Sacconi et al. 2005, Sacconi et al. 2010 and Sacconi et al. 2011) may need to be reviewing in order to take into account the limited effect, also across time, of a (laboratory) non-binding agreement behind a veil of ignorance. Indeed, although the reciprocal conformity model (Grimalda

et al. 2005) is verified, because round by round the compliance level is directly correlated with beliefs, the veil of ignorance cannot boost beliefs to keep compliance high across rounds[21].

Furthermore, the results concerning the tax compliance levels in our experiment shed light on a portion of the "slippery slope" curve (Kirchler 2008 and Kirchler et al. 2018), which might have been misrepresented. In particular the "slippery curve" should be reshaped and conceived as an asymmetric curve, because it has not been proved yet that full tax compliance can be based on the sole "trust in authorities (people)".

Given that in the two treatments we observed no differences in compliance levels, we are allowed to pool the data together in order to identify other interesting empirical regularities which were not taken into account by the predictive hypothesis. In Table 2 we report the coefficients of two linear models where the individual tax compliance, expressed in percentage terms, is the dependent variable. The second column of the Table checks for a treatment effect (model 1), while the coefficients of the third column are obtained pooling together the data of the two treatments (model 2). The errors of both estimations are clustered for group and for round.

$$individualcompliance_{i;t} = \alpha + \beta_1 treatment_{i;t} + \beta_2 round\,agreement_i + \beta_3 taxscheme_i + \beta_4 endowment_i$$
$$+ \beta_5 round_{i;t} + \beta_6 averagebelief + \beta_7 treatment\,x\,average\,belief_¿ + \beta_8 payoff\ ECU_{i;t-1} + controls_{i;t} + \varepsilon_{i;t}$$

For example, we can observe significant negative effects of the game round and of the payoff earned in the previous round. Each round compliance is reduced of about 1%, wile earning 1000 ECU more from one round to the next one decreases compliance of about 0.2%.

On the contrary, the assigned endowment and the predictions on the others' choices have a strong positive effect on tax compliance. In particular, believing that the other two players of the group will increase their compliance level of 1%, raises the individual compliance of 0.7 percentage points. If we then focus on the extremes, we can notice that according to the data, when the participants think that the other players will contribute with zero ECU to the common tax revenue (6% of the observations), they pay on average 5% of the due taxes. On the contrary, when an individual believes that the other participants of the group will fully comply with the chosen tax scheme (21% of the observations) they pay on average 89% of the due taxes. Therefore we can confirm again how behaviours are mainly driven by the expectations on the others' compliance (Croson et al. 2004 and Grimalda et al. 2005).

21 More specifically, the dynamic concerning the update of beliefs in the should be enquired (Battigalli et al. 2005)

Table 2[a] – Determinants of individual compliance

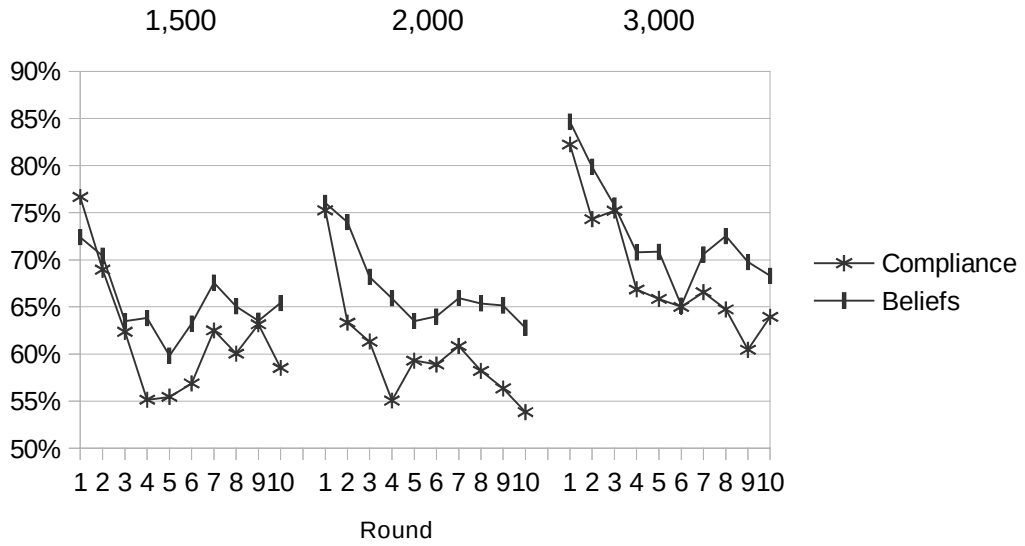
	Model 1 - treatment	Model 2 – pooled data
treatment	- 0.093 (0.056)	
round agreement	- 5.9e-04 (0.015)	- 0.003 (0.016)
tax scheme	0.009 (0.020)	0.005 (0.019)
endowment	0.118 (0.0367) **	0.110 (0.0037) **
round	- 0.009 (0.004) **	- 0.009 (0.004) **
average belief	0.725 (0.049) **	0.763 (0.036) **
*treatment * average belief*	0.077 (0.076)	
payoff ECU (round -1)	- 0.181 (0.054) **	- 0.178 (0.055) **
experimental experience	- 0.002 (0.003)	- 0.001 (0.003)
gender	0.076 (0.030) *	0.074 (0.030) *
age	0.005 (0.006)	0.005 (0.006)
field of studies	0.017 (0.034)	0.014 (0.033)
Participants 153		
Number of observation 1530		
* significant at 5%,		
** significant at 1%		
Adjusted R-squared	0.4674	0.4647
F-statistic p-value	2.2e-16	2.2e-16

a The variable *round agreement* refers to the round the agreement was reached (ranging from 1 to 6); *tax scheme* is the chosen tax regime, ordered from the flat tax scheme (1) to the most progressive (4); *endowment* is the gross amount in ECU expressed in thousands; *round* refers to the round of the tax compliance game (from 1 to 10); *average beliefs* takes in consideration the predicted average tax compliance of the other two players in the game and it is expressed in percentage; the interaction term tries to capture the additional effect of beliefs in predicting the compliance in the veil treatment; *payoff ECU* is the lagged term which captures the individual ECU payoff of the previous round; *experimental experience* is simply the number of previous experiments; *gender* is a binary dummy for the gender effect (male = 0); *age* is the personal age; *field of studies* refers to the undertaken studies (a dummy variable, where the value 1 corresponds to economic disciplines, whereas 0 to any other field, the second one is a discrete variable).

Besides, across the rounds compliance was higher for the players endowed with the highest level of income. Taking into account also the related predictions, it emerges the rich players pay more taxes (in percent terms) than the poorer ones because they think the latter will comply more.

The opposite occurs for the lowest levels of income. In other words, since the poor players think that the richer ones will contribute less, they comply less. Adding to the regression (model 2) an interaction variable between the income level and the beliefs in order to try to capture the mentioned extra effect does not provide any further significant result. However, this particular phenomenon can be observed in Chart 10.

Chart 10 – Compliance and income level (EMU)



Finally, and consistently also with previous studies (Kastlunger et al. 2010), we can observe a gender gap in tax compliance. The difference between male and female tax compliance is captured by the term *gender*, and being female increases tax compliance of about 8%.

Instead, there are not remarkable differences concerning the field of studies (economic field versus all the other disciplines). The accumulated "experimental experience" is not significant either. In the veil treatment beliefs become more predictive, but not in a significant measure.

Conclusions

Moving from the exclusion game design, we tested some further hypotheses related to Rawls's social contract theory. In particular we focused our attention on the pure veil effect and on the compliance dynamic of an impartial and non-binding agreement. In particular, we designed our experiment as a tax compliance game. The data collected from our game (with three players and

four possible tax regimes), showed that the veil of ignorance procedure has an important effect on the ex-ante votes allocation and a non-negative effect on compliance. However, the laboratory veil could not generate a path of compliance stable over time,

Going more in detail, the consideration that behind the veil of ignorance about 25% of groups moved from a less progressive towards a fairer tax regime (compared to the baseline treatment), together with the observation that this change had no significant impact on the average level of tax compliance should be considered a result in favour of the veil of ignorance. This is of course a week result, but it demonstrates that from a certain point of view the veil of ignorance is not a purely hypothetical tool: reasoning behind the veil of ignorance helps us to change our redistributive concerns without jeopardizing the economic outcome.

From a more general perspective our result shows that there is room for fighting inequalities without compromising the result in terms of efficiency (interpreted here in its broadest meaning). In other words our result shows that we can exploit the veil of ignorance to avoid, up to a certain degree, the trade-off between equality and efficiency in voluntary contribution mechanisms. Indeed, in our experiment, despite moving towards fairer tax regimes the average (considering all the rounds) contribution diminishes only of 3%, passing from a mean contribution of 446 ECU to 431 ECU.

Nevertheless, the laboratory veil did not produce any encouraging result as for the second main aim is considered, that is it could not generate a stable effect across time. Without the introduction of enforcement mechanisms, round by round compliance showed to be monotonically decreasing, as broadly documented for different voluntary contribution mechanisms. These specific result have relevant implications not only for the Rawlsian moral psychology and its experimental literature, but also for the "slippery slope" design. Indeed, if no arrangements are possible to generate full voluntary tax compliance, the slippery curve should be redesigned as asymmetrical.

Thus, a possible improvement of our tax game might consider the introduction of some sanctioning mechanisms in the compliance phase not so much to compensate the level of compliance for negative effects induced by the veil of ignorance, as for sustaining the level of compliance over time. This is a conclusion we can reach also reading more in depth John Rawls's words. In one passage, dealing with equal liberties, he admitted that tax compliance cannot survive over time reling on the sole sense of justice:

> "even in a well-ordered society the coercive powers of government are to some degree necessary for the stability of social cooperation. For although men know

> that they share a common sense of justice and that each wants to adhere to the existing arrangements, they may nevertheless lack full confidence in one another. They may suspect that some are not doing their part, and so they may be tempted not to do theirs. The general awareness of these temptations may eventually cause the scheme to break down. The suspicion that others are not honoring their duties and obligations is increased by the fact that, in the absence of the authoritative interpretation and enforcement of the rules, it is particularly easy to find excuses for breaking them. Thus even under reasonably ideal conditions, it is hard to imagine, for example, a successful income tax scheme on a voluntary basis. Such an arrangement is unstable. The role of an authorized public interpretation of rules supported by collective sanctions is precisely to overcome this instability. By enforcing a public system of penalties government removes the grounds for thinking that others are not complying with the rules. For this reason alone, a coercive sovereign is presumably always necessary, even though in a well-ordered society sanctions are not severe and may never need to be imposed. Rather, the existence of effective penal machinery serves as men's security to one another" (Rawls 199, p. 211).

In synthesis, "[t]he need for the enforcement of rules by the state will still exist even when everyone is moved by the same sense of justice" (Rawls 1999, p. 236).

Appendix: instructions of the veil treatment

Good morning,

We kindly ask you to read these instructions carefully. The instructions will be also read aloud by one of the experimenters. If at the end of the instructions you have any doubt, please raise your hand and wait for one of the experimenters to answer your questions.

EXPERIMENT

You are about to take part in an experiment which aims to investigate the tax attitudes of individuals who receive an income. During the experiment you will not be allowed to communicate in any way with the other participants. If you violate this rule you will be excluded from the experiment without being paid. The amount of money you can earn will depend on your decisions and on those of the other participants. The decisions you make will remain completely anonymous and no one will be able to associate your choices to your name. At the end of the experiment payments will be made by a bank transfer to your bank account.

The experiment will be run using Experimental Currency Units (ECU). 4.000 ECU are equivalent to €1.00. You will also earn €3.00 as show-up fee (SF) and €2.00 if you win the bonus (B).

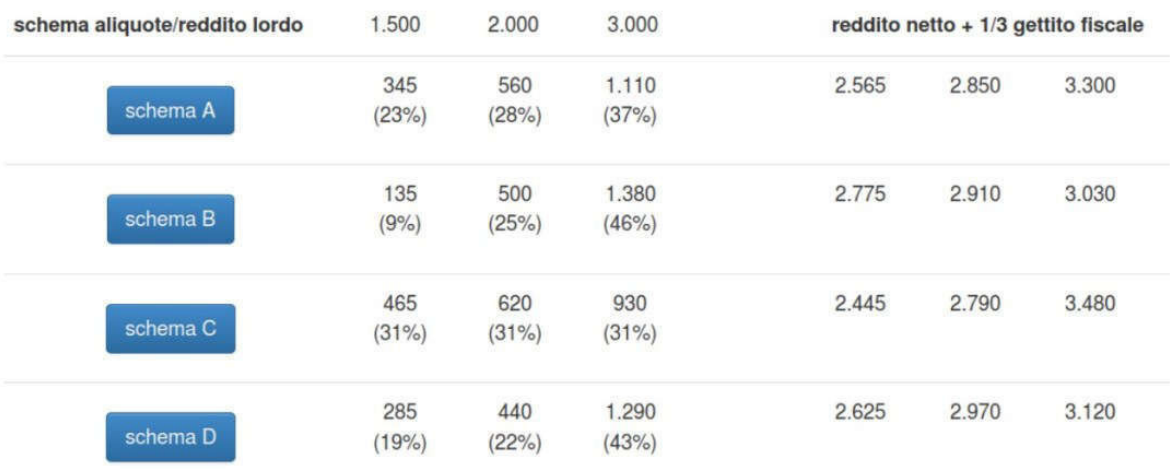
PHASE 1

At the beginning of phase 1 you will be randomly assigned to a group of people composed of other two participants. Each group will therefore be made up of three subjects. The group will be permanent, that is its participants will remain the same until the end of the experiment.

In phase 1, along with the other two players of your group, you will have to vote for the tax rates scheme you would prefer to adopt on three levels of gross income. The possible gross incomes are equal to 1.500 ECU, 2.000 ECU and 3.000 ECU and within the group there will not be two subjects with the same level of income. In phase 1 you will not know what level of income that will be assigned to you (this information will be communicated to you only in phase 2). Therefore you will have to choose the scheme of tax rates before you are assigned a level of income. The tax regimes

you can choose from are shown in the paper table next to the keyboard of your computer station (Figure 2) and which you can consult at any time during the experiment. All values (except percentages) are in ECU.

Figure 2 – Screenshot voting phase



schema aliquote/reddito lordo	1.500	2.000	3.000	reddito netto + 1/3 gettito fiscale		
schema A	345 (23%)	560 (28%)	1.110 (37%)	2.565	2.850	3.300
schema B	135 (9%)	500 (25%)	1.380 (46%)	2.775	2.910	3.030
schema C	465 (31%)	620 (31%)	930 (31%)	2.445	2.790	3.480
schema D	285 (19%)	440 (22%)	1.290 (43%)	2.625	2.970	3.120

The scheme of tax rates must be approved unanimously, i.e. all the subjects belonging to a group have to express the same choice about the type of tax to adopt in phase 2. In phase 1 you will have 6 rounds to reach unanimity:

- if unanimity is reached in any of the 6 available rounds you and your group will proceed immediately to phase 2 of the experiment, where only the voted scheme will be available;
- if at the end of the round number 6 you do not reached unanimous agreement on the tax scheme for you and your group the experiment will end here, and you will be paid only the show-up fee of €3.00.

PHASE 2

Phase 2 is made up of a predetermined number of rounds that will not be announced, therefore none of you will know it. However, all the groups that will access phase 2 will have the same number of rounds available. At the beginning of phase 2 you will be randomly assigned one of the expected income levels: 1.500 ECU, 2.000 ECU or 3.000 ECU. Within the group there will not be two subjects with the same income level and the income that will be assigned to you at the beginning will remain the same throughout all the rounds of phase 2.

At each round of phase 2 you will be asked to decide how much tax to pay according to your own income and the tax rates scheme voted during phase 1. The total amount of taxes paid by each participant of the group will constitute the tax revenue of the group. The tax revenue will be then multiplied by a capitalization factor of 2.1 and after that it will be redistributed in the same proportion (one third) to each participant of the group.

In the same screen where you will declare the amount of taxes you wish to pay, you will be also asked to predict the behaviour of the other two participants in your group. The player who will provide the best predictions on all rounds of phase 2 will get a bonus (B) of € 2.00 which will be added to his final payment. If two (or three) players provide equally correct predictions, the bonus will be awarded to both (or all three).

If you and your group access phase 2, the amount in € you will earn will be determined as follows:

Σ(assigned income in ECU – tax paid ECU + 1/3* (tax revenue ECU*2,1))/ 4.000 ECU* € 1.00 + € 3.00 SF + [€ 2.00 B]

Before proceeding with the experiment you will be asked to answer some brief control questions.

References

Ackert, L. F., Charupat, N., Church, B. K. & Deaves, R. (2006). An experimental examination of the house money effect in a multi-period setting. Experimental Economics, 9(1), 5-16.

Aguiar, F., Becker, A. & Miller, L. (2013). Whose impartiality? An experimental study of veiled stakeholders, involved spectators and detached observers. Economics & Philosophy, 29(2), 155-174.

Allingham, M. G. & Sandmo A. (1972). Income tax evasion: A theoretical analysis. Journal of public economics, 1(3-4), 323-338.

Alm, J., Jackson, B. R. & McKee, M. (1993). Fiscal exchange, collective decision institutions, and tax compliance. Journal of Economic Behavior & Organization, 22(3), 285-303.

Alm, J., McClelland, G. H. & Schulze, W. D. (1999). Changing the social norm of tax compliance by voting, Kyklos, 52(2), 141-171.

Alm, J. & Torgler, B. (2011). Do ethics matter? Tax compliance and morality. Journal of Business Ethics, 101(4), 635-651.

Andreoni, J., Erard, B. & Feinstein, J. (1998). Tax compliance. Journal of economic literature, 36(2), 818-860.

Antinyan, A., Corazzini, L. & Neururer, D. (2015). Public good provision, punishment, and the endowment origin: Experimental evidence. Journal of Behavioral and Experimental Economics, 56, 72-77.

Attanasi, G. & Nagel, R. (2006). Actions, Beliefs and Feelings: An Experimental Study on Dynamic Psychological Games.

Attanasi, G. & Nagel, R. (2008). A survey of psychological games: theoretical findings and experimental evidence. Games, Rationality and Behavior. Essays on Behavioral Game Theory and Experiments, 204-232.

Battigalli, P. & Dufwenberg, M. (2005): “Dynamic Psychological Games”, mimeo, August 2005 (previous version: IGIER working paper 287)

Becker W., Büchner H. J. & Sleeking, S. (1987). The impact of public transfer expenditures on tax evasion: an experimental approach. Journal of Public Economics, 34(2), 243-252.

Bicchieri, C. (2008). The fragility of fairness: an experimental investigation on the conditional status of pro-social norms 1. philosophical issues, 18(1), 229-248.

Bicchieri, C. & Xiao, E. (2009). Do the right thing: but only if others do so. Journal of Behavioral Decision Making, 22(2), 191-208.

Bicchieri, C. & Chavez, A. (2010). Behaving as expected: Public information and fairness norms. Journal of Behavioral Decision Making, 23(2), 161-178.

Bordignon, M. (1993). A Fairness approach to income tax evasion. Journal of Public Economics, 52(3), 345-362.

Bortolami, F. & Mittone, L. (2009). Does Participating in a collective decision affect the levels of contributions provided? An experimental investigation. Working paper, Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia, (No. 0902).

Bortolami, F. (2009). How to explain the participation effect: is it a question of different expectations and communication? A preliminary investigation. Working paper, Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia, (No. 0904).

Braithwaite, V. (Ed.). (2017). Taxing democracy: Understanding tax avoidance and evasion. Routledge.

Casal, S., Kogler C., Mittone, L. & Kirchler, E. (2016a). tax compliance depends on voice of taxpayers, Journal of Economic Psychology, 56, 141-150.

Casal, S. & Mittone, L. (2016b), Social Esteem Versus Social Stigma: The Role of Anonymity in an Income Reporting Game, Journal of Economic Behavior & Organization, 124, 55-66.

Chang, J. J., & Lai, C. C. (2004). Collaborative tax evasion and social norms: why deterrence does not work. Oxford Economic Papers, 56(2), 344-368.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Experimental Economics, 14(1), 47-83.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88-97.

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. American Economic Review, 92(4), 1218-1221.

Cherry, T. L., Kroll, S., & Shogren, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. Journal of Economic Behavior & Organization, 57(3), 357-365.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. Annu. Rev. Psychol., 55, 591-621.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance.

Clark, J. (2002). House money effects in public good experiments. Experimental Economics, 5(3), 223-231.

Croson, R. T. A. and Miller, M. (2004): Explaining the Relationship between Actions and Beliefs: Projection vs. Reaction, mimeo

Dal Bó, P., Foster, A., & Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. American Economic Review, 100(5), 2205-29.

Degli Antoni G., Faillo M., Francés-Gómez P. & Sacconi, L. (2016), Distributive Justice with Production and the Social Contract: An Experimental Study, Econometica Working Papers, N.60 September.

Esarey, J., Salmon, T. C., & Barrilleaux, C. (2012). What Motivates Political Preferences? Self-Interest, Ideology, and Fairness in a Laboratory Democracy. Economic Inquiry, 50(3), 604-624.

Faillo M., Ottone S. & Sacconi L. (2008), Compliance by Believing: An Experimental Exploration on Social Norms and Impartial Agreements, available at SSRN 1151245.

Faillo M., Ottone S. & Sacconi L. (2014), The Social Contract in the Laboratory: An Experimental Analysis of Self-Enforcing Impartial Agreements. Public Choice, 163(3-4), 225-246.

Fehr E., Fischbacher U., & Gächter S. (2002), Strong reciprocity, human cooperation, and the enforcement of social norms. Human nature, 13(1), 1-25.

Feld, L. P., & Frey, B. S. (2007). Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. Law & Policy, 29(1), 102-120.

Feld L. P. & Kirchgässner G. (2000), Direct Democracy, Political Culture, and the Outcome of Economic Policy: a Report on the Swiss Experience, European Journal of Political Economy, 16(2), 287-306.

Feld L. P. & Tyran J. R. (2002), Tax Evasion and Voting: An Experimental Analysis, Kyklos, 55(2), 197-221.

Filippin, A., Fiorio, C. V., & Viviano, E. (2013). The effect of tax enforcement on tax morale. European Journal of Political Economy, 32, 320-331.

Fischbacher, U., & Gachter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. American economic review, 100(1), 541-56.

Fischbacher, U., Schudy, S., & Teyssier, S. (2014). Heterogeneous reactions to heterogeneity in returns from public goods. Social Choice and Welfare, 43(1), 195-217.

Frey, B. S., & Torgler, B. (2007). Tax morale and conditional cooperation. Journal of Comparative Economics, 35(1), 136-159.

Frohlich, N., Oppenheimer, J. A., & Eavey, C. L. (1987). Laboratory results on Rawls's distributive justice. British Journal of Political Science, 17(1), 1-21.

Gangl K., Hofmann E., & Kirchler E. (2015), Tax authorities' interaction with taxpayers: A conception of compliance in social dilemmas by power and trust. New ideas in psychology, 37, 13-23.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. Games and economic Behavior, 1(1), 60-79.

Gintis H. (Ed.). (2005), Moral sentiments and material interests: The foundations of cooperation in economic life (Vol. 6). MIT press.

Grimalda, G., & Sacconi, L. (2005). The constitution of the not-for-profit organisation: reciprocal conformity to morality. Constitutional Political Economy, 16(3), 249-276.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. Science, 312(5770), 108-111.

Harrison, G. W. (2007). House money effects in public good experiments: Comment. Experimental Economics, 10(4), 429-437.

Harsanyi, J. C. (1978). Bayesian decision theory and utilitarian ethics. The American Economic Review, 68(2), 223-228.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists?. Behavioral and Brain Sciences, 24(03), 383-403.

Jackson B. R., & Milliron V. C. (1986), Tax compliance research: Findings, problems, and prospects. Journal of accounting literature, 5(1), 125-165.

Kahan D. M. (1997), Social Influence, Social Meaning, and Deterrence, Virginia Law Review, 349-395.

Kastlunger, B., Dressler, S. G., Kirchler, E., Mittone, L., & Voracek, M. (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2D: 4D). Journal of economic psychology, 31(4), 542-552.

Kastlunger, B., Kirchler E., Mittone L. & Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. Journal of Economic Psychology, 30(3), 405-418.

Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. The Scandinavian Journal of Economics, 102(1), 23-39.

Kirchler E. (2007), The economic psychology of tax behaviour. Cambridge University Press.

Kirchler E., Hoelzl E., & Wahl I. (2008), Enforced versus voluntary tax compliance: The "slippery slope" framework. Journal of Economic Psychology, 29(2), 210-225.

Kroll S., Cherry T. L. & Shogren, J. F. (2007), Voting, Punishment, and Public Goods, Economic Inquiry, 45(3), 557-570.

Laury, S. K. 2006. "Pay One or Pay All: Random Selection of One Choice for Payment." Georgia State University, Economics Center Working Paper Series 2006–24.

Ledyard, O. (1995). Public goods: some experimental results. In J. Kagel & A. Roth (Eds.), Handbook of experimental economics. Princeton: Princeton University Press (Chap. 2).

Masclet, D., Noussair, C., Tucker, S., & Villeval, M. C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. American Economic Review, 93(1), 366-380.

Mittone L. (2006), Dynamic behaviour in tax evasion: An experimental approach, The Journal of Socio-Economics, 35(5), 813-835.

Mittone, L., & Ploner, M. (2012). Asset legitimacy and distributive justice in the dictator game: An experimental analysis. Journal of Behavioral Decision Making, 25(2), 135-142.

Muehlbacher, S., Kirchler, E., & Schwarzenberger, H. (2011). Voluntary versus enforced tax compliance: Empirical evidence for the "slippery slope" framework. European Journal of Law and Economics, 32(1), 89-97.

Nikiforakis, N., & Normann, H. T. (2008). A comparative statics analysis of punishment in public-good experiments. Experimental Economics, 11(4), 358-369.

Nikiforakis, N., Noussair, C. N., & Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. Journal of Public Economics, 96(9-10), 797-807.

Orviska M. & Hudson J. (2003), Tax Evasion, Civic Duty and the Law Abiding Citizen, European Journal of Political Economy, 19(1), 83-102.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. American political science Review, 86(2), 404-417.

Pickhardt M., & Prinz A. (2014), Behavioral dynamics of tax evasion–A survey. Journal of Economic Psychology, 40, 1-19.

Pommerehne W. W. & Weck-Hannemann H. (1996), Tax Rates, Tax Administration and Income Tax Evasion in Switzerland, Public Choice, 88(1-2), 161-170.

Rawls J. (1999), A Theory of Justice, Revised Edition, Harvard University Press.

Richardson G. (2006). Determinants of tax evasion: A cross-country investigation. Journal of International Accounting, Auditing and Taxation, 15(2), 150-169.

Sacconi L. & Faillo, M. (2005), Conformity and Reciprocity in the 'Exclusion Game': An Experimental Investigation, University of Trento Economics working paper, (12).

Sacconi L. & Faillo, M. (2010), Conformity, Reciprocity and the Sense of Justice: How Social Contract-Based Preferences and Beliefs Explain Norm Compliance: the Experimental Evidence, Constitutional Political Economy, 21(2), 171-201.

Sacconi L., Faillo M. & Ottone S. (2011), Contractarian Compliance and the Sense of Justice': A Behavioral Conformity Model and Its Experimental Support, Analyse & Kritik, 33(1), 273-310.

Sefton, M., Shupp, R., & Walker, J. (2007). The effect of rewards and sanctions in provision of public goods. Economic Inquiry, 45(4), 671–690.

Spraggon, J., & Oxoby, R. J. (2009). An experimental investigation of endowment source heterogeneity in two-person public good games. Economics letters, 104(2), 102-105.

Tammi T. (2011), Contractual Preferences and Moral Biases: Social Identity and Procedural Fairness in the Exclusion Game Experiment, Constitutional Political Economy, 22(4), 373-397.

Torgler, B. (2002). Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. Journal of Economic Surveys, 16(5), 657-683.

Tyran, J. R. and Feld, L. P. (2001) Why people obey the law. Experimental evidence from the provision of public goods. Working Paper, University of St. Gallen.

Tyran, J. R., & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. The Scandinavian Journal of Economics, 108(1), 135-156

Wahl I., Muehlbacher S. & Kirchler, E. (2010), The Impact of Voting on Tax Payments, Kyklos, 63(1), 144-158.

Wenzel M. (2004), An analysis of norm processes in tax compliance. Journal of economic psychology, 25(2), 213-228.

Wenzel M. (2005), Motivation or rationalisation? Causal relations between ethics, norms and tax compliance. Journal of Economic Psychology, 26(4), 491-508.

4. Economics of Climate Change and Social Contract Theory: Intergenerational Insights From a Laboratory Experiment in a Rawlsian Perspective

Klaudijo Klaser, Lorenzo Sacconi and Marco Faillo

Abstract: Many actions we take today will show some of their consequences in the future. Therefore future generations, although they cannot have a real voice, should be considered as direct stakeholders of some of our present decisions. As far as this intertemporal misalignment between actions and outcomes is concerned, climate change is the most evident example we have of negative externality towards the future. This paper looks at the climate change problem and the related international agreements on the reduction of greenhouse gas emission through the social contract perspective.. We apply John Rawls's veil of ignorance decision-making model within an experimental setting. In particular, we implement a sequential group dictator game where generations (groups of players) are located on a chain representing the time line. The (laboratory) veil of ignorance induces a fair ex-ante perspective regarding the distribution of resources between generations, however ex-post compliance to the agreement remains an open issue.

JEL Code: D63, D64, F64, Q54

Keywords: Experimental Economics, Climate Change, Intergenerational Allocation of Resources, Veil of Ignorance, Social Contract Theory

"Society is indeed a contract [...] it becomes a partnership not only between those who are living, but between those who are dead, and those who are to be born"

Edmund Burke

"The dominant reason for acting on climate change is not that it would make us better off. It is that not acting involves taking advantage of the poor, the future, and nature"

Stephen M. Gardiner

Introduction

Climate change is a threat which looms over future generations but that is triggered by some careless actions of the present one, who enjoys the benefits of those actions. In a technical language "[c]limate change is an instance of an externality—when one agent's activities have costs or benefits for other agents that are not reflected in the prices the first agent faces" (Clements 2015, p. 263). This means that only the present generation can act in advance, constraining its own behaviour, in order to limit negative externalities, or said otherwise, the risk of bad consequences of global warming on future generations.

More importantly, climate change is a global issue and it cannot be tackled by the commitment of a minority of virtuous agents, that is environmental issues like the global warming cannot be solved through single community actions; instead they require a certain degree of international cooperation (Stern et al. 2006, p. 512). Therefore, if soon nations do not reach a widely shared agreement on how to coordinate themselves in order to limit or to avoid, through today's actions, the global warming and its consequences there is the risk to harm seriously future generations.

However, "notwithstanding more than twenty years of international negotiations to establish limits, emissions of greenhouse gases continue to rise" (Gardiner et al. 2016, p. 137)[1]. In other words, even if mankind understood the dangers intrinsic in climate change long time ago (Nordhaus

1 The first formal attempt to address climate change towards a common solution was the United Nations Framework Convention on Climate Change in 1992. However, according to the world meteorological organization, in 2019 greenhouse gas levels reached a new top (https://public.wmo.int/en/media/press-release/greenhouse-gas-levels-atmosphere-reach-new-record).

1993), it is not still capable to pursue the common goal of containing the risk to harm future generations. Two are the main reasons of this constant failure of international negotiations on climate actions.

On the one hand the success of agreements concerning the reduction of global greenhouse gases (the main cause of the global warming) is essentially conditioned upon the distribution of costs between nations: by and large some of them are expected to make grater sacrifices in containing emissions, or said more explicitly, some nations are supposed to pay higher costs in pursuing the common goal (Gardiner 2011a, Gardiner et al 2016). However, nobody is really willing or available to pay more than the others. This makes difficult any agreement, because it is not really clear which nations exactly should bear higher costs, that is there is not a shared intragenerational principle regulating the distribution of costs. This prevents any intragenerational commitment.

On the other hand, even if some actual agreements were formally reached (e.g. the Kyoto protocol), compliance to such agreements is known to be extremely fragile. Since reducing emissions is costly and since in the current geopolitical frame there are not institutions which can monitor and sanction defectors, single nations have a clear economic incentive to free ride[2]. Therefore, although it might be collectively rational to cooperate, from the individual (national) point of view it is rational to deviate from the formal agreements.

However, without an international agreement to reduce greenhouse emissions the world will head a tragedy of commons (Hardin 1968) because the atmosphere, despite being very big, can contain a limited quantity of greenhouse gases before these show their harmful effects on mankind. In addition to this concern we have to take into consideration that the common-pool resource dynamic (Ostrom et al. 1994) which usually drives the appropriation of natural resources over time (and the resulting parallel creation of noxious waste) is amplified by a strong present bias, since action and consequences occur at different times and on different people. Again, given the asymmetric relationship between generations, overexploitation of resources most of the times benefits the present generation at the expenses of the future ones.

For example, producing electricity through nuclear power stations we currently benefit of energy generated at a lower cost. However within some areas the radioactive slags will constrain and will jeopardize future generations for hundreds of years. The consequences of an unlimited use of fossil fuels might be even more tragic because they involve the whole globe. The accumulation

2 Baseically, without supranational institutions it is not possible to change the payoff structure, that is we cannot introduce incentives to induce or sanctions to enforce cooperation between nations.

of carbon dioxide in the atmosphere generates the so called greenhouse effect and this, through the increase of global temperatures, will negatively condition the existence of future generations all over the world.

Therefore how can the current generation, in its own decisions, take into account in a fair way the interests of future generations which have no voice (because not existing yet) at the table discussion but which are clearly direct stakeholders of the present actions? The answer is not easy at all if we take into consideration many specific features which distinguish allocation of resources between generations (Meyer 2016) from the circumstances which characterize the more familiar redistributive issue between contemporaries (Lamont and al. 2016 and Tremmel 2009, p. 147)[3].

The standard economic approach deals with the intertemporal allocation of resources assuming that the utility function to maximize depends positively not only on the bundle of consumption of the present (person) generation, but also on the consumption or on the utility of future (people) generations (Solow 1974). Within this kind of functional forms a discount rate is introduced to represent the degree of concern that one generation bears for the next ones, so that the maximization of the present utility keeps balanced with the (conjectured) interests of future generations. In this way a positive discount rate is supposed to avoid overconsumption behaviours of the present generation which might unfairly damage future generations.

However the assumption of an intergenerational perspective embodied in the utility function through a discount rate seems to be quite limiting for different reasons. For example that approach contradicts the classic pillar of the purely selfish homo oeconomicus who is supposed to care exclusively about his own consumption level and not at all about the welfare of his offspring[4].

Besides, within that economic approach very much depends on the social rate of discount and on the utility functional form. However there is not unanimity on how much exactly future generations are supposed to count in our present decisions. In the environmental economics field

3 "[I]ntergenerational [distributive] justice is saddled with puzzling difficulties, such as the nonidentity problem [...], the cooperation between generations [...], motivational considerations, conceptualisation of duties and toward future generations, lack of information, uncertainty and asymmetries of power" (Gabor 2013, p. 301). To not take into consideration the rationale of deriving intertemporal norms from a purely intragenerational context (Heyd 2009, p. 177). In other words a group of existing people (the contemporaries) is supposed to derive distributive principles and practices which take into consideration interests of other non-existing groups (future generations) that formally cannot claim anything. Said otherwise, we have "to build and rationalise a problem of cooperation, duties, rights, compliance, between non-existent individuals who lived, live or will live in different moments of time" (Gabor 2013, p. 304).

4 Moreover, it is not clear why the other regarding concern is usually assumed to be a one way vertical component. Thus the following asymmetry is inexplicable: why would it be legitimate to assume that an economic agent cares only about her children, but not about the welfare of her siblings or her friends, or even of her ancestors (like parents or grandparents)?

there is still a heated debate regarding the appropriate weight (social discount rate) to assign to future generations' welfare (Nordhaus 2008, Moore et al 2004 and Stern 2008). In the same way, there is complete uncertainty about future generations' utility functions. Said otherwise, future people do not exist yet for definition and they cannot reveal their preferences (Beckerman 2006 and Parfit 1984), therefore we cannot really know what is better for them (Barry 1977).

It is within this (considered inadequate) theoretical framework that the social contract theory can provide a useful and innovative tool to deal with climate change and more generally with distribution of resources between generations.

John Ralws (1999) was the pioneer in extending in a structured way the social contract model to the allocative problem between generations. In his theory Rawls claims that the set of currently existing people (therefore not all the generations of the history), instrumentally rational and free of any other regarding preference, have to design the principles to regulate the intergenerational distribution of resources (Rawls 1999, pp. 118-123). Preventing then, by means of a veil of ignorance, the current generation from knowing the specific moment of the history it belongs induces the (present) parties to design a fair principle for the allocation of resources through the human history.

The most interesting feature of the Rawlsian intragenerational setting for intergenerational principles is exactly the designed decision-making model. In particular in Rawls's intergenerational theory individuals taking part in the agreement are only contemporaries and they have to choose an allocative configuration which does not affect exclusively themselves. Instead the present generation is called to evaluate distributive principles that will produce effects also on third parties (future generations) who formally cannot take part in the contract, cannot make demands, cannot make objections, cannot threat and cannot punish the actual decision maker: this is exactly the modern climate change issue[5], which is characterized by a dictatorship of the present.

A further relevant feature of a Rawlsian approach to the specific climate change issue is then the ex-post compliance to the social contract. Indeed the agreement behind the veil of ignorance is not conceived by Rawls as binding. In other words, although the distributive principles are the outcome of a formally fair procedure, once the veil is dropped a dictatorship of the present generation over the future ones remains a concrete possibility. This is a strong analogy with what

5 The social contract on intergenerational principles is very different from the standard distributive issue where a set of (contemporary) individuals has to agree on the way to divide resources among themselves: whether I have to agree with you on how to split $10 between us or whether I have to agree with you on how to split $10 between ourselves and a third person who has absolutely no voice on the issue are two extremely different decision-making frameworks. In the latter case individuals involved in the contract are supposed to take into consideration some people who remain outside of the contract moment.

was previously shown: compliance to the international agreements on greenhouse gases reduction is a big real problem, because they are not enforceable.

In the experimental field compliance to non-binding agreements on intragenerational distributive principles was explained with the Rawlisan idea of the sense of justice (Degli Antoni et. al 2016, Faillo et al. 2008, Faillo et al. 2014, Rawls 1963, Rawls 1999, Sacconi and Faillo 2005, Sacconi and Faillo 2010, Sacconi et al. 2011 and Tammi 2011). Thus we can try to amplify the same concept of the sense of justice to the intergenerational context. In particular, we can extend its validity, verifying if a different decision-making frame (with the agreement that does not formally and substantially include the main stakeholders, that is the least advantaged party), concerning a slightly different distributive problem (resources have to be distributed not within group but between groups), leads to the same positive conclusions on the sense of justice.

Within the general framework described so far, we apply Rawls's intergenerational social contract model within an experimental setting in order to provide some useful insights concerning the modern climate change issue and the related agreements on the reduction of noxious emissions. The next Sections are therefore organized as follows.

Section 1 introduces Rawls's social contract theory (Rawls 1999), with a specific focus on its intergenerational extension. The aim of this part is to provide the theoretical background for the analysis of distribution of resources between generation in a Rawlsian perspective. Although the slant given by John Rawls to his intergenerational social contract theory (Rawls 1999, pp. 251-267 and Rawls 2001, pp. 158-160) requires some prudential clarifications (see Appendix A), his idea of an intragenerational agreement behind a veil of ignorance (Rawls 1999, pp. 118-123) is considered an adequate model for inquiring the modern climate change issue.

Section 2 describes the experimental design which captures the main features of the modern climate change problem in a Rawlsian perspective. The game is structured as a group dictator game (Kahneman et al. 1986) played sequentially (Bahr et al. 2007 and Casol et al. 1998) and it is run in two distinct conditions, with and without a preliminary voting stage simulating an agreement behind a veil of ignorance.

Section 3 provides the predictive hypothesis. Section 4 analyses the data of the experiment.

Appendix A analyses in depth the Rawlsian social contract theory extended to the intergenerational issue. Appendix B contains the complete instructions provided to the participants and read aloud during each experimental sessions.

4.1 John Rawls's social contract theory on allocation of resources between generations

Rawls designs his ethical system (Rawls 1999) in order to identify the main principles which should lead the human society and its institutions, with particular concern to the division of benefits generated by cooperation between individuals (Rawls 1999, p. 4). Rawls's theory establishes then a procedure inspired by the social contract tradition, that is the principles are the outcome of an agreement between those individuals involved in the cooperative scheme.

Within the Rawlsian decision-making procedure the impartiality in the choice of the principles is guaranteed by a veil of ignorance (Rawls 1999, pp. 118-123). This is specific tool is thought to excludes the access to any particular information to those parties who take part in the agreement. Therefore, in the agreement phase,

> "no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism" (Rawls 1999, p 118).

Furthermore it is important to highlight how according to Rawls "persons in the original position have no information as to which generation they belong" (Rawls 1999, p 118).

"In this manner the veil of ignorance is arrived at in a natural way", since it is excluded "the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices" (Rawls 1999, p. 17). In this way, behind the veil of ignorance no one can take advantage of personal contingencies to design principles (norms or institutions) which might favour her own particular position: in the original position everybody is equally represented since everybody has to choose in the same situation of perfect (mis)informational symmetry and the involved parties reach an agreement only on the basis of impartial and general considerations (Rawls 1999, pp. 118-123).

Although the distribution of resources between generations is particularly challenging and it "subjects any ethical theory to severe if not impossible tests" (Rawls 1999, p. 251) John Rawls (1999 and 2001) does not fail to extend his contractarian system in order to contemplate this topic. Indeed, Rawls is aware how the account of his social contract theory "would be incomplete without some discussion of this important matter" (Rawls, 1999, p. 251).

When Rawls moves from the intragenerational context to the analysis of principles that are supposed to regulate the allocation of resources between generations he adds a reasonable specification concerning the decision-making procedure described so far. In particular, Rawls specifies that although people are deprived of the information concerning the generation they belong (Rawls 1999, p. 118 and p. 254), that is even if they ignore the historical period and the economic development of the society they represent, the parties behind the veil of ignorance are all contemporaries (Rawls 1999, p. 121). Said otherwise the people involved in the agreement, even though focused on intergenerational principles, belong to the same generation (and they know it as a general fact).

With this specification Rawls substantially constrains his intergenerational decision-making model to the physically existing people (Dierksmeier 2006, p. 74). Indeed, in his opinion it would be unrealistic to conceive an agreement (although hypothetical) which gathers together all the possible generations of the human history: this all-inclusive approach would stretch imagination too much, that is it would require a too high level of abstraction[6] (Attas 2009 pp. 195-7, Rawls 1999 p. 120, Rawls 2001, p. 160 and Tremmel 2009, p. 156).

Notwithstanding the clarification concerning the contemporaneity of the parties (see Appendix A for further details), Rawls's social contract theory seems to be promising for dealing with the allocation of resources between generations (Tremmel 2013, p. 484) because with the veil of ignorance the present generation substantially looses its privileged position (dictatorial powers) towards the future ones. Indeed the veil of ignorance guarantees that the parties involved in the agreement, despite being and knowing to be contemporary, are encouraged to propose principles for the division of resources between generations which are impartial[7].

However, in Rawls's opinion the standard difference principle, which requires to maximize the expectations of the worst-off (Rawls 1999, p. 56, p. 69, and p. 72), is not a suitable tool to deal with redistribution of resources between generations because it apparently produces some ethically undesirable consequences (Gardiner 2009, Rawls 1999, pp. 253-255 and Appendix A). Therefore the issue concerning the intergenerational allocation of resources "must be treated in some other manner" (Rawls 1999, p. 254). Thus Rawls proposes the just saving principle as the normative rule

6 Indeed Rawls's model is to be "understood as a purely hypothetical situation" (Rawls 1999, p. 11). In other words, Rawls's original position coincides with the adoption of a particular perspective, and therefore the agreement is conceived as a simple mental experiment.

7 According to Rawls, even though it is formally the present cohort to decide about the allocation of resources between generations, the veil of ignorance procedure induces the parties to take into appropriate consideration also future generations and to choose "a path over time which treats all generations justly during the whole course of a society's history" (Rawls 1999, p. 257)

to regulate the distribution of resources over time: “in following a just savings principle, each generation makes a contribution to those coming later and receives from its predecessors” (Rawls 1999, p. 254).

Nevertheless Rawls does not describe in detail the peculiar features of the just saving principle (like for example providing a specific saving rate or a schedule of rates). Instead he limits himself in sketching some general ethical restriction which the contractual parties should take into account in defining the saving path (Rawls 1999, pp. 255-6). However, those are not of interest for the most immediate aims, and more details about the derivation and the configuration of the just saving principle are provided in Appendix A.

For the purposes of the experiment and its Rawlsian interpretation it is sufficient to highlight the working mechanism of the just saving principle. Basically, according to Rawls, every (present) generation is expected to give up a share of its own resources in order to pass it to the following generation. This is the outcome even though the next generations do not take part in the present agreement.

Therefore, with Rawls’s social contract model (and without discussing about social rates of discount or about future people’s preferences), it seems to be possible to justify in a compelling way the idea that a closed set of self-regarding contemporary individuals can take into due consideration people belonging to future generations. Again, this is the modern climate change issue.

4.2 Experimental design of the baseline treatment

Many different theoretical approaches have been proposed to address the environmental issue in a Rawlsian perspective, going from considering health and environment as social primary goods and including animals in original position (Gardiner 2011b and Thero 1995), to running a third level original position (Clements 2015). In the same way many experimental works aimed at testing empirically rither the assumptions or the conclusions of Rawls’s social contract theory (Gaertner and Schokkaert 2012). Nevertheless, there was not a systematic research which explicitly tried to merge the two fields. To the best of our knowledge, there was only one single attempt aimed at testing Rawls’s intergenerational theory within an experimental setting (Wolf and Dron 2015).

Wolf and Dron’s design is very intuitive. A common endowment is provided to a group of five participants. The single players are then randomly assigned to a position within a sequential dictator game (Bahr and Requate 2007 and Cason and Mui 1998). Starting from the player occupying the

first position, the participants sequentially enter the dictator role where they are asked to claim a share of the common endowment for themselves[8] until either the fourth player takes a decision (therefore the fifth pparticipant becomes a dummy player) or the common endowment is exhausted.

The underling idea of the game is that each single player in the sequence formally represents a (non-overlapping) generation, because each person: a) (except the first one) is subject to the consequences of the decisions taken by all previous generations; b) with her or his own decision can influence only the welfare of the following generation(s). Besides, in order to represent realistically the "dictatorship of the present" issue, every participant who enters the dictator role is allowed to claim the 100% of the (remaining) endowment, with the consequence that nothing would be left to the following players.

In one of the proposed treatments Wolf and Dron introduce a preliminary stage where players are asked to agree on a rule to share the common endowment between the five generations in the sequence. However, they are asked to do that behind a Rawlsian veil of ignorance, that is before knowing the position they will occupy in the actual sequential dictator game. Indeed, after players agree on a sharing rule they are assigned to a position (generation) and they sequentially enter the dictator role exactly as in the baseline treatment described above.

However, Wolf and Dron's (2015) attempt is to be considered unsatisfactory and not very representative of Rawls's theory because of two main reasons.

First of all the veiled agreement, even though it included the least advantaged subject (the last generation), did not produce a (significant) more equal distribution between the five players compared to the baseline treatment where there was not any kind of agreement before the players entered the sequential dictator game. Thus, although the participants (representing each one a generation) had the chance to discuss together about the ex-post distribution of the common endowment, players located in privileged positions (first generations) profited anyway of their contingency (Wolf and Wagner 2016) as the standard economic theory would predict.

Second, and even more importantly, their experimental design is to be considered inconsistent in relation to Rawls's theory strictly interpreted. Indeed in Wolf and Dron's experiment all the generations are put behind the same veil of ignorance, as if different generations could reach an intergenerational agreement (Anderson 2013). Instead John Rawls is really careful and clear to specify how the agreement for intergenerational principles is intragenerational, that is only contemporaries, generation by generation, are involved in the deliberative process behind the veil

8 In the experiment, the share that the players decide to withdraw from the common endowment during their turn constitutes their final payment.

(Rawls 1999 pp. 118 -121). Therefore, since Wolf and Dron's design represents generations with single individuals it also excludes the possibility of an agreement within generations.

It is true that according to Rawls the agreement behind the veil of ignorance is a mere mental experiment (see footnote number 6), so a formal agreement should not be necessary at all in order to derive principles of justice. Therefore, that each generation is represented by only one subject might be considered a reasonable and representative simplification, because the veil of ignorance turns the choice in the original position into an individual choice (or a choice from the point of view of a representative individual). However, at the same time, from an experimental point of view, it seems to be also too ambitious to simulate a fictitious intragenerational agreement with the unilateral decision of one single person, above all if we consider that without a formal agreement we cannot apply the model of conformist preferences.

Notwithstanding the simplifying and, in a certain degree, imprecise design and its discouraging outcome, Wolf and Dron provide a valid basis to inquire Rawls's intergenerational theory, also because they follow a widespread practice within the economic experimental literature, that is to simulate generations assigning players to different positions in a sequential game[9].

Thus, in order to design our experiment we started from Wolf and Dron's sequential dictator game. We improved it taking into account the further specifications made by John Rawls concerning the nature of the impartial agreement for intergenerational principles. In particular we focused our attention on the structure of the agreement: even though behind the veil of ignorance contractual parties assume an intergenerational perspective (since they do not know the generation of the history they belong), all of them strictly belong to the same generation, that is the impartial agreement is intragenerational and involves only contemporaries.

In our particular design the generations are constituted by groups of three participants[10]. The groups are then randomly assigned to a position in chains (sequences) of different lengths[11]. Essentially, every group is meant to represent a (non-overlapping) generation of contemporaries. Starting from the first generation players are asked to play a group sequential dictator game.

9 This practice has indeed occurred in trust games (Schotter et al. 2006), within public good games (Baggio et al. 2018 and Chaudhuri et al. 2006), with ultimatum games (Schotter et al. 2007) and with common pool resource games (Chermak et al. 2002 and Fisher et al. 2004). Usually, in this kind of intergenerational experiments, there is not any strategic interaction between players belonging to different positions in the sequence, because later generations cannot directly influence the payoff of the previous ones - while the opposite is true.

10 According to the introductory framework, in the experiment the single parties should be considered representatives of nation-states, like in the Law of People (Rawls 2001a, p. 10).

11 The shortest chain was made of one group (generation), the longest one of a sequence of five groups (generations).

During the game nobody can know the total length of her own chain because it is never communicated[12]. However players can deduce how many generations exist (how many groups play) before their own enters the dictator role since all chains start with the generation number one. For example, if a group is assigned to the generation number three, the players belonging to this group do not know how many generations there might are after theirs, but they know for sure that other two groups have to play before they can possibly take any decision.

The decision-making task for the dictator group is then designed as follows (for further details it is possible to consult Appendix B, which contains the instructions provided to the experimental subjects).

The first group (generation) of each chain has at its disposal a common endowment of €21. Each of the three participants of the group individually has to decide how much money to withdraw from the common pool, choosing an integer value between €0 and €7[13]. The amount that each player claims for himself in this stage constitutes her individual final payment. Finally, after a player makes a choice and before he is revealed the outcome of the group decision he is asked to guess, through an incentivized scheme[14], the decisions taken by the other two players of his group.

After all the 3 players decide how much money to withdraw from the common endowment, one of the two following scenarios occurs:

- if the common pool is left with at leas €6 in total, the chain continues and the next group in the sequence enter the decision-making phase becoming the dictator group. The common endowment is refilled up to the initial value of €21[15] and the new generation faces the same identical decision-making problem described so far;

12 This is a standard practice in experiments of this kind (see for example Fischer et al. 2004 or Hauser et al. 2014). This hidden information basically avoids that generations think about the last one as a pure dummy which is not supposed to take any decision.

13 It is important to remark that given the structure of the game players belonging to the same group are endowed with symmetric opportunities, that is among the players who belong to a single generation there are neither formal nor substantial differences. In the game differences between players are exclusively relevant with regard to the group position in the chain. Therefore within this design one of the two problems linked to the unsuccessful international agreements on climate actions, that is the distribution of individual costs, is basically put aside. Indeed within a situation of perfectly symmetric roles there are not formal reasons to distribute costs unequally. However, this simplification does not make the experiment less useful in order to solve the climate change issue: if nations were not able to reach an agreement in a situation of symmetry, a fortiori we could not expect a widely shared agreement in the case of asymmetric costs and benefits. Therefore our experiment constitutes an important preliminary step in understanding international agreements on the reduction of greenhouse gases.

14 Players with the best guess were rewarded with €2 extra. Given the symmetry of the roles within each group, in order to determine the player(s) with the best guess we adopted a simple sum of absolute distances between the guess and the actual choices. See Appendix B for further details.

15 The technology is known and identical for every group in the sequence.

- if in the common pool players leave a total of €5 or less, the common pool is emptied and the chain breaks up, with the consequences that all the following generations cannot take any decision and do not get paid.

An experimental session lasts up to the point that all chains either get to their natural end or break up.

The minimum material threshold of €6 has a clear interpretation. It simulates the threat embodied in the climate change: if the present generation overexploits the environment and does not constrain itself in consuming some available resources which actually increase its own welfare, it does that at the expense of all the future generations. On the contrary, if the players of the group called to take a decision (present generation) coordinate for not overexploiting the environment and for leaving a minimum amount of resources for the next generation, the latter can enjoy the same opportunities as the former. However, some clarifications are due about what was described so far.

It might be argued that the endowment level and the minimum threshold should not be fixed amounts, but they should be rather represented by a spectrum of values, proportionally adjusted according to the resources left by the previous generation. Thus, if one generation complies with the minimum threshold, the next generations should deal with the same situation (same endowment amount and same threshold). And this is true also for our experiment. However, if the current generation saves more (less) than what required by the minimum threshold, the next one should have a higher (lower) endowment and face a higher (lower) threshold. In other words, the effect of the savings should not have a binary effect of zero or one on the continuation of the chain, but they should affect the possibilities of the following generation in a proportional manner.

This kind of design, despite being less dramatic and at the same time more realistic than ours, leads to two complications that are not indifferent from the experimental point of view. First of all, in the veil treatment (see the next paragraphs) participants should vote a set of multiple principles and thresholds, depending on the possible levels of the endowment. This would complicate the agreement phase, unless we introduce very general principles like “withdraw the amount which makes your chain sustainable over time” against “save as much as you wish being unconcerned about the effects that your decision might have on your chain”.

Second, varying the endowment levels, and therefore the thresholds across generations would make heavier the interpretation of the data with regards the ex-post phase. Indeed, we could not really compare, for instance, the choices of a generation with an endowment of €21 and a saving threshold of 6€, one with an endowment with €24 and with a sustainable threshold of €7 and a

generation with only €12 as endowment and a target of only €3, because they would depend too much on the choices made by the previous generations.

Thus we modelled the experiment in order to avoid the mentioned issues and the choice of refilling the pool up to the same initial level (however not new in the experimental literature, Hauser et al. 2014), despite it might appear unrealistic, it was made to facilitate the agreement framework in the veil treatment and to facilitate the interpretation of the compliance task. Indeed, having one unique and certain level of endowment ensures that each group reaches the agreement in the same structural conditions. Besides, in this way the compliance decision is less dependent from the actions taken by the previous generations.

In short, we tried to simplify as much as possible de decision-making frame, such that the participants could entirely focus their attention on the main concern: in order to take into account the interests of future generations in a fair way, avoiding the global warming consequences, each generation must coordinate (constrain itself) to reduce today the consumption natural resources. Instead, if active players (those who enter the dictator role time after time) do not take into sufficient consideration the interests of the following groups, the former can seriously harm all future generations.

At the end of the experiment a general socio-demographic questionnaire was provided.

4.2.1 The veil treatment and its interpretation

The veil treatment, which adopts the Rawlsian insights to address the concern for future generations, adds to the baseline treatment described above a preliminary stage where the three players of every single group have to reach an (intragenerational) agreement in order to enter the (intergenerational) group dictator game. In particular, at the beginning of the experimental session every group is asked to unanimously agree on one the two following rules dealing with the management of the common endowment:

- Continuation of the chain: each participant of my group should withdraw a maximum of €5 from the common account, ensuring in this way a minimum total saving of €6 that allows the chain to continue[16]

16 This rule is meant to represent scheme consistent with Rawls's just saving principle.

- Interruption of the chain: having the possibility to do it each participant of my group should withdraw from the common more than €5, even if that means interrupting the chain.

As mentioned, in order to enter the sequential game every group is asked to reach a unanimous agreement behind a veil of ignorance, that is before being assigned to a position in a chain. Therefore, consistently with Rawls's setting, while groups of contemporaries vote for a principle aimed at managing the appropriation of common resources over time, they do so not knowing the generation (position) they belong in the history (chain).

Before moving to the predictive hypothesis, two specific features of the veil treatment deserve attention.

In the first instance, we need to clarify the interpretation we give either about those groups who might not reach an agreement in the voting phase of the veil treatment or about those groups who formally reach an agreement, but cannot de facto play any game because a previous group of players left less than €6 in the common pool, breaking in this way the chain up.

With regard to the first case, the interpretation seems to be quite intuitive. Groups (generations) that do not reach an agreement end up with living in the so called "state of nature". In other words generations who do not agree to enter a society built on the cooperative attitude and on mutual advantage enter anyway the intergenerational chain (the history), but they do not put themselves in the minimum essential condition to exploit the available resources (the common endowment of €21). We have to imagine a situation where fossil fuels are fully available in the nature and ready to be exploited. Nevertheless, the generation of people who did not reach a preliminary agreement can only look at those resources without being able to "touch" them, because the interested parties did not agree on how to cooperate in order to organize their extrapolation. Thus, people who did not reach an agreement come to the existence but they live in poverty because they cannot exploit the available resources.

As far as the second case is concerned, the interpretation seems to be even more straightforward. Even if a group agreed to enter a society based on the cooperative ttitude, they end up with living in minimal conditions as well. However this time it is not for their own (missing) willingness too coordinate and too cooperate, but because some of the previous generations did not leave enough resources to allow their society to be wealthy enough. In this case we have to imagine a situation where the unlucky society, despite having reached an agreement, observes that fossil fuels (the common endowment of €21) are not physically available because previous generations

overexploited the nature. In this perspective the unlucky generation basically pays the consequences of the global warming generated by the previous unconstrained behaviours.

Thus, the two mentioned situations are identical about the substantial material consequences of the existing generations. In none of the cases they can enjoy the common endowment. However, they are the result of different causes: they do not coordinate in the first case; they suffer the decisions of other groups in the second case.

Second, it is important to remember how the agreement reached behind the veil of ignorance is not conceived by Rawls as binding. Said otherwise, after the veil is dropped and groups are assigned to a chain and to a position, the outcome of the agreement is not automatically implemented (like it was did in other intergenerational experiments, e.g. Hauser et al. 2014[17]).

Thus, in our design, generations that are called to take a decision are not constrained by any external enforcement mechanism to apply the principle of the agreement reached behind the veil of ignorance. This implies that in the veil treatment the sequential dictator game exactly replicates the baseline treatment, and the compliance to the agreement is left to an individual choice. Again, this is a realistic structure since in the real world we have no formal institutions which can substantially constrain the present generation to care about the future ones, even if formal intergenerational norms are the outcome of a fair procedure.

4.3 Predictive hypothesis

Given the theoretical framework and the experimental design described so far it is possible to formulate the predictive hypothesis of our game. Our first hypothesis regards the baseline treatment and it follows from standard economic assumptions[18]. Without any other formal element in the

17 The mentioned practice is considered unrealistic. Even though the authors justify the binding vote as a good proxy for informal institutions which usually enforce cooperative attitudes (like punishments or rewards), those enforcement mechanisms work only when there are repeated interactions among the same subjects, so that paying a cost now (punishing) can generate long-term benefits. Indeed, this kind of institutions cannot be as much effective in one-shot games as in repeated games. In their Intergenerational Good Game there are no reasonable motivations (except maybe spitefulness) to punish another player who did not comply with the approved rule since there will not be a second chance to interact. A player could only loose utility by materially punishing somebody else belonging to the same group (generation). Therefore, the pretension to assimilate a binding vote to an informal institution which can enforce cooperation has to be considered ambitious, at least for a context simulating an intergenerational game played sequentially.

18 Except for special cases (Bardsley 2008, Cherry et al. 2002 and List 2007) threshold established in our experiment (6/21) mirrors exactly the average amount of money left by dictators (28.5%) to dummy players (Engel 2011). However, the impossibility to coordinate represents a non.-indifferent obstacle to allow chains to continue, since

game, the sub-perfect equilibrium is represented by the triple (€7, €7, €7) for every (so also the first) generation in any chain. Therefore players in the first generation participants are expected to appropriate the total available endowment, leaving no resources in the common pool. This means that the chains will not be sustainable, that is they will not continue after the first generation, because the choices of the first group undermine the entire scheme of cooperation over time. Thus,

H1: in the baseline treatment the generations number 1 will mostly break all the chains up

Our second hypothesis concerns the veil treatment ant it directly follows from the Rawlsian theory. As we have seen, according to Rawls, groups behind the veil of ignorance should agree "on a path over time which treats all generations justly during the whole course of a society's history" (Rawls 1999, p. 257). More precisely, players should agree on a just saving principle, according to which "each generation makes a contribution to those coming later and receives from its predecessors" (Rawls 1999, p. 254). Therefore, in order to guarantee a positive amount of savings to each generation, allowing in this way the chains to continue, behind the veil of ignorance most of the groups are expected to agree on a rule which somehow constrains a pure self-interest behaviour. In short, participants will agree that each individual should withdraw a maximum €5 from the common endowment.

H2: in the veil treatment, because of the impartial perspective, during the voting phase groups will mostly agree on the rule representing the Rawlsian just saving principle which guarantees the continuation of the chain

The third assumption follows as much from Rawls's social contract theory as from the experimental literature based on it: although the agreement behind the veil of ignorance is not conceived as binding, individual compliance to the chosen principle is expected to be high even in those cases where players face a counter-maximizing situation. Instead, this is not true for the standard economic theory. Indeed, since the agreement does not introduce any formal constrain mechanism, every individual in the decision-making phase should follow his purely selfish impulse, claiming €7 regardless of the chosen principle behind the veil of ignorance.

In particular, the so called exclusion game (Degli Antoni et. al 2016, Faillo et al. 2008, Faillo et al. 2014, Sacconi and Faillo 2005, Sacconi and Faillo 2010 and Tammi 2011) inquired from an

one purely selfish dictator can nullify the effort of the other two.

experimental point of view the Rawlisan concept of the sense of justice (Rawls 1999). The exclusion game is a one-shot resource allocation game contemplating a preliminary voting stage carried out behind the veil of ignorance, where the participants are prevented from knowing their role in the actual game (dictator or dummy).

In the game the agreement concerning the sharing rule is not binding, therefore players who are assigned the dictator roles are supposed to pursue their own interest regardless of the rule agreed in the voting stage. This is a clear analogy with the situation faced by the present generation in our intergenerational experiment. The experimental evidence of the exclusion game showed how the (unconstrained) ex-post compliance with the ex-ante chosen distributive norms is unexpectedly high even in those cases where groups agreed on an egalitarian (counter-maximizing) distributive rule[19]. Therefore we expect that

H3: in the veil treatment the individuals will comply with the rule agreed with the intragenerational agreement

The fourth and last hypothesis becomes a logical sum of the previous two: if groups agree on an intergenerational sustainable behaviour (just saving principle), and if they comply with the chosen norm, chains will continue up to the last generation.

H4: compared to the baseline treatment, in the veil treatment a significantly higher number of chains will continue until their natural end

The just mentioned hypothesis might be also formulated in other two equivalent manners

H4a: the proportion of people claiming an amount of €5 or less will be significantly higher in the veil treatment than in the baseline condition

H4b: the average individual claim will be lower in the veil condition than in the baseline treatment

19 However, it is important to keep in mind the dissimilarity between the exclusion game and our intergenerational agreement: the agreement of the former includes all the least advantaged individual (the dummy player), while the latter in the voting phase leaves out the direct stakeholders. This might have an impact on the reciprocal conformity and therefore on compliance. See also footnote number 5.

4.4 Data analysis and comment

All the experimental sessions took place in the Computable and Experimental Economics Laboratory (CEEL) of the University of Trento. They were run using the free software for economic experiments zTree (Fischbacher 2007). All the participants took part in the experiment after a public call.

In the experimental laboratory participants were randomly assigned to a computer terminal. All the emplacements were isolated by separation walls to avoid communication between the individuals. The participants were given paper instructions and the instructions were also read aloud to ensure common knowledge. In the final questionnaire participants declared on average that the provided instructions were very clear (4.6 on average, where 1 = not at all clear, and 5 = very clear). Before the actual experiment could start, in the baseline (veil) treatment 4 (6) control questions about the structure of the game were asked.

The experiment involved a total of 141 participants (60 in the veil treatment and 81 in the baseline treatment). On average the participants were 22 years old, 54% of them were females and 46% of the total participants were enrolled in programs related to the economic discipline, the rest in other fields going from humanities to natural sciences. The participants were privately paid in cash at the end of each session and on average they earned about €6 (included the show-up fee of €3). Each experimental session lasted most 50 minutes.

In the baseline treatment we run 4 sessions with a total of 8 chains distributed as shown in Table 1.

Table 1 – Distribution of chains per session in the baseline treatment

Session	Number of chains	Chains' length (n. generations)
1	2	2 and 5
2	2	2 and 4
3	2	2 and 5
4	2	3 and 4

In the baseline treatment 6 chains out of 8 (that is 75%) broke up after the choices made by players belonging to the first generation. In the two remaining chains (session 1 and 3), despite having only two generations, the second generations substantially behaved such that they would

have broken their chains up. Thus out of 81 participants only 30 (37%) of them played in the active role of the game taking an actual decision. These first data are sufficient to support ***H1***: in the baseline treatment chains (mostly) brake up after the first generation.

In the baseline treatment active players withdrew on average €5.30 from the common endowment believing that the other two players in the group claimed €5.20 on average. These last data shows an interesting empirical regularity which was not taken into account by the predictive hypothesis: generations in the baseline treatment waste resources[20], because on average they left in the common pool 1.70€. In the game this amount of money is substantially destroyed since it is not distributed to anybody. Indeed, despite left by the active players with the hope to support the next generation, that average amount was not enough to allow the chains to continue.

Seen from another point of view, we can claim that some players are altruistic individuals. Indeed, they renounce to consume (to withdraw) a part of their individual endowment without any possibility to be directly reciprocated by the future generations. More specifically, it can be said that they are intergenerationally altruistic, because they take into account the interests of possible future generations at expenses of their own material payoff. However their good purposes are nullified by the actions of a minority of players who do not do their part in contributing to the savings.

With regards to the preliminary voting stage of the veil treatment all groups reached a unanimous agreement on a rule by the second round (out of six available). In particular, 18 out of 20 groups agreed on the sustainable rule labelled “continuation of the chain”. Therefore ***H2*** is supported by the data too, because most of the participants (54 people, representing the 90%) voted for the just saving principle.

In the veil treatment there was a total of 6 chains divided in 3 sessions (Table 2).

Table 2 – Distribution of chains per session in the veil treatment

Session	Number of chains	Chains' length (n. generations)
1	2	2 and 5
2	2	2 and 4
3	2	3 and 4

20 In general all active groups left in the common pool at least €2. A total of €36 was left to the experimenters.

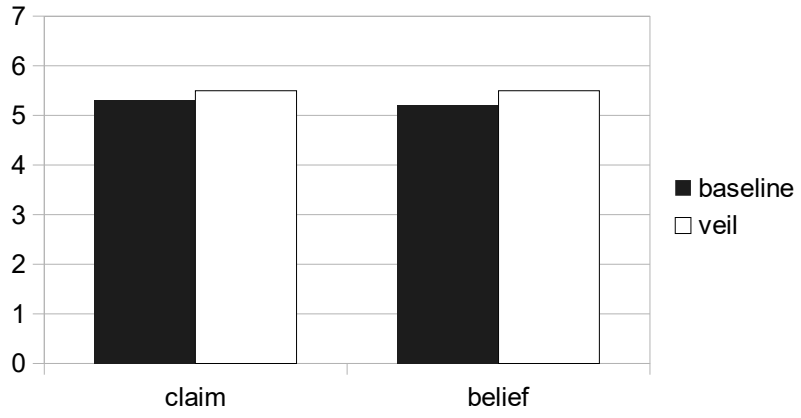
After the agreement phase 80% of those participants who could take a decision complied with the rule chosen behind the (laboratory) veil. Thus, data support also ***H3***, because the majority of the participants followed the approved rule even when that was against their own material interest.

Nevertheless, although participants mostly choose the just saving principle and even though they mostly complied with that rule, as in the baseline treatment in the veil treatment no one chain continued after the second generation. Therefore, ***H4*** is rejected by the empirical data, because no one chain reached its natural end as predicted.

That ***H4*** is not supported follows also from the rejection of the equivalent hypothesis. ***H4a*** is rejected since between treatments the proportion of players claiming an amount of €5 or less is exactly the same and corresponding to 0.7.

In the same way ***H4b*** is rejected because in the veil treatment the players claimed for themselves an average share of the common endowment not statistically different (two tails Welch t of student test is t= -0.78, with a corresponding p-value of 0.44) from the baseline treatment, withdrawing on average €5.50[21]. The comparison between treatments is shown in Chart 1.

Chart 1 – Individual average claim and average belief (€) per treatment



The undifferentiated behaviour which leads to the rejection of ***H4b*** and therefore of ***H4*** can be highlighted also with a simple linear model (1) where the individual claim is regressed on the treatment (dummy variable, where the baseline assumes the value 0), on the generation (position in the chain), on the average belief (in €) plus a series of control demographical factors. The model clusters the standard error per single groups (Table 3).

21 In the veil treatment the broken chains left to the experimenter a total of €21. Therefore in the veil treatment active players wasted on average €1.40.

$$(1)\ individualclaim_i = \alpha + \beta_1 treatment_i + \beta_2 generation_i + \beta_3 averagebelief_i + controls_i + \varepsilon_i$$

From the econometric estimation of the model we can see that there is not any statistically significant difference in the average claim between the two treatments. Among the main predictors, only the average beliefs on the others' withdrawal and the generation number had a strong and significant effect on the individual choice. In other words, a higher expectation of €1 on the others' withdrawal and belonging to the second generation (the last generation who actively played in our game) in both cases increased the average claim of about €0.5.

Even though the veil treatment did not show the expected effect on the individual claim and therefore on the chains sustainability, there is a further interesting comparison between the two conditions we may look at. In particular, in the veil treatment only 50% of the chains were interrupted by the first generation (compared to the 75% of the baseline treatment). Thus, thanks to the agreement 45% (27 out of 60) of the participants who took part in the veil treatment had the possibility to make an active choice, while that ratio lowers to 37% (30 out of 81 subjects) with regards to the baseline treatment. However, in the perspective of our experiment, these results do not lead to different conclusions given that they are not statistically significant (chi square test 0.29 on the proportion of participants who actively chose).

4.5 A less radical experimental design

The results showed in the previous sections highlighted what should be considered a limit of our experimental design: even though in the veil treatment of our experiment the average compliance rate (about 80%) is high and consistent with the previous empirical evidence provided by the exclusion game, this percentage seems to be low compared to the 100% of compliance (to the just saving principle) necessary in the intergenerational game to allow the chains to continue.

In other words, in our experiment a society is sustainable, that is a chain can continue only if compliance to the just saving principle is total. This situation is very interesting from an ethical point of view, because climate change may present us with a singular situation wherein a high percentage of compliance is not enough to protect the future generations against possible disasters, such that a little degree of deception can cause very bad consequences for the whole humanity. However, this theoretical result seems to be too demanding from a pragmatic point of view: it is utopian to expect perfect voluntary compliance, as much in an experiment as in the real world.

Table 3 – Determinants of individual claims

	Full model	Pooled model
treatment	0.18 (0.296)	
generation	0.53 (0.267) *	0.53 (0.266) *
average belief	0.48 (0.113) **	0.50 (0.119) **
clearness instruction	0.35 (0.134) **	0.38 (0.125) **
number previous experiments	0.04 (0.005) **	0.03 (0.007) **
age	- 0.07 (0.080)	-0.08 (0.066)
gender	- 0.39 (0.264)	-0.37 (0.268)
nationality	0.75 (0.307) **	0.81 (0.265) **
field of studies	- 0.4 (0.218)	-0.07 (0.206)
years of studies	0.11 (0.108)	0.13 (0.106)
risk attitude	0.04 (0.037)	0.04 (0.038)
family yearly income	0.24 (0.110) **	0.25 (0.111) **
general economic situation	0.55 (0.213) **	0.45 (0.200) **
Participants 141		
Number of observations 48[b]		
* significant at 5%		
** significant at 1%		
Adjusted R-squared	0.6013	0.6052
F-statistic p-value	6.5e-07	2.9e-07

a The variable *instructions* refers to the comprehensibility of the initial instructions (ranging from 0 to 5), *experiments* is the number of previous experiments, *age* the personal age, *gender* is a binary dummy for the gender (male = 0), *nationality* is a dummy variable for the nationality (0 Italian, 1 foreigner), *field of studies* and *year studies* refer to the undertaken studies (the first one is a dummy variable, where the value 0 corresponds to economic disciplines, whereas 0 to any other field, the second one is a discrete variable), *risk* is a subjective statement about the risk attitude (1 risk adverse, 10 risk seeker), *income* and *economic* situation measure respectively the yearly family income and the current economic status (both are discrete variables and range from 1, very low/bad, to 5, very high/excellent).

b The number of the observations is lower than stated so far because we could not collect the demographic statements of three participants in the baseline treatment, so the model drops those data. However, those three observations did not influence the outcome since the model considering only the three fundamental variables (treatment, average belief and generation) do not differ from the one presented here

Besides this speculative consideration, our non-positive result requires a deeper reflection which goes beyond the simple compliance rate. In particular, if we look closer at the distribution of the withdrawals in the two treatments (Table 4) we can better understand why our institutional mechanism of the veil of ignorance did not work as predicted. In the Table 4 within-group claims were ordered in ascending triples order, moving from the left column (player with a low claim) to the right (player with a medium and high claim). Furthermore, we highlighted in grey the participants who withdrew €6 or €7.

Table 4 – Distribution of individual claims (average belief) per treatment and group

BASELINE TREATMENT		
Player low	**Player med.**	**Player high**
4 (4.5)	4 (4.5)	4 (4.5)
4 (4)	5 (2.5)	7 (6)
5 (5)	7 (6.5)	7 (7)
4 (5)	5 (5)	7 (6)
4 (5)	5 (4.5)	7 (6)
5 (5)	5 (5)	5 (7)
5 (5)	5 (5)	6 (7)
4 (4)	5 (5)	7 (7)
5 (2)	5 (5.5)	6 (6)
5 (5)	5 (5)	7 (7)

VEIL TREATMENT		
Player low	**Player med.**	**Player high**
5 (5)	5 (5)	5 (5)
5 (5)	5 (5)	7 (5)
4 (4.5)	5 (5)	5 (5)
5 (5)	5 (5)	7 (5)
5 (5)	5 (7)	7 (7)
4 (5.5)	5 (5)	7 (5)
5 (5)	5 (5)	5 (5)
5 (5)	5 (6.5)	7 (7)
7 (7)	7 (7)	7 (7)

As pointed out also earlier in the Chapter, given the structure of our game one high claim (€6 or €7) in a group is sufficient to undermine the entire scheme of intragenerational cooperation necessary for the intergenerational continuation of the chain. In particular, from the Table 4 we can clearly observe how the agreement phase of the veil treatment did not produce any effect on the participants who seek to withdraw a high amount from the common pool. Therefore there is a share of subjects which is somehow immune to the impartial agreement since the last one cannot change

their psychological equilibrium. This observation, together with the consideration that it is necessary a total compliance for making the chains sustainable over time, is probably at the basis of the unwanted results of our experiment.

In order to mitigate the veto power in the ex-post phase described above, we designed a new experiment. In the new design the chains were not broken definitely up by the defection of one player in the group. Instead, every chain was shortened proportionally to the savings left for the next generation. This time the common fund was of €18 and the threshold to leave the chain unaltered was settled at €9 (50%of the initial endowment). If the group of one generation saved less than the threshold level, the chain was shortened of one generation if they saved between €9 and €6 (excluded), of two generations if the group saved between €6 and €3 (excluded), of three generations if the group saved €3 or less.

This structure implies that the decisions of a single player could have limited consequences on the destiny of all future generations. We also modified marginally the two rules on which the participants were asked to vote to adapt them to the new design. We left unaltered the remaining parts of the initial design (groups formation, elicitation of beliefs etc.).

This time we involved 81 participants of the University Milano Bicocca to take part in the new veil treatment. 60% of the participants were male and 75% of them enrolled in courses related to the economic discipline. On average they were 24 years old.

We run the experiment in the computer lab of the Department of Economics of the University and paid the participants with book coupons they could spend in a library close to the mentioned Department. We run three separate sessions including two chains constituted of two generations, three chains of five generations and two chains of four generations. At the end of the experiment the average payoff was about €6.

This time in the preliminary voting stage only 16 groups out of 27 (59%) agreed on the just saving principle, casting doubt on the solidity of our ***H2***. Besides, among the groups who unanimously agreed of the just saving principle and took an active decision, only 14 participants of 33 (42%) complied with the rule. Thus the average claim was of €4.30[22], not sufficient to allow the generational chains to continue until their natural end: indeed none of the chains continued after the second generation.

Given these preliminary results, we decided not to run the baseline treatment for this version of the experiment.

22 The average belief of those participants was of €3.50.

Conclusions

Relying on Rawls's intergenerational social contract theory and some of its experimental evidence, through our laboratory experiment we tried to address the modern issue regarding the climate change agreements. In doing so we left aside the problem concerning the distribution of costs between contemporaries (nations) assuming symmetrical situations between the players belonging to the same generational group. Therefore we focused on the pure intergenerational problem, trying to see if the Rawlsian theory could help to structure a fair intragenerational agreement for the intergenerational distribution of resources.

The experimental results showed that a laboratory veil of ignorance induces people to reach an ex-ante fair agreement concerning the management of common resources over time and that the voted rule was consistent with Rawls's just saving principle. At the same time, however, in the first version of the experiment the compliance to that principle, despite being high, was not sufficient to allow chains to survive significantly longer than in the baseline treatment, where no agreements were possible.

Even after we modified the experimental design in order to mitigate the dramatic effects of the partial compliance observed in the first experimental version, the general outcome did not change, because the chains always interrupted within the first two generation. However, the common result of the two experiments leads to two conclusions that somehow differentiate. In particular, in the first design the conformity model seemed to work properly, but the general structural conditions were too demanding to produce the expected results. Vice versa, in the second version of the experiment the model of conformist social preferences did not seem to work as supposed, because beliefs on others' compliance did not play a significant role. Thus, in both cases the veil of ignorance was not able to produce the sufficient conditions for a sustainable society over time, but for different reasons.

In general however, it seems that the sense of justice to an impartial agreement is not triggered when a set of people *X* agrees to undertake a specific redistributive action towards a set of people *Y* (different from *X*). In other words, when the subjects representing the weakest part are not included in the formal agreement, the conformist preference model is made lane and cannot show its full power. As highlighted several times, within the Rawlsian theory a set of contemporaries is called to evaluate distributive principles that will produce effects on third parties (future generations) who formally cannot take part in the contract. And this structure exactly implies that

the ex-post redistributive actions of mutual advantage (the savings) are not directed towards the same set of people (the contemporaries) who agreed on the intergenerational principle.

In conclusion, the Rawlisan sense of justice, based on mutual expectation of compliance and which is supposed to be the glue that fosters the general compliance to an impartial redistributive principle, does not enter into play when a set of individuals has to choose distributive actions towards some subjects who are left outside of the agreement itself. This might be do to the fact that in the ex-post phase the outside-parties cannot enter a mutually beneficial relationship with the agreement-parties. This is quite evident in the intergenerational setting, where the future generations cannot directly cooperate with the present one, and the absence reciprocal expectations of compliance between generations might undermine the intergenerational fairness.

These are only preliminary conclusions we can draw from the analysis of our experiment, but further research seems to be necessary to disentangle better the sense of justice between the two situations where the weakest parts are included or not in the agreement moment, because without the sense of justice there cannot be compliance to the redistributive principle; and without compliance we cannot expect chains to continue.

At this stage, the big issue on how to structure international agreements on climate actions keeps open.

Appendix A: Rawlsian intergenerational justice and derivation of the just saving principle

The starting point of the reflection of John Rawls concerning the allocation of resources between generations is an extension of the main hypothesis of his social contract theory, which portrays the human society as a cooperative venture for the mutual advantage (Rawls 1999, p. 4). Thus Rawls assumes that "life of a people is conceived as a scheme of cooperation spread out in historical time" (Rawls 1999, p. 257, emphasis added, Rawls 1977, p. 161 and Rawls 2001). Therefore, according to Rawls, it is necessary to agree "on a path over time which treats all generations justly during the whole course of a society's history" (Rawls 1999, p. 257).

Since for Rawls "persons in different generations have duties and obligations to one another just as contemporaries do" (Rawls 1999, p. 258) and since according to him justice between generations "is to be governed by the same conception of justice that regulates the cooperation of contemporaries" (Rawls 1999, p. 257) it might be reasonable to extend the standard (intragenerational) principles of justice (Rawls 1999, pp. 47-101)[23] over the time dimension.

Even more relevant for the present discussion about redistribution of resources, given the just mentioned similarities, it might be intuitive to adopt the canonical formulation of the so called difference principle[24] to regulate the allocation of resources between generations. After all Rawls himself explicitly claims how the "appropriate expectation in applying the difference principle is that of the long-term prospects of the least favored extending over future generations" (Rawls 1999, p. 252). Thus it seems that the difference principle, when fully applied, has to take into consideration and to operate on two dimensions, space and time.

However, almost contradicting his own claims, Rawls remarks through many passages how the difference principle's prescriptions have to be realized exclusively within an intragenerational context. Indeed, in Rawls's opinion, the difference principle is inadequate to discipline the allocation of resources between generations because of its undesirable consequences: "for when the difference principle is applied to the question of saving over generations, it entails either no saving at all or not enough saving to improve social circumstances sufficiently so that all the equal liberties can be effectively exercised" (Rawls 1999, pp. 253-254)

23 "The first principle simply requires that certain sorts of rules, those defining basic liberties, apply to everyone equally and that they allow the most extensive liberty compatible with a like liberty for all" (Rawls 1999, p. 56), meanwhile the second principle of justice, the so called difference principle, prescribes to "maximize the expectations of the least favored position" (Rawls 1999, p. 69).

24 For a complete presentation of the difference principle see (Rawls 1999, pp. 52-65, pp.130-9 and pp. 153-160).

In other words, “since the persons in the original position know that they are contemporaries [...] they can favor their generation by refusing to make any sacrifices at all” for the others (Rawls 1999, p. 121). Here Rawls tries to suggest that since the parties involved in the agreement (who are contemporaries) are instrumentally rational and they desire to maximize first of all their own expectations[25], it is not legitimate to expect, from the generation involved in the agreement, any renounce of resources which could benefit (the least advantaged) people in another generation (Rawls 1999, pp. 254-255, Attas 2009, p. 190 and Buchanan 1987, p. 250). However, this way of reasoning cannot be compatible with the most deep meaning of the difference principle itself (Dasgupta 1974, pp. 330-337)

Besides, even if a difference principle was conceivable for the intergenerational framework, there would be no way to act on the past (Brandstedt 2017, p. 270), that is, the criterion could be applied only from the moment of the “entry in society”[26] onward, while it would be impossible to carry out its prescriptions towards any previous generation. For example, if the least advantaged subjects, after the veil is dropped, were located in the past (in a moment “before” the agreement), there would not be any concrete way to fully realize the difference principle’s (intergenerational) prescriptions[27]. As for this point is concerned Rawls is extremely clear: “there is no way for later generations to help the situation of the least fortunate earlier generation” since “it is a natural fact that generations are spread out in time and actual economic benefits flow only in one direction” (Rawls 1999, p. 254).

Therefore in Rawls’s opinion the difference principle is not a suitable tool to deal with the redistribution of resources across time: “thus the difference principle does not hold for the question of justice between generations and the problem [...] must be treated in some other manner” (Rawls 1999, p. 254). Rawls therefore proposes the just saving principle as the normative solution to guarantee intergenerational redistributive fairness. Thus “the difference principle holds within generations” while “the principle of just saving holds between generations” (Rawls 2001, p. 159).

“The just savings principle applies to what a society is to save as a matter of justice” (Rawls 1999, p. 255) and “in following a just savings principle, each generation makes a contribution to

25 In terms of primary social goods, that "are things which it is supposed a rational man wants whatever else he wants" (Rawls 1999 p. 79). See also (Rawls 1999, pp. 78-81)

26 As highlighted many times by Rawls, the original agreement is hypothetical, therefore the words "moment", "before", "after" and so on and so forth have to be taken with the right caution and interpreted coherently with the context.

27 This way of reasoning is in line with the idea that within an intergenerational context “ought implies can” (Partridge 2017) and the worst-off should be accessible (Gaspart et al. 2007, p. 203), no matter the generation they belong .

those coming later and receives from its predecessors" (Rawls 1999, p. 254). However Rawls does not describe in detail the particular features of the mentioned intergenerational principle (like for example providing a specific saving rate or a scheme of rates). Instead he limits himself to sketch some general ethical restriction that the contractual parties should take into account in defining the saving path (Rawls 1999, pp. 255-6).

Nevertheless here it is not of particular interest to linger on those, although reasonable, ethical concerns. Instead it is relevant to understand which is the positive reasoning offered by Rawls to substitute the inadequate difference principle with the just saving principle. He essentially proceeds in two parallel steps:

- first of all Rawls restates the intergenerational redistrivutive problem. The parties behind the veil of ignorance are aware of the natural flow of the economic benefits (which is a general and unalterable circumstance), therefore the new issue becomes to understand how the generation involved in the agreement can fairly treat not all the possible generations of the history but only the subsequent ones;

- second, in order to induce the subjects involved in the agreement to think not only as contemporaries but to take into consideration also the future generations, Rawls amends his own theory and adds an intergenerational motivational interest assuming that "the parties represent family lines", that is, they "care at least about their more immediate descendants" (Rawls 1999, p. 255 and Brandstedt 2017, p. 276).

These are the further specifications introduced by John Rawls in order to deal with the peculiarities regarding the distribution of resources between generations within his social contract theory. Thus, reminding how the subjects behind the veil of ignorance are unaware of the historical period they belong and adding to this premise a carefulness for the closer future generations (the "family line" assumption), the parties involved in the agreement naturally derive the just saving principle[28].

One important feature of the just saving principle is its duration. In fact, the principle is not required to apply forever, but the resources have to be moved towards future generations only until the specific task which the principle was designed for is accomplished. In particular "once just

28 Rawls concludes his intergenerational theory adding a really important elucidation concerning the just saving principle, outlined as a formula to represent the duty to sustain just institutions across time. In particular he specifies how "the difference principle includes the savings principle as a constraint" (Rawls 1999, p. 258). That means that before applying the difference principle it is necessary to fulfil the requirements of the just saving principle. Said otherwise "the just savings principle demands that we leave enough capital and resources for future generations while making transfers to our contemporary poor (as required by the difference principle)" (Heyd 2009, p. 171).

institutions are firmly established and all the basic liberties effectively realized, the net accumulation asked [by the just saving principle] falls to zero" (Rawls 1999, p. 255). Therefore, "the just savings principle can be regarded as an understanding between generations to carry their fair share of the burden of realizing and preserving a just society" and "the end of the savings process is set up in advance" (Rawls 1999, p. 257)[29].

In this perspective, the Rawlsian intergenerational theory shows to be essentially structured in two distinct stages: a temporary phase of accumulation where the just saving principle applies; a steady state where it is not required to apply any particular intergenerational (redistributive) principle (Gaspart et al. 2007, pp. 193-197, Gosseries 2008, pp. 18-19 and Gosseries 2016, pp. 79-85).

While one fringe of the secondary literature almost uncritically accepted Rawls's conclusions on the just saving principle, taking them as the basis for further theories (Arrow 1973, Dasgupta 1974 and Solow 1974), the majority of the authors showed instead some perplexities about the approach adopted by Rawls to deal with the redistribution of resources between generations within his social contract theory.

For most of the authors Rawls's approach to intergenerational justice appeared to be limited and unsatisfactory in its deductions (Mathis 2009, Paden 1997 and Partridge 2017). With his intergenerational framework Rawls was considered to reach "a modest conclusion" (Heyd 2009, p. 187) since he did not provide any particularly elaborated intergenerational distributive theory (Gosseries 2016, p. 87). Thus, for the critical literature Rawls substantially failed to apply the veil of ignorance design to the intergenerational context (Tremmel 2009, pp. 149-154).

The general disappointment is then ascribable to different specific critiques. For example, Gardiner (2009, p. 81) claims that the just saving principle does not treat fairly all the generations because more concern is paid to the future generations. Indeed, the accumulation phases might violate the maximin prescriptions (Gosseries 2016, p. 79) because a very high price is paid by the first generations (Agius 2006, p. 324): this is an implicit utilitarian conclusion that Rawls tries to avoid throughout all his contractarian theory of justice (Rawls 1999).

Besides, the motivational altruistic assumption is to be considered an ad hoc construct (Wall 2003, p. 81) that reflects a conception of the good (English 1977, p. 93) and that undermines Rawls's whole theory since it generates tensions between the Rawlsian intragenerational system and

29 This structure implies that the just saving principle does not pay any direct attention to the worst-off (like the difference principle does) and more in general it is not concerned with the pure redistribution of resources between generation. Instead, its main goal seems to be exclusively to secure the conditions for the realization of just institutions and of a just society (Attas 2009, p. 211, Heyd 2009, p. 187, Gabor 2013 p. 305, Gosseries 2016, p. 80, Paden 1997, pp. 28-29 and p. 38, Wall 2003, p. 93).

his theory supposed to regulate justice between generations (English 1977 and Wall 2003): “the postulate of altruistic interest within the original position therefore compromises the whole systematic derivation from contract theory” (Mathis 2009, p. 54).

Furthermore the artificial trick of the family’s chains substantially eludes the real intergenerational problem since according to some authors it is not possible to derive an adequate concern for the whole future from the thin interest for the own offspring (Heyd 2009, p. 175 and Mathis 2009, p. 54). In other words the “concern for the future cannot be understood in individualistic terms” (Norton 1989, p. 151).

Rawls was not indifferent to some of those critiques and tried to improve his social contract approach to the allocation of resources between generations. In particular he simplified the framework (Wallack 2006, p. 91), but he did not change the main outcome concerning the just saving principle. Thus, following some hints provided by other philosophers (Rawls 2001, p. 160, footnote 39), Rawls dropped the most controversial hypothesis within his intergenerational system, that is the altruistic intergenerational concern.

In the last version of his intergenerational theory Rawls assumes that the full compliance condition of his ideal theory (English 1977, Heyd 2009, p. 179 and Attas 2009 p. 220) is sufficient to guarantee intergenerational fairness: now the parties in the original position are intergenerationally disinterested but they “are to ask themselves how much [...] they are prepared to save at each level of wealth as society advances, should all previous generations have followed the same schedule” and “the correct principle, then, is one the members of any generation (and so all generations) would adopt as the principle they would want preceding generations to have followed it” (Rawls 2001, p.160)[30].

However, despite the new theoretical frame, nothing new was added to the substance of the just saving principle.

30 It is interesting to highlight how more than thirty years before the final version Rawls essentially reached the same conclusion: “the correct principles for the basic structure are those that the members of any generation (and hence all generations) would agree to as the ones their generation is to follow and as the principles they would want other generations to have followed and to follow subsequently, no matter how far back or forward in time” (Rawls 1977, p. 161)

Appendix B: instructions for the experiment in the veil treatment

The instructions for the baseline and the veil treatment are exactly the same for what concerns the sequential game. The latter integrate the former only with the agreement phase.

Good morning,

You are about to take part in an experiment on economic decisions. By participating in the experiment you will be able to earn an amount of money that will depend on your decisions and on those of other participants. The decisions you make will remain completely anonymous and no one will be able to associate your choices with your name. During the experiment you will not be allow to communicate in any way with other participants. In case of communication you will be excluded from the experiment without being paid.

We ask you to read carefully the instructions that have been provided to you and which you can consult at any time during the experiment. The instructions will also be read aloud by one of the experimenters. If at the end of the instructions you will have doubts, raise your hand and wait for one of the experimenters to answer to your questions. At the end of the experiment you will be paid privately in cash. To the payment depending on your decision you will earn an extra €3 as show-up fee.

EXPERIMENT

At the beginning of the experiment you will be randomly assigned to a group of 3 participants (you included). The experiment will then be divided into two phases and will start from the phase 1. However, for clarity, the instructions first show the details of the phase 2

PHASE 2

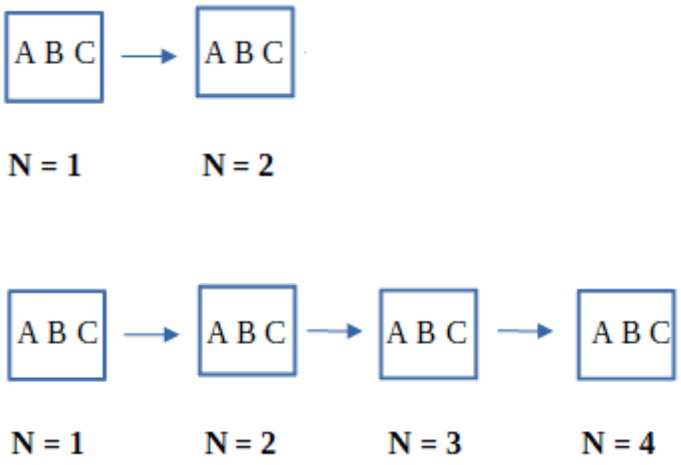
Only groups that pass the phase 1 will have access to phase 2. At the beginning of phase 2, each participant in each group will be assigned an identification letter (A, B or C). All groups that will access the phase 2 will be randomly ordered to constitute various "chains". Within each chain each group of three participants will represent a generation. The chains will have variable length. The

shorter chain will be long 1, and therefore will include only one group, but there may be chains of two, three, four, and more groups.

Each group in each chain will be then assigned a number (N) corresponding to the position of the group within the chain. All chains will start with a group in position N = 1, which will represent the first generation of the chain. To the second generation, if there will be, will be assigned the number N = 2, to the third the number N = 3 and so on for all the groups which compose the chain. Even if you will know the position of your group (N) in your chain, you will not be notified of the total length of the chain, that is you will not know the number of generations that will be there after yours. You will only be certain that within your chain there are other N-1 groups before your group. For example, if your group is assigned to the position N = 3, you know that in your chain there are other two groups in the previous positions, but you do not know how many groups there will be in the subsequent positions.

Below is an example of chains of length 2 and 4

Figure 1 – Example of chains



PHASE 2 – DECISION

Each group that will take a decision in the phase 2 will face the same type of choice. Participants of the group assigned to the first generation (N = 1) will be the first to make a decision within each chain.

At the time of the decision, the group will have €21 available on a common account. Each participant in the group will be asked to decide how many euros s/he wants to withdraw from the common account by inserting an integer between €0 and €7. The sum that you decide to withdraw on this occasion will constitute your final payment, to which we will add the €3 of the show-up fee. In addition, when you will have to decide how much to withdraw from the account you will not know how much the other members of your group have claimed for themselves.

When all the participants will have made their choices, depending on the amount of euros left on the common account two distinct scenarios might happen:

- If the total amount left on the common account by the group (N) will be at least €6, the chain will continue and the next group (N + 1) will enter the decision phase. In this case the sum left by the group (N) will be completely integrated and on the common account of the next group (N + 1) there will be again available €21.

- If the total amount left by the group (N) on the common account will be less than €6, the chain will be interrupted and the common account will be emptied of any remaining euro. In this case, all participants of any subsequent groups (N + 1, N + 2, N + 3 etc.) will not be able to take any decision and their will be paid only the participation fee of €3.

If your group came at the decision phase, after the choice of how much to withdraw from the common account you will also be asked to make a prediction on the behaviour of the other two participants of your group. You will have to indicate the forecast by entering an integer between €0 and €7 for each of the other 2 participants in your group. The participant of the group who will provide the best forecast will get a bonus of €2 that will be added to their final payment. If two (or three) players provide equally accurate predictions, the bonus will be awarded to both (or all three) participants. The best forecast will be defined according to the following rule (imagining that you are the player A):

SCORE_A= (distance between forecast_on_player_B and decision_of_player_B) + (distance between forecast_on_player_C and decision_player_C)

Thus the SCORE can vary between a minimum of 0 and a maximum of 14. The bonus will be won by the participant(s) whose SCORE will be smaller (or equal) than that of the other two members of the group.

The phase 2 of the experiment will end when all the chains are either exhausted or interrupted.

PHASE 1 – AGREEMENT

In the phase 1, along with the other two players in your group formed at the beginning of the experiment, you will have to vote to decide which rule to adopt on the management of the common account in case your group comes to the decision moment in the phase 2. However, during phase 1 you will have to vote not knowing which generation (N position in the chain) your group belongs. This information will be provided to you only at the beginning of the phase 2. Therefore in the phase 1 your group will have to agree on a resource management rule before knowing which N position the group belongs within a chain.

Each subject in your group will have to vote for the rule you prefer, choosing between the following two:

Continuation of the chain: each participant of my group should withdraw a maximum of €5 from the common account, ensuring in this way a minimum total saving of €6 that allows the chain to continue

Interruption of the chain: having the possibility to do it each participant of my group should withdraw from the common more than €5, even if that means interrupting the chain

The resource management rule of the common account must be approved unanimously, that is the agreement will be reached only if all the subjects belonging to the same group have voted for the same rule. In the phase 1 you will have 6 rounds to reach unanimity. Groups that do not reach unanimity within the 6 rounds will not be able to access the phase 2 and they will be paid only the show-up fee.

In the phase 2 you will then decide whether to apply the rule chosen in the phase 1, choosing to withdraw an amount compatible with this rule, or withdraw another sum.

SYNTHESIS

PHASE 1 - In the phase 1 you will have to unanimously vote for the rule concerning the management of the common account in the phase 2 without knowing which generation (position) of the chain your group belongs. The agreement on a rule is an essential prerequisite to access the phase 2.

PHASE 2 - In the phase 2 you will know the generation (N) your group belongs in the chain and you will have to decide whether to apply the rule unanimously chosen by the participants of your group in the phase 1 or to withdraw a different amount.

Your final payment will therefore depend on the choices made by you and your group during the phase 1 and the scenario in which you will be during the phase 2 and it will be determined as follows:

- In case your group does not reach unanimity during one of the 6 rounds of the phase 1 you will receive only the €3 of the show-up fee

- In the event that your group enters the phase 2 but you have not made any choices on how much to withdraw from the common because your chain was interrupted before your group entered could take any decision you will receive only the €3 of the show-up fee

- In case that your group enters the phase 2 and you take a decision on how much to withdraw from the common account you will be paid €3 of the show-up fee as + the amount of money that you have decided to withdraw from the common account + [€2 bonus in case yours is the best forecast]

Before starting with the experiment you will be asked to answer some brief control questions.

References

Agius E. (2006), Intergenerational justice, Handbook of intergenerational justice, 53-71, edited by Tremmel J. C., Edward Elgar

Anderson M. W. (2013), Intergenerational Bargains: Negotiating Our Debts to the Past and Our Obligations to the Future, Futures, 54, 43-52.

Arrow, K. J. (1973), Rawls's principle of just saving, The Swedish journal of economics, 75(4), 323-335.

Attas D. (2009), A Transgenerational Difference Principle, Intergenerational justice, 189-218, edited by Axel Gosseries and Lukas H. Meyer, Oxford University Press.

Baggio, M., & Mittone, L. (2018). Grandparents Matter: Perspectives on Intergenerational Altruism and a Pilot Intergenerational Public Good Experiment. Homo Oeconomicus, 1-22.

Bahr, G., & Requate, T. (2007). Intergenerational fairness in a sequential dictator game with social interaction. Kiel University.

Bardsley N. (2008), Dictator game giving: altruism or artefact?, Experimental Economics, 11(2), 122-133.

Barry B. (1977), Justice between generations. In: Hacker P, Raz J (eds) Law, morality, and society. Essays in honour of H.L.A. Hart. Clarendon Press, Oxford.

Beckerman W. (2006), The impossibility of a theory of intergenerational justice, Handbook of intergenerational justice, 53-71, edited by Tremmel J. C., Edward Elgar.

Brandstedt E. (2017), The Savings Problem in the Original Position: Assessing and Revising a Model, Canadian Journal of Philosophy, 47:2-3, 269-289.

Buchanan J. M. (1987), The Constitution of Economic Policy, The American economic review, 77(3), 243-250.

Burke E. (1993), Reflections on the revolution in France, edited with an introduction of Mitchell L. G., Oxford University Press.

Cason T. N. & Mui V. L. (1998), Social influence in the sequential dictator game, Journal of mathematical psychology, 42(2-3), 248-265.

Chaudhuri A., Graziano S. & Maitra P. (2006), Social learning and norms in a public goods experiment with inter-generational advice, The Review of Economic Studies, 73(2), 357-380.

Chermak J. M. & Krause K. (2002), Individual response, information, and intergenerational common pool problems, Journal of Environmental Economics and Management, 43(1), 47-70.

Cherry T. L., Frykblom P., & Shogren J. F. (2002), Hardnose the dictator, The American Economic Review, 92(4), 1218-1221.

Clements P. (2015), Rawlsian Ethics of Climate Change, Critical Criminology 23, pp 461–471, https://doi.org/10.1007/s10612-015-9293-4

Dasgupta P. (1974), On Some Problems Arising from Professor Rawls' Conception of Distributive Justice, Theory and Decision, 4(3), 325-344.

Degli Antoni G., Faillo M., Francés-Gómez P. & Sacconi, L. (2016), Distributive Justice with Production and the Social Contract: An Experimental Study, Econometica Working Papers, N.60 September.

Dierksmeier C. (2006), John Rawls on the rights of future generations, Handbook of intergenerational justice, 72-85, edited by Tremmel J. C., Edward Elgar.

Engel C. (2011), Dictator games: A meta study, Experimental Economics, 14(4), 583-610.

English J. (1977), Justice Between Generations, Philosophical Studies, 31(2), 91-104.

Faillo M., Ottone S. & Sacconi L. (2008), Compliance by Believing: An Experimental Exploration on Social Norms and Impartial Agreements, available at SSRN 1151245.

Faillo M., Ottone S. & Sacconi L. (2014), The Social Contract in the Laboratory: An Experimental Analysis of Self-Enforcing Impartial Agreements. Public Choice, 163(3-4), 225-246.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental economics, 10(2), 171-178.

Fischer, M. E., Irlenbusch, B., & Sadrieh, A. (2004). An intergenerational common pool resource experiment. Journal of environmental economics and management, 48(2), 811-836.

Gabor A. (2013), The Savings Principle and Inter-Generational Justice. European Journal of Science and Theology, 9(2), 299-308.

Gaertner, W., & Schokkaert, E. (2012). Empirical social choice: questionnaire-experimental studies on distributive justice. Cambridge University Press.

Gardiner S. M. (2009), A Contract on Future Generations?, Intergenerational justice, 77-118, edited by Axel Gosseries and Lukas H. Meyer, Oxford University Press.

Gardiner, S. M. (2011a). A perfect moral storm: the ethical tragedy of climate change. Oxford University Press.

Gardiner, S. M. (2011b). Rawls and climate change: does Rawlsian political philosophy pass the global test?. Critical Review of International Social and Political Philosophy, 14(2), 125-151.

Gardiner, S. M., & Weisbach, D. A. (2016). Debating climate ethics. Oxford University Press.

Gaspart F., & Gosseries A. (2007), Are generational savings unjust?, Politics, Philosophy & Economics, 6(2), 193-217.

Gosseries A.(2008), Teorie della giustizia intergenerazionale: Una sinopsi, NOTIZIE DI POLITEIA. - ISSN 1128-2401. - 24:91, pp. 7-26.

Gosseries, A. (2016), La Cuestión Generacional y la Herencia Rawlsiana, Revista Electrónica Instituto de Investigaciones Jurídicas y Sociales AL Gioja, (8), 71-90.

Hardin, G. (1968). The tragedy of the commons. Science 162: 1243–1248.

Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. Nature, 511(7508), 220.

Heyd D. (2009), A Value or an Obligation? Rawls on Justice to Future Generations, 167-188, edited by Axel Gosseries and Lukas H. Meyer, Oxford University Press.

Kahneman D., Knetsch J. L. & Thaler, R. H. (1986), Fairness and the assumptions of economics, Journal of business, S285-S300.

Lamont J. and Favor C., "Distributive Justice", online resource, The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/justice-distributive/>, entered August 2017.

List J. A. (2007), On the interpretation of giving in dictator games, Journal of Political economy, 115(3), 482-493.

Mathis K. (2009), Future Generations in John Rawls' Theory of Justice, Archiv für Rechts-und Sozialphilosphie, 95(1), 49-61.

Meyer L., "Intergenerational Justice", online resource, The Stanford Encyclopedia of Philosophy (Summer 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2016/entries/justice-intergenerational/>, entered August 2017.

Moore, M. A., Boardman, A. E., Vining, A. R., Weimer, D. L., & Greenberg, D. H. (2004). “Just give me a number!” Practical values for the social discount rate. Journal of Policy Analysis and Management, 23(4), 789-812.

Nordhaus, W. D. (1993). Reflections on the economics of climate change. Journal of economic Perspectives, 7(4), 11-25.

Nordhaus, W. D. (2008). A question of balance: Weighing the options on global warming policies. Yale University Press.

Norton B. G. (1989), Intergenerational equity and environmental decisions: A model using Rawls' veil of ignorance, Ecological Economics, 1(2), 137-159.

Ostrom, E., Gardner, R., Walker, J., & Walker, J. (1994). Rules, games, and common-pool resources. University of Michigan Press.

Paden R. (1997), Rawls’s just savings principle and the sense of justice, Social Theory and Practice, 23(1), 27-51

Parfit D. (1984), Reasons and Persons, Oxford University Press.

Partridge E. (2017), Beyond “just savings”, online resource, Environmental ethich and public policy, The online gadley, http://gadfly.igc.org/Unpublished/70/bjs.htm, entered July 2017.

Rawls, J. (1963). The sense of justice. The Philosophical Review, 281-305 Vol. 72, No. 3.

Rawls J. (1977). The basic structure as subject. American Philosophical Quarterly, 14(2), 159-165.

Rawls J. (1999), A Theory of Justice, Revised Edition, Harvard Uiversity Press.

Rawls J. (2001), Justice as fairness: A restatement, Harvard University Press.

Sacconi L. & Faillo, M. (2005), Conformity and Reciprocity in the 'Exclusion Game': An Experimental Investigation, University of Trento Economics working paper, (12).

Sacconi L. & Faillo, M. (2010), Conformity, Reciprocity and the Sense of Justice: How Social Contract-Based Preferences and Beliefs Explain Norm Compliance: the Experimental Evidence, Constitutional Political Economy, 21(2), 171-201.

Sacconi L., Faillo M. & Ottone S. (2011), Contractarian Compliance and the Sense of Justice’: A Behavioral Conformity Model and Its Experimental Support, Analyse & Kritik, 33(1), 273-310.

Schotter A. & Sopher B. (2006), Trust and trustworthiness in games: An experimental study of intergenerational advice, Experimental Economics, 9(2), 123-145.

Schotter A. & Sopher B. (2007), Advice and behavior in intergenerational ultimatum games: An experimental approach, Games and Economic Behavior, 58(2), 365-393.

Solow, R. M. (1974). Intergenerational equity and exhaustible resources. The review of economic studies, 41, 29-45.

Stern, N. (2008). The economics of climate change. American Economic Review, 98(2), 1-37.

Stern, N., Peters, S., Bakhshi, V., Bowen, A., Cameron, C., Catovsky, S., ... & Edmonson, N. (2006). Stern Review: The economics of climate change (Vol. 30, p. 2006). London: HM treasury.

Tammi T. (2011), Contractual Preferences and Moral Biases: Social Identity and Procedural Fairness in the Exclusion Game Experiment, Constitutional Political Economy, 22(4), 373-397.

Thero, D. P. (1995). Rawls and environmental ethics. Environmental Ethics, 17(1), 93-106.

Tremmel J. C. (2009), A theory of intergenerational justice, Earthscan.

Tremmel J. C. (2013), The Convention of Representatives of All Generations Under the 'Veil of Ignorance', Constellations, 20(3), 483-502.

Wall S. (2003), Just savings and the difference principle, Philosophical Studies, 116(1), 79-102.

Wallack, M. (2006), Justice between generations: the limits of procedural justice, Handbook of intergenerational justice, 72-85, edited by Tremmel J. C., Edward Elgar.

Wolf, S., & Dron, C. (2015). Intergenerational sharing of non-renewable resources: An experimental study using Rawls's Veil of Ignorance (No. 01-2015). The Constitutional Economics Network Working Papers.

Wolf, S., & Wagner, K. (2016). If Future Generations Had a Say: Experimental Evidence on Resource Sharing with Veto Power of a Future Generation.

General Conclusions:

Main Results, Limits and Insights for Future Research

The general purpose of my thesis was to model the behaviour of economic agents within three areas – European Union integration, tax compliance and environmental sustainability – using a Rawlsian framework. In particular, the thesis aimed at answering to the following overarching research question: is an impartial and non-binding agreement, conceived in a Rawlsian frame, sufficient to generate fair and stable redistributive institutions?

Before entering the specific research fields, in the Introduction of the dissertation I stated the necessity to contextualize each economic analysis within an ethical frame. In the Appendix to the Introduction I corroborated this thesis showing the hierarchy between ethics and economics established by Adam Smith, largely considered as the founding father of the modern economic science. In particular, according to Smith's all-encompassing project, the figure of the sympathetic but impartial spectator, who represents the ethical touchstone, plays a fundamental role also in regulating the market mechanism. The last one is based not only on mutually advantageous material exchanges, but also on messages of moral approval or disapproval. Said in other words, the so called *homo oeconomicus* has to be always conceived operating within a larger sphere of ethical nature. Thus, the precise relationship between ethics and economics identified by Adam Smith shows how it is fundamental to define a broader context made of ethical connotations before proceeding with an economic inquiry.

Then I moved to set out the specific ethical frame adopted throughout the economic analysis, that is John Rawls's social contract theory. It was shown that in Rawls's ethical system the constitution of a well-ordered society is essentially based on a (hypothetical) agreement between individuals. The device of the veil of ignorance guarantees then a condition of impartiality given that it excludes the availability of any particular information to the parties taking part in the agreement. In this way the moral norms that arise from the agreement and that are supposed to shape the main society's institutions are chosen independently from any subjective and accidental condition. In particular, in the agreement under the veil of ignorance applies the maximin criterion,

so Rawls identifies two specific principles of justice which must be complied in their lexicographic order: 1) to implement a scheme of individual liberties as broad as possible; 2) to redistribute resources in order to maximize the expectations of the most disadvantaged by the natural distribution of social circumstances and individual talents. The second principle of justice was labelled by Rawls himself as difference principle.

I then dedicated the reminder of the Introduction to contextualize the two main areas analysed in depth through the following Chapters: the Rawlsian approach to the international distribution of resources, with a particular focus on the European Union case, and the empirical Rawlsian justice. As for the latter is concerned, I specified the model of social conformist preferences at the basis of the two experimental works

From the Chapter 1 onward I engaged in the analysis of three different economic issues under the assumptions and the conclusions of the Rawlsian theory.

In the Chapter 1 the analysis started with ascertaining the existence of an institutional gap between the European economic integration and the European social integration. Within this general context I applied Rawls's ethical frame to the European Union. The analysis compellingly showed that the constitutional elements which currently characterize the European Union, that is 1) a scheme of mutually advantageous cooperation between its member states and 2) a common basic structure, according to Rawls's domestic theory imply a precise redistributive scheme consistent with the difference principle. In particular, the European difference principle requires that the resources generated by the European economic cooperation are redistributed in order to maximize the expectations of the (group of) European citizen(s) least advantaged. A corollary that followed this conclusion concerned European Fiscal Union, representing a possible way to pool the European surplus and to implement the derived redistributive scheme.

The Chapter 2 was developed together with Luigi Mittone. Within the "slippery slope" framework we adapted the Rawlsian social contract theory and some of its experimental applications to the tax compliance context. In particular, implementing a veil of ignorance in a laboratory experiment our study aimed to analyse in depth the effects over time on tax compliance of a non-binding agreement reached in an anonymity condition. We designed a repeated tax game with three players and four tax regimes. With our experiment we tested two main hypothesis related

to Rawls's theory: 1) behind the veil of ignorance players should vote for the tax regime most consistent with the difference principle and 2) ex-post compliance should be non-declining over time, even in the absence of enforcement mechanisms. The result showed that the veil of ignorance procedure had an important effect on the votes allocation concerning the preferred tax regime and a non-negative effect on compliance. However, the laboratory veil could not generate a stable effect of compliance across time compared to the baseline treatment. Indeed compliance showed to be monotonically decreasing round by round.

Chapter 3 was written with the collaboration of Lorenzo Sacconi and Marco Faillo. It addressed the modern climate change issue within a Rawslian framework. In particular the paper focused on the two main reasons of decades where international negotiations on climate actions failed in being implemented: 1) there is not a broad consensus about the distribution of costs between nations; 2) given the absence of institutions which can monitor and sanction defectors, compliance is fragile. It is within this framework that we implemented Rawls's social contract theory in order to provide useful and innovative insights to deal with climate actions agreements. In particular, we approached the problem by means of a laboratory experiment designed on Rawls' intergenerational justice. In his setting the individuals taking part in the agreement behind the veil of ignorance are only contemporaries, simulating quite well the modern climate change issue. We put groups of experimental subjects on chains of different length, asking them to choose an intergenerational distributive principle before letting them know the position occupied in the chain (veil of ignorance). The results showed that most of the groups agreed on a fair distributive rule and

After highlighting the conclusions reached in the specific areas of inquiry it is possible to provide a common answer to the general research question at the basis of thesis. Again, the research question was stated as it follows: is an impartial and non-binding agreement, conceived in a Rawlsian framework, sufficient to generate fair and stable social and economic institutions? The answer to this question may be introduced quoting Immanuel Kant, who once wrote: "this may be true in theory but it does not apply in practice".

In other words the Rawlsian economic agent works in theory. However its pragmatic observation is less immediate. Indeed, on the one hand, the formal theoretical constructions always result compelling and harmonious, and they are followed by interesting speculative conclusions. This makes possible to provide a positive answer to the research question. Nevertheless, on the other hand, the practical implementation of the procedures aimed to replicate the theory is less

immediate because of many hurdles. This means that according to the empirical tests run in the Chapters 3 and 4 the research question proves to be wrong.

Thus, the general conclusion is that the Rawlsian theory, and in particular the simulation of the veil of ignorance procedure, is not universally applicable, that is it is not validated always and anywhere. Instead, the situations in which it is possible to observe a Rawlsian economic agent seem to be limited. This is also true when compliance is concerned: the model of social conformist preferences apparently is not as general as claimed. This means that in the experimental field, in order to obtain the predicted results, the conditions to implement the procedures have to be carefully tailored.

Nevertheless this general conclusion does not have to be interpreted as a warning message to give up doing empirical research on Rawls's social contract theory. Instead, it is exactly the opposite. It is an incitement to work more in the direction of the experimental research, in order to understand the conditions under which Rawls's social contract theory may produce useful practical results, because "empirical insights are necessary if one wants to apply any theory of justice in the real world"(Gaertner et al. 2012, p.7).

For example, with regards to the European difference principle, a counterfactual empirical analysis might be undertaken to measure the actual amount of the European surplus. Indeed, without an empirical evaluation of a possible European budget the European difference principle keeps being only an elegant theoretical result.

Tax compliance might be approached in a more simplified framework, trying to disentangle whether it is the agreement itself which fosters tax compliance or whether it is the impartiality condition offered by the veil of ignorance that produces a high compliance rate. However, none of the two procedures seems to be able to trigger and to sustain empirical expectations over time. Therefore the psychological mechanism behind the compliance should be analysed in depth also from a dynamic point of view.

In the same way the intergenerational distribution of resources might be studied considering a design which takes into account that generations overlap, so that the least advantageous subject is somehow included in the agreement. Indeed, up to a certain point it is plausible a mechanism of direct reciprocity between generations which might sustain compliance

Many are the possible improvements which can be implemented to understand better the practical implications of the Rawlsian social contract theory. This thesis was another step in that direction.

Appendix to the Introduction: Why According to Adam Smith Ethics has to Play a Major Role in the Economic Theory

Abstract: I sustain that the *Adam Smith Problem* is a fictitious problem. There is not any real inconsistency or discontinuity between Adam Smiths' *Theory of Moral Sentiments* (*TMS*) and his well known *Wealth of Nations* (*WN*). On the contrary, there is a precise relationship between the two books, with the latter being a subset of the former. Focusing on Smiths' words, I show how the ethical system developed in the *TMS* constitutes the (moral) ground on which Smith could conceive his *homo oeconomicus* at the basis of the *WN*. Thus, within the Smithian project, which concerns the theme of human interaction in its different courses, the market mechanism can work on the sole axiom of "self-interest" (*WN*) if, and only if, we preliminarily assume the human capacity of "sympathy" (*TMS*): following this reasoning, I reinforce the interpretation according to which the "invisible hand" of the *WN*, usually meant to regulate mere market exchanges, has to be interpreted from a broader perspective as the "invisible hand" of the "sympathetic but impartial spectator" that disciplines human behaviour on a more general level. I conclude claiming that this precise hierarchy assigned by the father of the economic science to the two mentioned fields implies the necessity of an ethical contextualization in order to have a reliable economic analysis.

JEL Code: B12, B40

Keywords: Adam Smith, Adam Smith Problem, Economics, Ethics

Introduction

The Adam Smith Problem refers to the idea of a substantial discontinuity and of a more general asymmetry between the two books written by Adam Smith, *The Theory of Moral Sentiments* (Smith 1976, since now *TMS*) and the *Wealth of Nations* (Smith 1994, since now *WN*). The debate about the Problem can be traced back to the late 19th century, when some German authors started observing how Adam Smith (apparently) conceived two distinct (apparently irreconcilable) representations of the human nature within the two mentioned works: the *homo oeconomicus* uniquely led by self-interest drafted in the *WN* and the sympathetic individual, spontaneously capable to take into account the others' interests, meticulously described in the *TMS*. From those first objections the literature has constantly contributed to examine the Problem in depth by means of different approaches, either defending its actual existence or denying any apparent inconsistency.

The aim of the present paper is neither to review slavishly the debate occurred since the first formulations of the Adam Smith Problem, nor to reorganize in a systematic paradigm the different attempts which tried to disentangle the claimed inconsistencies. Other authors made a great work in pursuing similar goals (Gocen 2007, Montes 2003, Otteson 2000, Teichgraeber 1981). The aim of the paper is rather to deny the existence of any inconsistency in Adam Smith proposing a compelling interpretation of the exact relationship between the two books (human natures) conceived by Adam Smith. This interpretation mainly relies on the explicit words written by the Scottish author himself. There is room to undertake similar direction since, although the majority of authors agrees on the fact that the Adam Smith Problem is a fictitious problem, "there is still no widely agreed version of what it is that links [those] two texts, aside from their common author" (Wilson & Dixon 2006, p. 251). Thus it is not enough to reach the conclusion that Adam Smith was not inconsistent. In order to make the analysis complete it is also necessary to demonstrate, in a compelling way, in which terms Smith was consistent, understanding his general project: that is the main aim of the paper.

In particular I will sustain that once we acknowledge how in putting forward the *WN*'s *homo oeconomicus* Adam Smith implicitly took for granted the thesis and the conclusion stated in the *TMS*, it becomes natural to interpret the figure of the sympathetic but impartial spectator (developed within his ethical reflection) as the (moral) regulating mechanism for the market exchanges, the latter providing the basis of the economic development of nations. Through a detailed analysis of the two textbooks it will be possible to conclude how the *TMS* substantially constituted the ethical

foundation on which the Scottish author was able to shape his meagre (in terms of human connotations) economic agent described in the *WN*. This specific conclusions will be then extended sustaining, as Sen did (1987), that the development of an ethical ground becomes an essential step in developing a plausible economic analysis.

Starting, for expository convenience, with a brief review of the *WN*, Section 1 will highlight how in Smith's economic analysis the division of labour, together with the existence of a market system, constitute the propelling engine of the economic development. The motive to the division of labour is then identified by Smith in what can be considered an axiom, since the Scottish author does not linger to explain its anthropological origins: the self-interest innate in every human being (economic agent).

Section 2 will outline the dense analysis concerning the moral principles of sympathy and approval contained in the *TMS*. According to Smith the sympathetic but impartial spectator becomes the prescriptive mechanism which regulates human interactions occurring in society: assuming his perspective everybody can take into consideration the others' interests in the most equitable way.

Section 3 has the aim to recall the Adam Smith Problem which might emerge reading in an approximate manner the two above mentioned books. The Problem was approached from different perspectives, but in general it concerns the idea of an irreconcilable inconsistency between the two distinct representations of the human being given by Smith in the two texts. On one side there is the dispassionate economic agent moved by the sole self-interest who merely weigh material costs and benefits. On the other there is the individual endowed with sympathetic capacities indirectly expressing other-regarding preferences. Within the present paper the (fictitious) Problem will be overcome identifying the precise relationship which ties the *TSM* and the *WN*. It will be shown how the self-interest axiom is not sufficient in itself for the market mechanism to work and to produce public virtues from private vices. Indeed Smith adopts that postulate only as a complementary assumption of his moral theory previously developed in the *TMS*. In particular, the market institution that, together with the division of labour (the last one motivated by the self-interest), leads to the economic development of nations is founded by Adam Smith on the ethical system based on the sympathetic but impartial observer. That ideal figure, by means of his "invisible hand", regulates and coordinates personal interests of all economic agents in such way that positive benefits are produced for the whole community (nation).

In the light of similar conclusions and from a broader perspective it is possible to interpret the *TMS* as the ethical background of the Smithian economic theory. The last one is not self-sufficient and cannot operate in autonomy without any further (moral) assumption. Thus Section 4 will

generalize the obtained result concerning the existing relationship between Adam Smith's *WN* and *TMS*. In particular, it will be corroborated the thesis that ethics and economics constitute an inseparable binomial and are to be in a precise hierarchy: before carrying out any economic analysis it is recommended a broader contextualization in ethical terms. This is a way to restate a conclusion reached by the Nobel prize Amartya Sen (with a less analytical approach) 30 years ago.

A.1 The *Wealth of Nations*

In his most famous work, the *Wealth of Nations*, Adam Smith (1994) proposes a systematic analysis of the main economic phenomena which characterize a modern economy. Many of those topics are still nowadays the subject of intense debate and extensive research[1]. Those themes were not unknown when Smith investigated them; however, Smith was the first author who attempted to give to those economic forces a coherent representation and who tried to link them in a unitary system. In his *WN* Smith "took the inchoate economic literature generated between 1650 and 1750 and fashioned it into an intellectual discipline he called *political economy*" (Landreth & Colander 2002, p. 16). In other words Smith was the first who "had been able to integrate into a single volume an overall vision of the forces determining the wealth of nations" (Landreth & Colander 2002, p. 82).

The literature is therefore unanimous in recognizing how Smith's contribution to the (foundation of the) economic science in terms of original ideas is much less important then the successful attempt of codifying in a systematic representation the state of the economic knowledge of his time. Of course it would be improper to deny the existence of new notions within the Smithian economic analysis, however it is essential to remark how "his role was [mostly] to take up the best ideas of other men and meld them [...] into a comprehensive system" (Landreth & Colander 2002, p. 88) rather than having developed completely new concepts. Thus, the main contribution of Adam Smith's *WN* was to have introduced a new (scientific) method within the economic analysis: for this reason he is usually defined as the funding father of the economic science.

In his inquiry into the nature and the causes of the wealth of nations Smith lingers in an analytic and systematic manner on the cause-effect relationships which occur between the different

1 The book deals with topics like comparative advantage, free market institution, protectionism, perfect competition, monopoly, money, taxation. At the same time it is possible to find out traces of an infinite variety of themes nowadays assumed as obvious, going from inflation to capital markets, from opportunity-cost analysis to credit risk, from collusions to lotteries.

economic courses. With regard to the specific contents of the book, as mentioned in the note number 1, they encompass a countless number of thematic areas; however, this does not mean that within the Smithian economic project there are not a starting point, a common thread and some clear conclusions: it has to be remarked the scientific method of his economic inquiry. In particular the Scottish economist starts his reasoning dealing with the division of labour, which is assumed to be the engine of the economic development, since "so far as it can be introduced, occasions, in every art, a proportionable increase of the productive powers of labour" (Smith 1994, p. 5)[2].

However the increased productivity of individuals is necessary but not sufficient to produce general wealth. In order to achieve a general progress by means of the division of labour it is essential the existence of a coordinating mechanism which can "recompose" what through the division of labour "is divided": this is the market exchange system, which allows every individual to benefit of the higher productivity accomplished by everybody else. Thus

> "[w]hen the division of labour has been once thoroughly established, it is but a very small part of a man's wants which the produce of his own labour can supply. He supplies the far greater part of them by exchanging that surplus part of the produce of his own labour" (Smith 1994, p. 24).

Therefore a market mechanism, where it is possible to exchange productivities yet before products, is the symbiotic element which has to develop simultaneously to the progress of the division of labour if we want to achieve, according to Smith, an increase in general wealth. About these two categories of interpretation, Adam Smith is straightforward in claiming how the measure in which the division of labour is feasible essentially depends on the extent of the available market (Smith 1994, pp. 19-23): it follows an unavoidable relationship between the two elements at the basis of economic development. Smith moves then to highlight the inconvenience in exchanging every product(ivity) with every other product(ivity), and how this obstacle was overcome by the introduction of money as universal mean of exchange. However, deepening similar technical analysis becomes superfluous for the purposes of the present paper.

Instead it is relevant to linger on the binomial constituted by division of labour and market mechanism. In particular it is fundamental to understand which is the spark that turns on what in Adam Smith's opinion is the engine of the economic development, or said otherwise, which is the motive that pushes individuals to undertake the division of labour. Smith clarifies how the latter is

2 Smith attributes to three specific circumstances the higher productivity achieved through the division of labour: increase of dexterity, saving of time and invention of machines (Smith 1994, pp. 7-11)

neither the outcome of a structured agreement between people nor a sober reflection on the possible net benefits deriving from exchange; instead, Smith explicitly identifies the origin of the division of labour in what in his opinion is a peculiar feature which distinguish human beings, viz. "the propensity to truck, barter, and exchange one thing for another" (Smith 1994, p. 14), up to the point that "[e]very man [...] lives by exchanging, or becomes, in some measure, a merchant" (Smith 1994, p. 24).

But the quoted words, if considered in isolation, induce an interpretative difficulty since they generate a cycle of vicious nature about the relationship between division of labour and development of a market. Those words do not say much if they are not jointly read and integrated with those of an excerpt become famous within the economic discipline: in Smiths' opinion every human being

> "has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favour, and shew them that it is for their own advantage to do for him what he requires of them. Whoever offers to another a bargain of any kind, proposes to do this. Give me that which I want, and you shall have this which you want, is the meaning of every such offer; and it is in this manner that we obtain from one another the far greater part of those good offices which we stand in need of. It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity, but to their self-love, and never talk to them of our own necessities, but of their advantages. Nobody but a beggar chooses to depend chiefly upon the benevolence of his fellow-citizens (*WN* Smith 1994, p. 15)"[3].

In other words it is the propensity to truck and to barter which gives origin to the division of labour and the parallel development of a market, since without that inclination neither the former nor the latter would take place; however, similar propensity is not else then a derivative, a phenomenal display of a more primitive but existing selfishness characterizing the human being: only the self-interest induces the economic agent to undertake the exchange action, or equivalently, the division of labour[4]. And given that it is the personal interest to move the economic subject towards the specialization of labour, a question follows immediately: how can a multitude of

3 It is from those words that it is usually inferred that Adam Smith would have designed the prototype of the cold *homo oeconomicus*, solely focused on comparing own and other's people material costs and benefits

egoistic interests coexist and to be of use to a society? How is possible that there are growth and development there where everybody undertakes only those actions which are the reflection of their personal interests? It is exactly at this point that intervenes the second component concerning the economic progress developed within the Smithian theory: the market exchange mechanism.

The market institution is generally illustrated by Adam Smith as that natural mechanism which ensures that the division of labour can be realized and can find its (economic) convenience; yet, the market is designed as that spontaneous process of co-ordination which, to the extent it is left to act freely, leads to general positive outcomes that are not explicitly taken in consideration by the self-interest of economic agents. The latter idea can be made clearer through the words of Adam Smith himself, quoting another paragraph which became famous in the economic textbooks. With regard to the economic subject moved solely by his own self-interest, Smith writes that, within the market mechanism, he is "led by an invisible hand to promote an end which was no part of his intention.[...] By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it" (Smith 1994, p. 485)[5]

In this perspective the market is designed as that institution which, without any artificial constrain, can govern in a coordinated manner the innumerable individual interests, and that, by means of the invisible hand can lead to achieve an unintentional aim, that is the welfare of the whole society. But at this point Smith's reasoning can be interpreted also from another perspective, the most important in the present paper: it is not necessary to worry if the economic subjects are not motivated by regards which go beyond their strict personal interests (like for example altruism or reciprocity) or by evaluations that concern the well-functioning of the general economic system (like for example the total welfare); even though we assume that individuals are moved by the narrowest self-interest, the market mechanism is capable to coordinate conveniently those (egoistic) human actions in order to produce positive outcomes for the whole society.

4 "Man's propensity for exchange, which makes the division of labour and therefore specialization possible, and which according to Smith is the basis of the wealth of nations, is not based on altruism. Nobody, says Smith, makes an exchange for the sole purpose of making happy his neighbour" (Rommel & Winter 2001, p. 69, *my own translation*). Yet, "self-love is the definite source in the *Wealth of Nations* of exchange activity. And this is vital" (Macfie 1959, p. 227).

5 A few pages earlier, Smith anticipates the same concept with a more technical language: "[e]very individual is continually exerting himself to find out the most advantageous employment for whatever capital he can command. It is his own advantage, indeed, and not that of the society, which he has in view. But the study of his own advantage naturally, or rather necessarily, leads him to prefer that employment which is most advantageous to the society" (Smith 1994, p. 482).

And it is probably in reason of what just said that Adam Smith, in the initial part of his *WN*, does not linger too much on inquiring or deepening in detail the nature of those self-interested motives which are at the basis of the human (economic) behaviour, but he considers the egoism more like an axiom: the later claims on the physiognomy and the physiology of the market mechanism allow to minimize the earlier anthropological assumptions at the basis of the economic agent. In favour of this second interpretation it is worth to spend some more words with the purpose to express a reasoning that, although developed from another point of view, remains coherent with everything written so far.

Substantially, if from one side Smith limits himself in assuming as foundation of the division of labour, and consequently of the economic growth, the self interest as exclusive motive of the economic action, from the other side he can do that solely because similar postulate is legitimated by the further assumptions concerning the extraordinary coordination capacities of the market institution. Yet, outlined a certain configuration to the market mechanism, similar configuration subsequently allows the possibility to minimize the anthropological assumptions, taking the self-interest as unique motive. It is then a different issue to understand which specific features attributed to the market institution allow to validate the just suggested logic. Smith does not describe in the *WN* the behavioural assumptions at the basis of the market mechanism; to identify those elements it is necessary to look at his other book, the *TMS*.

A.2 *The Theory of Moral Sentiments*

Anticipating some of the later conclusions, it is possible to sustain how for Adam Smith the market mechanism is not infinitely virtuous in bestowing its positive benefits. More precisely the limit of the individual economic (self-)interests must coincide with the wider plot of a moral social conduct (approval of the sympathetic but impartial spectator): only under similar assumption motives of individualistic nature can generate, even involuntary, positive effects in favour of the community; only to the extent that every subject is capable to take in consideration (also indirectly) some meta-personal interests, it is possible a general progress for the human society under the market institution. And which are those social boundaries that are supposed to limit the pursue of

personal interests, so that the positive effects generated by the market are not offset by some other negative (social) consequences[6], is clearly remarked by Smith himself in his *TMS* (Smith 1976).

In the *TMS*, which deals in a broader perspective with human interaction (not limited to the economic exchange sphere), Smith provides a deep analysis of human moral action. And it is thanks to the *TMS* that it is possible to realize how Smith's real research had a wider perspective than inquiring the mere *homo oeconomicus*: his research deals more generally with the relationship of the single individual with the society; the Smithian research can therefore be defined as a research about the human being in its intersubjective dimension, with the economic environment representing only a particular case, a subset of all human interactions[7].

As for the *WN*, also in the *TMS* Smith sets a starting point around which to develop his system of moral regulation. In particular, the whole Smithian ethical analysis is founded on the concept of sympathy, meant as a (moral) capacity that allows individuals to take part, more or less intensely, in the feelings of other human being[8]; the sympathy itself then "does not arise so much from the view of the passion, as from that of the situation which excites it" (Smith 1976, p. 12). Indeed the Scottish author specifies "[a]s we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation" (Smith 1976, p. 9): in this way imagination, understood as that capacity to identify ourselves with a certain third-party circumstance, assumes a central role within the Smithian moral system. Similar capacity is not neutral, that is, it does not consist in a pointless representation of the considered situation; on the contrary, it allows the observer to "relive" the observed state of affairs and thus to sympathize with the passions of the observed individual.

Thus Smith clarifies what is the purpose of the whole process unfolded through imagination and sympathy:

> "[w]hen the original passions of the person principally concerned are in perfect concord with the sympathetic emotions of the spectator, they necessarily appear to this last just and proper, and suitable to their objects; and, on the contrary, when,

6 In the *TMS* Smith "tried to establish the appropriate institutional framework so that particular [self-]interests could be expressed without harming [others] individuals" (Pena López & Sánchez Santos 2007, p. 75).

7 Sen's words are really clarifying with regard to that point: "Smith was concerned not only with the sufficiency of self-interest at the moment of exchange but also with the wider moral motivations and institutions required to support economic activity in general" (Sen 2010, p. 50).

8 Smith opens the *TMS* with the following claims: "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it". (Smith 1976, p.9)

> upon bringing the case home to himself, he finds that they do not coincide with what he feels, they necessarily appear to him unjust and improper, and unsuitable to the causes which excite them" (Smith 1976, p. 16).

Therefore the aim of the sympathetic process is to express a judgement in terms of moral appropriateness or inappropriateness of the affections of other subjects with regard to the situation that generated them; appropriateness established according to the harmony or dissonance of spectator's moral feelings with those of the observed individuals.

In short, an observer, by means of sympathy and following the imaginary exchange of circumstances, will judge whether or not the passions of another individual are appropriate with regard to the situation that generated them according to a correspondence or a discrepancy with his own feelings. And since also the observed subject is capable of sympathy, being able to grasp the approval or disapproval of his own feelings by the spectator, but above all since everyone is, according to Smith, in search of approval[9], the observed individual will be induced to "level" his feelings, and then his behaviour, up to that point in which the latter are considered appropriate and are approved by the spectator. That is the empirical balance mechanism proposed by Adam Smith within his ethical system.

However, similar system of moral evaluation, if limited to what was said so far, would easily leave room to the objection of ethical relativism[10] since it does not conceive any objective prescriptive mechanism. To avoid similar critique, and in completing his moral system, the Scottish author adopts the expedient of the ideal sympathetic (but impartial) observer: all "passions of human nature" writes Smith, "seem proper and are approved of, when the heart of every impartial spectator entirely sympathizes with them, when every indifferent by-stander entirely enters into, and goes along with them" (Smith 1976, p. 69). Thus,

> "though man has, in this manner, been rendered the immediate judge of mankind, he has been rendered so only in the first instance; and an appeal lies from his sentence to a much higher tribunal, to the tribunal of the [...] impartial and well-informed spectator" (Smith 1976, p. 130).

9 "[N]othing pleases us more than to observe in other men a fellow-feeling [approval] with all the emotions of our own breast; nor are we ever so much shocked as by the appearance of the contrary [disapproval]" (Smith, 1976, p. 13).

10 It is almost superfluous to highlight how different spectators might express different degrees of approval with regard to the same circumstances.

Through the figure of the impartial spectator Adam Smith wants to state firmly the importance of the depart of the observer from all those personal peculiarities which could influence the approval judgement. In fact, the comparison with the feelings of an impartial and well-informed observer allows to discriminate objectively a correct (approved) behaviour from an incorrect one to the extent which the feelings of the subject under judgement are aligned with those of the impartial spectator: in this way the sympathy becomes a social bond (Macfie 1959, p. 212) and the sympathetic but impartial observer becomes a normative figure for the moral equilibrium within a society. Equilibrium which is required in every situation of social interaction, included the market exchange environment.

The system of the impartial but sympathetic spectator leads then to two important corollaries: first of all, the continuous exercise of evaluation of human conduct through the eyes of the impartial spectator induces the society to derive some general rules regarding what is appropriate to do or to avoid in some specific circumstance. In this way Smith tries to justify the so-called social norms: we are not supposed to enter the impartial observer perspective in every single moment or circumstance since "habit and experience have taught us to do this so easily and so readily" (Smith 1976, pp. 135-136 and for a more complete dissertation of the theme see Smith 1976 pp. 156-170); secondly, it enables the single individual to be an objective judge of himself, since it will be sufficient for him to refer to the impartial observer's sentiments to evaluate his own behaviour. But once again those technicalities go beyond the aim of the paper.

A.3 *TMS* and *WN*: the *Adam Smith Problem* and its solution

What is relevant at this point is to focus on how the ideas belonging to the two books (moral and economic sphere) are supposed to be read within a single framework. After having sketched both, the economic system relying on selfishness and division of labour proposed in the *WN*, and the moral apparatus based on the figure of the sympathetic but impartial spectator described in the *TMS*, it is necessary to identify a clear connection between these two discourses in which human interaction is the protagonist. And what might seem a mere academic exercise is in fact the result of a heated controversy over what was supposed to be an inconsistency of premises, hence also of the conclusions, developed by Adam Smith in his two main works, *WN* and *TMS*.

With "Adam Smith Problem" the literature labelled that debate concerning the (apparently) contradicting and incompatible anthropological views which emerge within the two Adam Smith's

texts. As stated in the introduction, it is not a purpose of the present paper to sum up the different positions, which dealt with the Problem through various methodologies: other contributions can serve better similar purpose (Gocen 2007, Montes 2003, Otteson 2000, Teichgraeber 1981). In general however, the debate concerned the existence of a dualistic human nature such as represented by Smith; dualism reflected in the two works according to the following scheme: an empirical (phenomenal) agent lead by mere self-interest placed at the foundation of the *WN*, which implies to take into account of a second subject only to the extent he is necessary for immediate personal aims; an ideal (noumenic) actor endowed with sympathetic capacities playing a central role in the *TMS*, which in turn requires to take into consideration another subject according to a broader perspective than treating him as a pure mean.

While it is not the main purpose of this paper to retrace the debate about the Adam Smith Problem, it is important to ascertain the point which the controversy has arrived to. Pack (1997) shows, by means of the reading of a third Smith's work[11], how the *WN* and the *TMS* are part of a broader common system. Evensky states how "the confusion lies not in the pen of Adam Smith, but in the eyes of those who profess to see an Adam Smith Problem" (Evensky 1987, p. 464); Raphael and Macfie quickly get rid of the Adam Smith Problem defining it as a "pseudo-problem based on ignorance and misunderstandings" (Smith 1976, p. 20); Montes makes a further clarification: despite the Adam Smith Problem is fictitious, it "had not yet been fully exhausted [because a] second stage in the debate [requires] also defending the consistency position" (Montes 2003, p. 79). Boff remarks the same point: "even if one accepts just one picture of man in both the TMS and the WN [...] there is still room in our analyses that invites a questioning as concerns the relation between both books" (Boff 2014, p.7). Wilson and Dixon are the most exhaustive in pointing out the status quo of the research:

> "[t]he old *Das Adam Smith Problem* is no longer tenable. Few today believe that Smith postulates two contradictory principles of human action: one in the *Wealth of Nations* and another in the *Theory of Moral Sentiments*. Nevertheless, an Adam Smith problem of sorts endures: there is still no widely agreed version of what it is that links these two texts, aside from their common author; no widely agreed version of how, if at all, Smith's postulation of self-interest as the organising principle of economic activity fits in with his wider moral-ethical concerns" concerning the sympathetic principle (Wilson & Dixon 2006, p. 251).

11 A collection of notes called *Lectures on Jurispiudence*.

The general conclusion is that a real "Problem" does not exist, that is, if in the representations of human nature proposed by Smith there is an inconsistency, this cannot be attributed to a deliberate lack of the author himself[12]. However, even though it does not exist any real inconsistency, there is still room to identify the most compelling way to unify the ideas that emerge in the two texts (once assumed that those do not represent two incompatible categories).

In this paper I claim that the best way to aggregate Smith's *TMS* and *WN* in a coherent framework is indirectly suggested by Montes (2003) when he reports the reactions of some authors to the first formulations of the Adam Smith Problem (Montes 2003, pp. 73-77). Summarizing here in a single concept those different positions, I sustain that the two systems, respectively based on self-interest and sympathy, are fully complementary, and therefore compatible, to the extent we assume the latter (modality) as the regulating principle of the former (motive): self-interest and sympathy are not exclusive (Witztum 1998); it is the capacity of sympathy (the approval of the impartial observer) that circumscribes the exact limit of individual self-interests in general and within the market exchange system in particular. Montes (2003) himself does not agree with this interpretations since "narrows Smith's concept of sympathy" (Montes 2003, p. 85), not considering the latter as a possible motive of action too; however it will be shown how the reading provided throughout the following analysis enlarges the role of sympathy instead of narrowing it.

Thus, the main thesis of the paper is that the "invisible hand" that coordinates the personal interests of economic agents within a market economy is to be essentially interpreted as the invisible hand of the impartial spectator, because "the economic man also is under the sway of social sympathy and the impartial rulings of the informed spectator" (Macfie 1959, p. 223). The impartial spectator, thanks to his sympathetic sentiment, balances the selfishness of the various agents acting within the market, producing positive benefits for the society as whole: as Smith himself clearly writes, "by acting according to the dictates of our moral faculties [through the perspective of the sympathetic but impartial spectator], we necessarily pursue the most effectual means for promoting the happiness of mankind" (Smith 1976, p. 166). Similar interpretation on the

12 If we accept the hypothesis of the existence of a Problem we are required to explain Smith's intention to ignore the conclusions reached in the *TMS* when writing the *WN*, and vice versa: indeed, if we take into account that Smith published six editions of the *TMS*, two of them after (1781 and 1790) the first publication of the *WN* (1776, and then four additional editions), with no significant structural changes introduced in the *TMS*, we can immediately conclude how the author did not write the two books in air locked rooms. From this perspective it should not surprise that Smith did not introduce any big revision of the two book from one edition to the other (Viner 1927, p. 201 and p. 217).

relationship between the impartial observer and the market institution, that is between the *TMS* and the *WN*, is supported by the intuition of other authors[13]:

> "from this point of view, the 'invisible hand theorem' becomes much more interesting than traditional microeconomics manuals suggest. In the market, economic agents do not exchange only goods, but also messages of approval or disapproval, so that individuals, while behaving in a self-interested way, tend to do so respecting the legitimate expectations of others [...] In this way opportunism is kept at bay and the cooperative behaviour is stimulated. In short, the invisible hand that contributes to the construction of the good of all seems to be that of the impartial spectator" (Screpanti & Zamagni 2004, p. 111, *my own translation*).

Therefore the interaction between Smith's two works does not give rise to a conflict, but rather to a coherent, broader system that assumes the perspective of the impartial observer at the basis of the market institution. Again, the thesis sustained in this paper is that the *TMS* theory constitutes a (moral) background on which Smith promoted his own economic reflections of *WN*[14]: "the *Wealth of Nations* is simply a special case - the economic case - of the philosophy implicit in the *Moral Sentiments*" (macfie 1959, p. 223). However it is important to remark how Screpanti and Zamagni did not provide a real justification on what drives their conclusions. Therefore, in order to reinforce the just proposed interpretation and to understand more exactly how, according to Smith himself, the moral sphere plays a major role with regard to the economic theory, it is useful to quote an extended paragraph taken from the *TMS* and which sustains the proposed interpretation:

> "[t]here can be no proper motive for hurting our neighbour, there can be no incitement to do evil to another, which mankind will go along with, except just indignation for evil which that other has done to us. To disturb his happiness merely because it stands in the way of our own, to take from him what is of real use to him merely because it may be of equal or of more use to us, or to indulge, in this manner, at the expence of other people, the natural preference which every

13 Another similar interpretation is provided by Pena López and Sánchez Santos: "the well-known metaphor of the "hand invisible" of the WN also appears in the TMS (IV, 2) alludes to an involuntary coordination of interests. That is, the self-regulation of the moral system plays a direct role either in the TMS or in the WN, because the imaginary [spectator] that sustains the liberal society supposes not only the coordination of the individual interests but also the of individuals as *homines ethici* or members of a social group. (Pena López & Sánchez Santos 2007, p. 83, *own translation)*.

14 As recalled in the note 11, the *TMS* was written before (1759) than the *WN* (1776).

> man has for his own happiness above that of other people, is what no impartial spectator can go along with. Every man is, no doubt, by nature, first and principally recommended to his own care; and as he is fitter to take care of himself than of any other person, it is fit and right that it should be so. Every man, therefore, is much more deeply interested in whatever immediately concerns himself, than in what concerns any other man: and to hear, perhaps, of the death of another person, with whom we have no particular connexion, will give us less concern, will spoil our stomach, or break our rest much less than a very insignificant disaster which has befallen ourselves. But though the ruin of our neighbour may affect us much less than a very small misfortune of our own, we must not ruin him to prevent that small misfortune, not even to prevent our own ruin. We must, here, as in all other cases, view ourselves not so much according to that light in which we may naturally appear to ourselves, as according to that in which we naturally appear to others. Though every man may, according to the proverb, be the whole world to himself, to the rest of mankind he is a most insignificant part of it. Though his own happiness may be of more importance to him than that of all the world besides, to every other person it is of no more consequence than that of any other man. Though it may be true, therefore, that every individual, in his own breast, naturally prefers himself to all mankind, yet he dares not look mankind in the face, and avow that he acts according to this principle. He feels that in this preference they can never go along with him, and that how natural soever it may be to him, it must always appear excessive and extravagant to them. When he views himself in the light in which he is conscious that others will view him, he sees that to them he is but one of the multitude in no respect better than any other in it. If he would act so as that the impartial spectator may enter into the principles of his conduct, which is what of all things he has the greatest desire to do, he must, upon this, as upon all other occasions, humble the arrogance of his self-love, and bring it down to something which other men can go along with. They will indulge it so far as to allow him to be more anxious about, and to pursue with more earnest assiduity, his own happiness than that of any other person. Thus far, whenever they place themselves in his situation, they will readily go along with him" (Smith 1976, pp. 82-83).

And Smith concludes his reasoning with the following words, which make extremely clear how to regulate the market (competition), that is to coordinate the countless self-interests of

economic agents acting within a market, it is not possible to rely on narrow behavioural premises, but only on a moral system embodied in the impartial spectator:

> [i]n the race for wealth, and honours, and preferments, he may run as hard as he can, and strain every nerve and every muscle, in order to outstrip all his competitors. But if he should justle, or throw down any of them, the indulgence of the spectators is entirely at an end. It is a violation of fair play, which they cannot admit of. This man is to them, in every respect, as good as he: they do not enter into that self-love by which he prefers himself so much to this other, and cannot go along with the motive from which he hurt him. They readily, therefore, sympathize with the natural resentment of the injured, and the offender becomes the object of their hatred and indignation. He is sensible that he becomes so, and feels that those sentiments are ready to burst out from all sides against him" (Smith 1976, p. 83).

From those words it is possible to understand how for Smith it is legitimate and natural that every human being, according to his own perspective, primarily prefers himself to anybody else. However, to make morally licit the pursuit of his own (economic) interests he must take in due consideration the interests of others. And the way in which every (economic) agent is supposed to take into consideration the claims of other subjects is specified by Smith himself in a subsequent passage (which partly recalls what was quoted above):

> "to the selfish and original passions of human nature, the loss or gain of a very small interest of our own, appears to be of vastly more importance, excites a much more passionate joy or sorrow, a much more ardent desire or aversion, than the greatest concern of another with whom we have no particular connexion. *His interests, as long as they are surveyed from this station, can never be put into the balance with our own, can never restrain us from doing whatever may tend to promote our own, how ruinous soever to him. Before we can make any proper comparison of those opposite interests, we must change our position. We must view them, neither from our own place nor yet from his, neither with our own eyes nor yet with his, but from the place and with the eyes of a third person, who has no particular connexion with either, and who judges with impartiality between us*" (Smith 1976, pp. 135-136, *italics added*).

Therefore, assuming the point of view of the impartial spectator becomes the perspective which leads to a (more general) ethical equilibrium in human relationships, but also to an economic equilibrium within the exchange market: in fact similar regulator mechanism allows the single agent to take into account, in an appropriate manner, the self-interest of those who he interacts with (also within the market), making in similar way his own selfishness (morally) licit. The self-interest assumed by Smith as the first principle of economic development is then limited up to that precise point in which it is approved by the impartial spectator: the "invisible hand" of the latter purifies the personal interests of economic agents from those components that could be detrimental to the interests of society in general.

In other terms, in the market system "[t]he impartial spectator censures our [excessive] selfish impulses and restores things more nearly to their correct proportions" (Cam 2008, p. 108), or said through the words of Smith himself "the natural misrepresentations of self-love [is] corrected only by the [hand] of this impartial spectator" (Smith 1976, p. 137). Since, according to this interpretation, market exchanges are directly regulated by the impartial spectator, it becomes also clear why for Smith motives of altruistic nature are not necessary at all for the good functioning of the economic system and for the wealth of nations, and beneficence becomes an "ornament which embellishes, not the foundation which supports the building" of the market economy (Smith 1976, p. 86).

A.4 In Smith's perspective: ethics as foundation of the economic theory

The specific conclusion related to the (non-existent) Adam Smith Problem allow to enlarge the considerations regarding the Smithian ethical-economic system. Before to proceed, I wish to quote once again the words of Screpanti and Zamagni, which effectively summarize what has been hitherto claimed: "the incriminating passage[15] of the *WN* presupposes in its enunciation the thesis of the *TMS*, and in particular those related to the existence of a system of 'norms of civic and economic morality' based on *sympathy*. This system of rules guarantees the orderly functioning of the market without individuals having to resort to violence and coercion to force the parties to respect 'the rules of the game'. So the aforementioned maxim simply says that a market economy could function *even if* the *additional* motivations of all the individuals belonging to it were exclusively of self-interested nature" (Screpanti & Zamagni 2004, pp 114-115, *my own translation*).

15 See footnote number 2.

That is, within the Adam Smith's project the ethical background plays a major role in the construction of his economic system based on the market institution; again, his ethical theory constitutes an indispensable prerequisite for a correct reading, interpretation and functioning of his economic system; the *TMS* constitutes the foundation of the *WN* (Macfie 1959). However, what is even more important is the possibility of extending those specific conclusions and the "Smithian logic" on a more general level: an ethical contextualization becomes essential before making assumptions within the economic analysis. This was the path followed by that author who is considered the father of the economic discipline. Smith developed a marvellous system that requires ethics to play a major role within his economic theory, and there are no reasons to deviate from similar practice. Ethical considerations have to be at the basis of every economic theory.

Similar intuition is in line with an idea expressed by Amartya Sen 30 years ago, who denounced the "impoverishment of welfare economics as a result of the distance that has grown between ethics and economics" (Sen 1987, p. 51). Sen was the first personality of international relevance to reaffirm strongly the limited view of the modern economic theory; limitation derived, in his opinion, from having constantly subtracted the study of economics to a comparison with considerations of ethical nature. Sen was convinced that an almost unanimous interpretation of Adam Smith achieved reading in isolation his economic theory (*WN*), implied that reflections of ethical importance, far from being ignored by Smith (*TMS*), lost their natural place within the standard economic theory.

In other terms, according to Sen, having ignored the moral reflection made by Adam Smith, and more generally, having misunderstood the importance of ethical considerations within the economic field, led to a depletion of any result obtained within the economic theory: Sen comes to similar conclusion showing how Adam Smith was consciously concerned with delineating in detail an ethical premise (*TMS*) before carrying out his pure economic analysis (*WN*). With the following words Amartya Sen's thought can be summarized: "it is precisely the narrowing of the broad Smithian view of human beings, in modern economies, that can he seen as one of the major deficiencies of contemporary economic theory" (Sen 1987, p. 28). In conclusion, ethical premises must be taken in account to produce reliable economic theories.

Conclusions

The primary aim of this paper was to highlight the necessity of lingering on ethical considerations before moving to develop any economic theory. The thesis that an economic analysis requires some precise ethical premises was defended through a particular methodology: in the paper I showed which hierarchy Adam Smith, considered the father of the economic science, clearly established between ethics and economics. In particular it was pointed out how Smith assumed, in the elaboration of the *WN*, the thesis developed in the *TMS*: the figure of the sympathetic but impartial spectator, put at the centre of his ethical system, plays a fundamental role also in regulating the market mechanism and the division of labour (which solely relies on self-interest). In other words, market exchanges are only a particular subset of all the possible human interactions.

The reason of lingering on the Adam Smith Problem to restate the importance of an ethical perspective within the economic theory is that the Problem "entails the relationship between individual and society and, more specifically, the interdependence of ethics and economics." (Montes 2003, p.82). The single model proposed to reconcile the two discourses (ethics and economics) in Adam Smith is to consider the *WN* as a strict subset of the *TMS*. In the Smithian system mutual economic benefits are possible only where there is mutual sympathy, and the market institution is only one branch of the society where the sympathy and the prescriptions of the impartial spectator take place. Since Adam Smith chose to proceed in that way, it is important to recognize the necessity of an ethical reflection before moving to pure economic considerations: premises of moral nature are not only ancillary, but rather necessary before engaging in an economic analysis.

However the analysis provided in this paper does not exhaust a further issue: "Smith did not refer even once to the TMS in the WN" (Wagner-Tsukamoto 2013, p .77) so "why should we think that economic behavior is rightfully subject to the impartial spectator's judgment if there is no mention of any such thing in WN?" (Otteson 2000, p.66). Given the interpretation provided in the present paper (the *WN* as subset of the *TMS*), it becomes important to provide a reliable justification to the decision of Smith to not mention explicitly similar relationship. Here there is no space to inquire that choice. A reliable answer probably requires to include in the analysis other Smith's works, beyond the *WN* and the *TMS*. What remains clear anyway is that Adam Smith's economic theory cannot work without the assumptions made throughout his ethical analysis. This conclusion is broadened and assumed as a method valid for encompassing my whole doctoral thesis.

References

Cam, P. (2008). The two Adam smiths. Think, 7(20), 107-112.

Boff, E. D. O. (2014). What'S The Problem, Mr. Smith? Sheddingmore Light (Than Heat) On Adam Smith'S View Of Man. In Anais do XL Encontro Nacional de Economia [Proceedings of the 40th Brazilian Economics Meeting] (No. 013). ANPEC-Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics].

Evensky, J. (1987). The two voices of Adam Smith: moral philosopher and social critic. History of political economy, 19(3), 447-468.

Gocmen, D. (2007). Adam Smith Problem: Reconciling Human Nature and Society in the'theory of Moral. IB Tauris.

Landreth, H., & Colander, D. C. (2002). History of economic thought. 4th Ed. Houghton Mifflin Company. Boston.

Macfie, A. L. (1959). Adam Smith's Moral Sentiments as foundation for his Wealth of Nations. Oxford Economic Papers, 11(3), 209-228.

Montes, L. (2003). Das Adam Smith Problem: its origins, the stages of the current debate, and one implication for our understanding of sympathy. Journal of the History of Economic Thought, 25(1), 63-90.

Pena López, J. A., & Sánchez Santos, J. M. (2007). Los fundamentos morales de la conomía: una relectura del problema de Adam Smith. Revista de economía institucional, 9(16), 64-87.

Otteson, J. R. (2000). The Recurring "Adam Smith Problem". History of Philosophy Quarterly, 17(1), 51-74.

Pack, S. J. (1997). Adam Smith on the virtues: a partial resolution of the Adam Smith problem. Journal of the history of economic thought, 19(1), 127-140.

Rommel, T., & Winter, H. (2001). La ricchezza delle nazioni, guida e commento. Garzanti. Torino.Screpanti, E., & Zamagni, S. (2004). Profilo di storia del pensiero economico, dalle origini a Keynes., vol.1, Carocci. Roma.

Sen, A. (2010). Adam Smith and the contemporary world. Erasmus Journal for Philosophy and Economics, 3(1), 50-67.

Sen, A. (1987), On ethics and economics. Blackwell Publishing. Oxford

Smith, A. (1976). The theory of moral sentiments. Edited D. D. Raphael and A. L Macfie. Claredon Press. Oxford

Smith, A. (1994). An inquiry into the nature and causes of the wealth of nations. Edited by Edwin Cannan. The Modern Library. New York.

Teichgraeber, R. (1981). Rethinking Das Adam Smith Problem. Journal of British Studies, 20(2), 106-123.

Viner, J. (1927). Adam Smith and laissez faire. Journal of political economy, 35(2), 198-232.

Wagner-Tsukamoto, S. (2013). The Adam Smith problem revisited: a methodological resolution. Journal des? conomistes et des? tudes Humaines, 19(1), 63-99.

Wilson, D., & Dixon, W. (2006). Das Adam Smith Problem: a critical realist perspective. Journal of Critical Realism, 5(2), 251-272.

Witztum, A. (1998). A study into Smith's conception of the human character: Das Adam Smith problem revisited. History of Political Economy, 30(3), 489-513.

Index of Charts and Tables

Acknowledgements

This thesis is not the mere outcome of my personal efforts. Many people, directly or indirectly, are part of it.

First of all I wish to thank my Supervisor, Professor Lorenzo Sacconi. With this dissertation the concrete conciliation of the two spheres of ethics and economics, representing respectively my passion and my academic skills, was possible thanks to his constant support and also the approval of the scientific committee of the Doctoral School of Social Sciences at the University of Trento. I am really thankful to all of them for accepting my ideas and for helping me to translate into a rigorous economic language my love for moral philosophy.

I am grateful to Università degli Studi di Trento, to its current Dean Paolo Collini, to Fondazione Cassa Rurale di Trento and to its president Rossana Gramegna. These institutions and their staff trustingly supported me throughout my degree.

I wish to thank Professor Luigi Mittone, the former director of the Doctoral School, and Professor Marco Faillo, who collaborated with me, working directly at my projects.

My academic gratitude goes also to Marco Tecilla, the handyman of CEEL. He introduced me to the lab procedures and followed me in the development of an experimental software.

A special thank goes to Philippe Van Parijs, to Daniel James Butt and Lucas Stanczyk, who accepted me for a visiting period respectively at Louvain-la-Neuve, at Oxford and at Harvard. Without their contributions my thesis and my life would have not been as rich as they are.

My gratitude goes also to the two referees, Professor Pedro Francés-Gómez and Professor Giacomo degli Antoni. They provided wise suggestions on how to improve my research.

I express a sincere thank to my PhD fellows Tatiana, Tam, Piero, Sergiu and Valeria. I shared with them not only my time in the office, but also my ideas and advices.

A sincere thank goes also to Cassa Centrale Banca and to my closest ex-colleagues there. They accepted my decision to follow my passions.

I am then mostly grateful to my parents, Gordana and Đuzepe, who literally spent their lives to make my dreams realizable. If I have achieved all my personal goals it is thanks to them and to their love.

A huge thanks goes also to my brother, Teodoro. He is the only person who has never judged me for my choices in my life.

I wish to thank also my grandparents Mara, Vaso, Viktoria and Teodor. I have not met them from a long time, but I am sure they have always cheered for me.

My gratitudes go also to my aunts Maria and Teodora, my uncles Mirko and Renzo, my cousin Dajana and all my closest family. I have the relatives who everybody wishes.

A special thanks goes to Margarita, Mauro, Alfredo, Elisa and Asia. I consider them my second family. Thanks to them nobody was luckier than me.

I am grateful to my neighbours, available to support my path and my family, constantly, day by day for more than 25 years. Thank to Gisella, Ivo, Giulia, Matteo, Luciana, Francesco, Roberto, Paolo, Rosanna and Enzo.

A passionate thanks goes to my childhood friends *i coscritti* Alessio, Andrea, Denny, Mattia, Simone and their current wives or girlfriends Marika, Tatiana, Gloria, Martina and Chiara. Without them nothing would have been the same.

Thanks to Jessica, Matteo, Jessica and Nicola. They are a true example for everybody, therefore also for me.

I want to express a special gratitude to Francesca *the redhead* who pushed me to pursue the unthinkable and to reach the impossible. She is the most inspirational person I met in my life.

I am also grateful to Francesca's mum, Daniela, who has constantly been patient with me.

A special gratitude goes to Felice and Gabriele. They have always allowed me to experience people and situations from an opposite perspective.

I wish to thank Danielson, Elibetta, Mary, Ludo, Saretta e Andrea. They have demonstrated to me that there are always new parts of ourselves we can discover and enjoy.

I am really thankful to Annabel, the most beautiful woman I have ever met. She always insisted that I did not waste the skills I acquired through almost ten years spent at the University. She taught me also that in life not always we have a second chance. Sometimes you simply loose.

My gratitude goes then to all those Professors, fellows, friends and employees I did not name here and I met in academic environments and travelling across the world for workshops, conferences and visiting periods: all of you are like pieces of a big puzzle which is now complete.

In life you never go alone. This is now a new beginning together, certainly not an end.