



## THE METAPHYSICS OF SHARED ACTION: AN ANALYSIS BASED ON THE IDEA OF COMMITMENTS

*Metafísica da ação compartilhada: uma análise a partir da ideia de comprometimentos*

Thiago Monteiro Chaves \*

---

**Abstract:** In this paper I consider two examples of social interaction and try to figure out what is distinctive of each example by pursuing an examination of the nature of the commitments present in each situation. I discuss the ideas of pre-commitment and joint commitment and pursue an understanding of their roles in joint or shared action. The main argument of the paper is that the metaphysics of shared action is better understood through the proper kind of commitments that are constitutive of the action.

**Keywords:** Commitments, Joint action, Agreements, Social contract, convention.

**Resumo:** Neste artigo eu considero dois exemplos de interação social na tentativa de descobrir o que é distintivo em cada exemplo. Será feita uma investigação da natureza dos compromissos presentes em cada situação. As ideias de pré-compromisso e compromisso conjunto são discutidas na tentativa de se compreender seu papel na ação conjunta ou compartilhada. O argumento principal do artigo é que a metafísica da ação compartilhada é melhor compreendida através do tipo apropriado de comprometimentos que são constitutivos da ação.

**Palavras-chave:** Compromissos, Ação conjunta, Acordos, Contrato social, Convenções.

---

\* Doutorando em Filosofia pela Universidade Federal de Minas Gerais (UFMG). E-mail: [monteirochaves.uespi@gmail.com](mailto:monteirochaves.uespi@gmail.com). Telefone: (86) 99826-1912. The author thanks the State University of Piauí for having granted him a sabbatical period for carrying out this research.

## I. Preliminaries<sup>1</sup>

In this article I seek an understanding of the nature of commitments and the role they play in two kinds of social interaction. By social interaction I mean individual actions that are in some sense interdependent. This relation of interdependency will be assumed as generated by the proper manner actors are committed to the interaction. These interactions may be clearly depicted by the following examples: a group of friends travelling together by car and a business man travelling by car with a professional driver. In both cases, *the actors have committed themselves to an agreement at some prior moment in order to carry out the joint action*. The group of friends got together at some moment and planned the trip they are going to have. The business man contacted the professional driver beforehand and they established the terms upon which they are going to carry out the appropriate transactions. These two kinds of situations are very common and important to our lives. Without them we could hardly live together in society and would hardly understand ourselves as social beings. It is also very important for any science which intends to understand ourselves as social animals to explain what is going on in each of these examples. I want to understand the difference between these two situations regarding the nature of the *commitments* established in each of them. I am assuming they are different in some way, and I think this distinction may be accounted for in terms of a distinction between commitments.

First of all, we need a definition. This definition does not purport to be a definitive one, but it gives some clarification so that we can know we are talking about the same thing. I am understanding by *joint action* some sort of *shared action*, as opposed to an *individual action* which is not *shared* with anyone. But sharing an action with someone is something more than acting interdependently or acting taking the other's actions into account<sup>2</sup>. When I take my car for a trip alone, I am not sharing this action with any person. But we may think that the simple fact that I am using the public road and that I am interacting with other drivers suggests that I am sharing my action in some way. Perhaps Locke had that in mind when he thought that I tacitly consent to the rules of society if I am following them<sup>3</sup>. It is not this sense of *sharing* an action that I intend to mean. There is a clear difference between having a car travel

<sup>1</sup> I could not accommodate all the insightful criticisms I have received from an anonymous referee, specially those demanding more precise definitions of important notions as *agreements*, *contracts* and *conventions*, which the article deals poorly with. I have tried, although, to provide an analysis of some of these notions through the notion of *commitments*, which I think is a more primitive one.

<sup>2</sup> *Shared* or *joint* action is something more than *social action*, as defined by Max Weber (“Economia e Sociedade”. Vol. I. Brasília: Ed. UNB, 1994).

<sup>3</sup> “And to this I say, that every man, that hath any possessions, or enjoyment, of any part of the dominions of any government, doth thereby give his tacit consent, and is as far forth obliged to obedience to the laws of that government, during such enjoyment, as any one under it; whether this his possession be of land, to him and his heirs for ever, or a lodging only for a week; or whether it be barely travelling freely on the highway; and in effect, it reaches as far as the very being of any one within the territories of that government” (LOCKE, John. “Second Treatise On Government”. Edited by C.B Macpherson. Indianapolis: Hackett Publishing Company, 1980, p. 64).

alone and having a car travel with someone. If I am a smoker, I surely will not restrict my freedom to smoke in the car if I am alone, but if I am accompanied I may bind myself to some rule of politeness. I must take the other's desires, intentions and beliefs into consideration. But shouldn't I take other's desires, intentions and beliefs into consideration when I am having a lonely car trip? Surely I do if I don't want to have a car accident. But merely coordination does not seem to be enough to capture the elements we want to capture when we talk about two friends having a car trip together or a business man travelling with a professional driver. A hint from Margaret Gilbert<sup>4</sup> will help us here. When two people are travelling *together*, or doing anything *together*, they will promptly say that “*we* are travelling”. By saying “*we* are travelling” when two friends are travelling together, we mean something different from what we mean when we say something about the drivers in São Paulo who are jammed in traffic. “*They*” in “*they* are jammed in traffic” do not grasp the strong collective meaning which is grasped in “*they* are travelling together” in the friends’ situation. There is a sense in which the coordinated actions of the drivers in São Paulo's traffic are taking place (individually) in parallel. The action of the group of friends and the action of the business man with his professional driver are taking place in a much more *shared way*<sup>5</sup>. The fact that individual actions are coordinated is not enough to characterize the *sharedness* present in joint action. So we can define, somewhat loosely, a *joint* or *shared* action as *an action two or more people carry out together, in a manner that they can say about it that “we are doing such and such”, and we can say about them that “they are doing such and such”*. I think that this definition captures an element of *sharedness* which is present in our two inaugural examples but is not present in coordinated activities as exemplified by the loose collective action of the lone São Paulo's drivers. If the reader accepts my very weak definition, some questions can then be proposed: what is this *sharedness element* that is present in our inaugural examples and is absent in mere coordination? Is it present *in the same way* in the two examples? Does it play the *same role* in the two examples?

Can the *sharedness element* present in the inaugural examples and absent in the description of the lone drivers coordinating on traffic rules be found in the *manner* which *agreements* were made in each situation? As I described them, in the inaugural examples the *agreements* had to be made explicitly. If the friends got together and planned the trip, they must have *agreed* upon some rules and goals. They must for instance have agreed upon their destination, but disagreed upon the restaurants where they would eat. But we expect that they are able to come into a consensus so that the trip can be at least pleased. The same

<sup>4</sup> GILBERT, Margaret. “Walking Together: a Paradigmatic Social Phenomenon”. *Midwest Studies in Philosophy*, vol. XV. Notre Dame: University of Notre Dame Press, pp. 1–14, 1990; “Searle on Collective Intentions”. In: Tsohatzidis, Savas L. (ed.), 2007, pp. 31-48.

<sup>5</sup> That is why, I tend to think, we cannot understand the idea of full-blown *sociality* only with the notion of conventions or coordinated action, as exposed ingeniously by David Lewis, “Conventions: a philosophical study”. Cambridge, Mass.: Harvard University Press, 2002.). For a very articulated critique of Lewis in these terms, see GILBERT, Margaret. “On Social Facts”. Oxford: Bodleian Library, 1978.

apply to our business man. He must have contacted the professional driver in order to establish the terms of the contract. But what about the lone (coordinating) drivers? Surely they have never gathered together in order to decide on which side of the road they should drive. In the vocabulary of political philosophy, we may say that they have *tacitly* agreed on the established traffic rules. As Hobbes put it, “signs of contract are either express or by inference”<sup>6</sup>. If a teacher says “we are going to have a test tomorrow, any objections?” and the class remains in silence, the teacher *infers* that the class has agreed upon the test being tomorrow. So the agreement on traffic rules and the proper sort of agreements present in our inaugural examples seem to differ on the *manner* they take place, expressly (by signals) or tacitly (by inference). But is it not the case that the silence of the class is a *signal of agreement*? And is it not the case that when someone says “Ok, let’s do it” I may *infer* from that sentence that she has agreed on something? I do not think we can clearly make the distinction between tacit and express agreements appealing to Hobbes’s distinction. Granted that there is a difference between inferring one’s intention when one expresses oneself with words and inferring one’s intentions when one expresses oneself by being in silence. But in both cases we are assuming that signalling one’s intentions (through words or silence) is prior to the establishment of the agreement. I think that the failure to notice this has led some philosophers to put too much weight on the distinction between tacit and express agreements. I tend to think that this distinction does not even exist, since an inference of one’s intention to agree has logical primacy on the agreement itself. So it seems, actually, that the *manner* agreements are made cannot fully explain what sort of *sharedness element* is present in our inaugural examples and absent in the picture of the lone (coordinating) drivers. We must pursue that difference in the *kind* of agreements and the role they play in social interactions of the sort we are considering.

Can the *sharedness element* that is present in the inaugural examples and absent in the description of the lone drivers coordinating on traffic rules be found in the *kind* of agreements that were made in each situation? I think it can. In the inaugural examples, the agreements are taking place as *attitudes* of the actors. I mean by *attitude* any sort of stance one’s mind has towards something. Believing and intending are paradigmatic examples. As an attitude, an agreement takes place in time and space. In the picture of the lone (coordinating) drivers, agreements are not attitudes in the mind of the actors, but they are a result of a complex interrelation of the actors’ attitudes. Let me call the first kind of agreement, which I think is the important *sharedness element* present in the first couple of situations, an *actual agreement*. And let me call the second kind of agreement, which I think is present in the situation of the lone (coordinating) drivers, a *spontaneous agreement*. In calling an agreement *spontaneous* I do not mean that the actors make

<sup>6</sup> HOBBS, Thomas. (1651). “Leviathan”, edited by M. Oakeshott. London: Macmillan, 1962, p. 94.

it without thinking carefully about it, or that actors make them out of the blue. On the contrary, I mean that the actors do not make any agreement conceived as *an act of the mind*<sup>7</sup>.

With these preliminaries, I am able to present my argument concerning the kind and role of commitments in the two examples with which we opened our discussion. Contrary to the interpretation of agreements present in macro-social coordination, as exemplified by the lone drivers, the agreements present in the two inaugural examples must be conceived as *actual agreements*, as *acts of one's mind*. These actual agreements seem to establish some sort of *sharedness element* which is absent in mere coordination of individual actions, as depicted by the lone (coordinating) drivers. So the question I propose to investigate is: can the distinction between the two situations exemplified by the inaugural examples be grasped by a difference between the *commitments* to these actual agreements?

## II. Commitments in two examples of Social Interaction

In order to say what I think is the difference between the situation in which a group of friends travels together and the situation in which a business man travels in the company of a professional driver, I need to introduce some game-theory vocabulary. I will first analyse the second situation. My question is what is the kind of commitment involved in the agreements that generate the joint action. I think everyone will agree that the second example is a situation which can be accounted for in strategic terms. The business man wants someone to get him to somewhere, and the professional driver wants to get paid for a job. The “general sense of common interest”, to quote Hume<sup>8</sup>, seems enough to generate the shared action. But as we have seen, there is a crucial difference between this situation and the situation of the lone (coordinating) drivers. The lone drivers do not fulfil the condition necessary for an action to be labelled *joint*. Remind our definition of a *joint* or *shared* action as *an action two or more people carry out together, in a manner that they can say about it that “we are doing such and such”, and we can say about them that “they are doing such and such”*. I think no one would say of the lone drivers which are coordinating with each other on traffic rules that *they* are doing such and such. Do the lone drivers together form a *we*? Only if they join together in order to do something (like a protest, or to form a convoy, etc.). Do the lone drivers form a *they*? In a loose sense, yes, as in the sentence “they are jammed in traffic”. But consider the following scenario. One driver stops at a gas station. After that, another driver who is driving on the same

<sup>7</sup> I am here avowedly making reference to the tradition of spontaneous order of which Hayek is a major representative: “It would be no exaggeration to say that social theory begins with—and has an object only because of—the discovery that there exist orderly structures which are the product of the action of many men but are not the result of human design” (HAYEK, F. A. “Law, Legislation and Liberty”. London: Routledge and Kegan Paul, 1973, p. 37).

<sup>8</sup> HUME, David. (1740). “A Treatise of Human Nature”, edited by L. A. Selby-Bigge (2<sup>nd</sup> edition). Oxford: Clarendon Press, 1960, p. 490.

road stops at the same gas station. An observer at the gas station, who saw the latter driver overpassing the former some minutes ago, is asked by the manager if *they were driving together*, since the first driver forgot to pay for the gas and the second driver is still there. The observer answers that she does not know, but she “guess they are not”. The observer may have got her opinion from many observations, like the fact that the drivers did not seem to be together because one overpassed another in a cold manner. In this important sense, *they* are not driving together. So, although the example of the business man travelling accompanied by the professional driver may be treated strategically, it is not clear if the mere coordination of actions explain what is socially shared in the situation. My guess is that the element of *sharedness* present in the business man example is captured by a proper sort of *social contract* established between the business man and the professional driver. This social contract is a *special kind of convention*, but not of the same form of the conventions of traffic rules which emerge (or at least could emerge) spontaneously as in the lone (coordinating) drivers’ example. In binding oneself to the prior agreement, the business man carries out a certain decision in order to be able to make another decision in a posterior time (the same applying to the professional driver). This sort of contract or agreement depends on the mechanism of *self-binding* or *pre-commitment*<sup>9</sup>. I will try to elaborate this claim within the following analysis. Eventually, commitments understood as *self-binding* and as *joint commitments* will be distinguished.

It is undisputed that the business man and the professional driver both think of their agreements as *binding*. The driver may have proposed some price for his or her services; the business man may have promptly accepted or tried to negotiate. But they must have reached an agreement and surely find it binding. If this were not true, why would they bother to even think of making contracts? In order to understand what is going on here, we may assume that prior to the agreement or negotiation, both the business man and the driver are in a situation which they rank as less preferable than the situation in which they form an agreement. Let me call this situation the *state of nature*. It is important to note that the preferences of the actors in the state of nature relative to the example of the lone drivers may generate a convention in Lewis's sense<sup>10</sup>, but the preferences of the actors in the state of nature relative to the example of the business man cannot generate a convention. This fact must not be surprising since we saw that the agreements involved in the example of the lone drivers are spontaneous, as opposed to actual

<sup>9</sup> ELSTER, Jon. “Ulysses and the Sirens: a theory of imperfect rationality”. *Social Science Information*, 1977; “Ulysses Unbound: studies in rationality, precommitment, and constraints”. Cambridge: Cambridge University Press, 2000.

<sup>10</sup> The general definition of a convention found in Lewis can be put as: “a regularity R in the behaviour of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if, in any instance of S among members of P: (i) everyone conforms to R; (ii) everyone expects everyone else to conform to R; (iii) everyone prefers to conform to R on condition that the others do”. Lewis comes up with this definition departing from an analysis of coordination problems and coordination equilibrium. The definition is elaborated on pages 42; 58; 76 to 79 of his *Convention* (1969) (the pages refer to the 2002 edition). I will not discuss coordination problems in this text, and will assume this general definition of conventions as a fundamental one.

agreements. The example of the business man seems to be an instance of the prisoner's dilemma, and even though agreements are possible and binding by a mechanism of pre-commitment or self-binding.

An interesting way of putting this view is to ask if the preferences of the business man and the driver in the state of nature permit a strategic equilibrium to occur with a Pareto-optimal combination of strategies. A combination of strategies is an *equilibrium* if each agent has done his best given the strategies followed by other agents<sup>11</sup>. A combination of strategies is said to be Pareto-optimal if there is no other combination which is weakly preferred by any agent *and* strictly preferred by at least one agent<sup>12</sup>. If a combination of strategies is not Pareto-optimal, it is said to be Pareto-inferior<sup>13</sup>. If an equilibrium that is also a Pareto-optimal combination cannot occur, so a convention in Lewis's sense cannot emerge spontaneously from a state of nature, and a state of nature relative to a social contract must be distinguished from a state of nature relative to a convention. In order to see why this is the case, consider the preferences of two agents in a Hobbesian state of nature. In a Hobbesian state of nature, men find themselves in a permanent state of conflict. It has been widely assumed by commentators that the choices and preferences of men in this state of nature match those of a prisoner's dilemma matrix<sup>14</sup>. Let us assume that each person in the state of nature has two courses of action: seeking peace (P) or keeping war (W). And let us assume that the preferences for each person hold as follows:  $a > b > c > d$ , transitivity being the case. So, in a Hobbesian state of nature we have the matrix<sup>15</sup> number 1:

	P	W
P	b;b	d;a
W	a;d	c;c

**Matrix 1: Preferences in the state of nature relative to a social contract**

The strategy of keeping war strictly dominates the strategy of seeking peace, so whatever one agent does, it pays for the other to choose war. What makes this state of nature a state of stable conflict is the fact that, although the agents would prefer the pair of strategies P,P; the Pareto-inferior pair of

<sup>11</sup> For a similar definition, see LEWIS, David. "Conventions: a philosophical study". Cambridge, Mass.: Harvard University Press, 2002, p.8; and TAYLOR, Michael. "The Possibility of Cooperation". Cambridge: Cambridge University Press, 1987, p. 14.

<sup>12</sup> A strategy *A* is said to weakly dominates *B* if there is at least one combination of strategies for which choosing *A* pays more than choosing *B*, and for all other combinations choosing *A* pays the same as choosing *B*. A strategy *A* strictly dominates *B* if for all combinations choosing *A* always pays more than choosing *B*.

<sup>13</sup> For a close definition, see TAYLOR, Michael. "The Possibility of Cooperation". Cambridge: Cambridge University Press, 1987, pp. 13-18.

<sup>14</sup> See, for instance, TAYLOR, Michael. "The Possibility of Cooperation". Cambridge: Cambridge University Press, 1987; ULLMANN-MARGALIT, Edna. "The Emergence of Norms". Oxford: The Clarendon Press, 1977 and HAMPTON, Jean. "Hobbes State of War". *Topoi*, 4, pp. 47-60, 1985.

<sup>15</sup> For the sake of simplicity, player one is represented by row, and player two is represented by column.

strategies W,W constitutes an equilibrium. This cannot happen in a convention, and that is precisely one of the characteristics that differentiates a social contract from a convention: in the former, the Pareto-optimal combination of strategies cannot be an equilibrium; on the other hand, in a convention the Pareto-optimal combination of strategies constitutes always a kind of equilibrium. The situation is clearer when we consider matrix number 2:

	P	W
P	a;a	d;b
W	b;d	c;c

**Matrix 2: preferences in the state of nature relative to a convention**

In this situation, it pays for each agent to choose P, if the other is also choosing P. Eventually both agents will choose P, since it is an equilibrium and a Pareto-optimal combination of strategies. An instance of this matrix could be Hume's example of two men pulling the oars of a boat. If both men are not pulling the oars in a coordinated way, they are in a worst situation than if they pull the oars in a coordinated way. A convention to row in an efficient way will emerge spontaneously since the equilibrium point is Pareto-optimal.

If Hobbes's state of nature is a prisoner's dilemma as depicted by the first matrix, the perpetual conflict inherent of that situation is due to the dominance of the strategy of keeping war. External coercion happens to be the only solution for achieving peace, since contracts, although possible, are not binding. Note that a social contract in Hobbes's terms cannot be a convention because although the agents would prefer achieving peace (through a certain kind of commitment to follow a social contract), each of them would prefer lone deviation from the social contract (which is the dominant strategy in the prisoner's dilemma game). In a convention the opposite just happens to be the case: each agent prefers general conformity to conformity of all but himself. Lone deviation is a poor strategy in a coordination game when a convention is occurring. This intuition was first captured by Hume when he thought of conventions as a “general sense of common interest”<sup>16</sup>.

But what if some sort of equilibrium that happens to be a Pareto-optimal combination of strategies be achieved in the state of nature relative to a social contract? If we put *time* in the story, this could be the case. And that is what Hobbes seems to have in mind when he disapproves what the Fool has to say about keeping one's promises. In his words:

<sup>16</sup> HUME, David. (1740). “A Treatise of Human Nature”, edited by L. A. Selby-Bigge (2<sup>nd</sup> edition). Oxford: Clarendon Press, 1960, p. 490.

The Foole hath sayd in his heart, there is no such thing as Justice; and sometimes also with his tongue; seriously alleaging, that every man's conservation, and contentment, being committed to his own care, there could be no reason, why every man might not do what he thought conduced thereunto: and therefore also to make, or not make; keep, or not keep Covenants, was not against Reason, when it conduced to ones benefit [...]. This specious reasoning is neverthelesse false. For the question is not of promises mutuall, where there is no security of performance on either side [...]. But either where one of the parties has performed already; or where there is a Power to make him performe; there is the question whether it be against reason, that is, against the benefit of the other to performe, or not. And I say it is not against reason<sup>17</sup>.

Hobbes is considering here if it is rational for someone to cooperate *conditionally on other's cooperation*. This situation should be thought as an iterated prisoner's dilemma. In this situation, although any particular shot of the game has a Pareto-inferior equilibrium (the combination of strategies “keeping war”), over an indefinite sequence of the game the players may achieve coordination equilibrium with the combination of strategies “seek peace if the other seeks peace”. A coordination equilibrium is a combination of strategies such that each player is made no better off have *any player* changed strategy<sup>18</sup>. If we put time into consideration in prisoner's dilemma interactions, a convention to reciprocate is highly probable to emerge, since it will lead to a coordination equilibrium. It will be in each player's best interest to cooperate and to have others cooperating. Lone deviation would be a very poor strategy in this dynamical scenario. Let us call *CP* a strategy to cooperate conditionally on other's cooperation. *CP;CP* is a Pareto-optimal combination, because even though unilateral defection is strictly preferred to unilateral cooperation by each agent in each shot of the game, if one player chooses to cooperate, unilateral defection by the other player will lead to dynamical combinations less preferred by both players, *since they have an interest for future pay-offs*. As Russell Hardin puts it, each player has a “double incentive” to choose *CP*<sup>19</sup>. As I understand it, one incentive is due to the agent's interest in future pay-offs, and the second incentive is due to her expectation of the other's choosing to cooperate. If each player is expecting certain coordination equilibrium to be the outcome, he or she has no incentive to change his or her strategy, and so has a double incentive to choose the cooperative strategy<sup>20</sup>.

If a state of nature is an iterated prisoner's dilemma, so a social contract may emerge *as a convention*. That is to say that the combination of cooperative strategies is not just preferred to mutual and unilateral defection, but is also achieved voluntarily since it consists in a Pareto-optimal combination<sup>21</sup>.

<sup>17</sup> HOBBS, Thomas. (1651). “Leviathan”, edited by M. Oakeshott. London: Macmillan, 1962, pp. 101-102.

<sup>18</sup> For a close definition, see LEWIS, David. “Conventions: a philosophical study”. Cambridge, Mass.: Harvard University Press, 2002, and TAYLOR, Michael. “The Possibility of Cooperation”. Cambridge: Cambridge University Press, 1987.

<sup>19</sup> HARDIN, Russell. “Collective Action”. Baltimore, Maryland: The Johns Hopkins University Press, 1982, p. 171.

<sup>20</sup> For a similar argument, see TAYLOR, Michael. “The Possibility of Cooperation”. Cambridge: Cambridge University Press, 1987, p. 67.

<sup>21</sup> That is to say that choosing *CP* will maximize global utility, taking time into account.

So, as a convention, the agents' commitment to follow the social contract is due to *strategic stability*, that is to say, no one has an incentive to defect due to *the mutual expectations that a coordination equilibrium will be the outcome*.

Social contract as conditional cooperation is being understood here as an agreement to which agents in a situation  $S$  would bind themselves in order to achieve a new situation  $S'$  that is preferred to  $S$ . Although an agent may rationally agree or contract to cooperate in iterated prisoner's dilemma situations, she will succeed in doing so only if she *self-binds* or *pre-commits* herself in some specific ways. As Elster has observed in his very influential work on rationality and commitment<sup>22</sup>, a rational agent may make a decision at a prior moment (let's say at  $t_1$ ) in order to be able to make a posterior decision (at  $t_2$ ), since she is able to foresee some inversion of her preferences over time. That is to say that rational agents are capable of taking a *strategic stance* towards their preferences over time. In Elster's famous example, we have Ulysses, who at  $t_1$  knows he wants more than everything to go back homeward; but he foresees that when passing by the sirens (at  $t_2$ ) he will prefer taking a hot bath with them than having dinner with Penelope at home. Being aware of the weakness of his will<sup>23</sup>, Ulysses asks his friends to tie him to the mast. He chooses (at  $t_1$ ) to restrict the set of his possible actions at *time 2* in order to achieve a maximal pay-off over time. Elster defines *self-binding* or *pre-commitment* as a kind of *rationality over time*, which meets the following requirements:

- (i) To bind oneself is to carry out a certain decision at time  $t_1$  in order to increase the probability of another decision being carried out at time  $t_2$ ; [...]
- (ii) If the act at the earlier time induces a change in the set of options that will be available at the later time, this counts as binding oneself only if the change is a reduction of the feasible set; [...]
- (iii) The effect of carrying out the earlier decision must be to set up some causal process in the external (material) world; [...]
- (iv) The resistance against carrying out the decision at  $t_1$  must be smaller than the resistance that would have opposed the carrying out of the decision at  $t_2$  had the decision at  $t_1$  not intervened. [...]
- (v) The act of binding oneself must be an act of commission and not of omission<sup>24</sup>.

Returning to the example of the business man travelling with the professional driver, it is easy to see that the agreement made between them must meet these conditions. (i) The establishment of the prior agreement considerably increases the probability of the decision to follow it be made at a posterior time.

<sup>22</sup> ELSTER, Jon. "Ulysses and the Sirens: a theory of imperfect rationality". *Social Science Information*, 1977; "Ulysses Unbound: studies in rationality, precommitment, and constraints". Cambridge: Cambridge University Press, 2000.

<sup>23</sup> The situation can be described as an awareness of the weakness of one's will, or as a prediction that one's preferences will be inverted over time. Inversion of preferences may occur by the mere passing of time, due to utility discounting. For a discussion, see ELSTER, Jon. "Ulysses Unbound: studies in rationality, precommitment, and constraints". Cambridge: Cambridge University Press, 2000.

<sup>24</sup> ELSTER, Jon. "Ulysses and the Sirens: a theory of imperfect rationality". *Social Science Information*, 1977, pp. 470-474.

(ii) The agreement restricts the set of possible actions at a posterior time. (iii) The agreement set up a causal process in the world which is responsible for the occurrence of the joint action. (iv) The resistance against making the agreement must have been smaller than the resistance against acting contrary to it, had it not been established. (v) The agreement must have been an *actual agreement* (an act of one's mind, an act of commission).

My claim is that in this example we have a case of social contract as a form of actual agreement bound by a mechanism of self-binding or pre-committing. And the role this sort of self-binding plays in generating the joint action is essentially strategic. In fact, in the kind of transaction depicted by this example, lone deviation is a dominant strategy. The business man would prefer having the driver cooperating but not himself, since he would maximize his own pay-off. But since it happens that this kind of transaction is very common, the business man reasonably acts prudentially, taking future interactions into account. He is not myopic. His pre-commitment to the contract is *a prudential stance towards his preferences over time*. An interesting observation here is that the way a *contract* is often conceived by jurists as “a promise or set of promises for the breach of which the law gives a remedy”<sup>25</sup> presumes that pre-commitment (conceived in the way discussed above) is the normal state, deviation being a pathological case. Following the same line of thought, Elster has claimed that we can think of *constitutions* as a set of pre-commitments against the preference inconsistencies which take place over time in the phenomenon of actual agreements<sup>26</sup>. The strategic inconsistencies analysed by Elster are exactly the ones that occur when someone tends to act in accordance with a strict dominant strategy in one shot of a prisoner's dilemma game, but at the same time has to consider the dynamical properties of the iterated game. The mechanism of self-binding permits a rational agent to keep an eye on global maximization over the time. When rationality fails, the law assumes its role.

If we assume the state of nature as an iterated prisoner's dilemma scenario, we may only think of the nature of the commitments involved as a form of *self-binding*. But if these commitments in the social contract are understood as self-binding, so the agents' preferences and goals must be conceived as *idiotic*<sup>27</sup> by principle, that is to say, they cannot be understood as *joint goals*, only as “*one's own*” or *idiotic goals*, which means that the way Hobbes and many other philosophers have conceived the social contract is not *that* social as it may appear. In contrast, the kind of commitments found in the example of the group of

<sup>25</sup> The quotation is from the *American Restatement of the Law of Contracts*, which I took from GILBERT, Margaret. “Is an Agreement an Exchange of Promises?”. *The Journal of Philosophy*, vol. 54, no. 12, 1993, pp. 627–649.

<sup>26</sup> ELSTER, Jon. “Ulysses Unbound: studies in rationality, precommitment, and constraints”. Cambridge: Cambridge University Press, 2000.

<sup>27</sup> I am using this expression according to Hannah Arendt's explanation in *The Human Condition*: “(...) a life spent in the privacy of 'one's own' (*idion*), outside the world of the common, is 'idiotic' by definition” (ARENDDT, Hannah. “The Human Condition”. Chicago: The University of Chicago Press, 1998, p. 38).

<i>intuitio</i>	ISSN 1983-4012	Porto Alegre	Vol.10 – Nº.1	Julho 2017	p. 107-122
-----------------	-------------------	--------------	---------------	---------------	------------

friends travelling together may be understood as *joint commitments*, what would in fact capture its social character.

As Margaret Gilbert has proposed in many articles on the subject<sup>28</sup>, it is exactly the sort of *fact* sociality is that cannot be accounted for in terms of *self-binding*. A kind of willingness in *joining forces* with each other in accepting a common goal cannot be reducible to unilateral pre-commitments, in which one binds oneself under expectations of the other's performance, but “upon whom one has no claim”<sup>29</sup>. In joint action, we form *joint commitments*, which mean that we are bound *simultaneously and interdependently*. If a group of friends are travelling together by car, they have obligations that are not mere prudential obligations, in the sense of conditional cooperation discussed above. They have mutual obligations that are generated because they form a *plural subject* by binding themselves and joining forces to achieve a common goal.

I think that in order to understand what is going on in the situation of a group of friends travelling together as opposed to the situation of the business man, we must come up with an adequate conception of a *collectivity* or *social group*, a task that is crucial to the epistemology of the social sciences. This sort of investigation should shed light on important ontological questions for the social sciences, as the understanding of the kind of entity a social group is: if an aggregate of individuals or something that form a “different substance”, which is bigger than the summation of the individuals involved<sup>30</sup>. An interesting strategy, followed by Gilbert, is to analyse small scale cases of what she calls *shared action*, exemplified as *going for a walk together, having a conversation together* and the like. These shared actions should embed the essential properties of *collectivity*.

The question which captures Gilbert's attention in this sort of analysis is: what is happening *in the world* when a group of friends is travelling together<sup>31</sup>? Gilbert has her own terminology for Lewis's strategy for answering this question: the *shared personal goal analysis* of joint or shared action. This analysis may come in a weak or a strong way, and she dismisses both. The weak version says that it is sufficient for a joint action to occur that each agent shares the same goal. If you and I are travelling together, you have the goal to travel with me and I have the goal to travel with you. But here we still do not have the general *sense* of common interest, since a cognitive link is missing. Although the action

<sup>28</sup> See GILBERT, Margaret. “Walking Together: a Paradigmatic Social Phenomenon”. *Midwest Studies in Philosophy*, vol. XV. Notre Dame: University of Notre Dame Press, 1990, pp. 1–14; “Is an Agreement an Exchange of Promises?”. *The Journal of Philosophy*, vol. 54, no. 12, 1993, pp. 627–649; “Searle on Collective Intentions”. In: Tsohatzidis, Savas L. (ed.), 2007, pp. 31–48.

<sup>29</sup> GILBERT, Margaret. “Walking Together: a Paradigmatic Social Phenomenon”. *Midwest Studies in Philosophy*, vol. XV. Notre Dame: University of Notre Dame Press, 1990, p. 7.

<sup>30</sup> This question may be put in methodological form: should we proceed with the investigation in the social sciences from individual agency to macro-structures or should we proceed inversely? An inaugural treatment of this question is found in POPPER, Karl. “The Open Society and Its Enemies” (2 volumes). London: Routledge, 1945.

<sup>31</sup> Gilbert's preferred example is *two people walking together*.

would be loosely joint, it would not be *intentionally* joint<sup>32</sup>. The strong version of the shared goal analysis maintains that it is sufficient for you and me to travel together if we have the same goals and this is *common knowledge* between us. We do not need here to investigate this notion any further, the assumption being that the event of going for a walk together, what characterises a joint or shared action, is, *according with the strong shared goal analysis*, fully captured by motivational and cognitive states. The capacity for sociality is thought here as a capacity for sharing goals and reasoning on common knowledge.

Gilbert thinks that a closer examination of our intuitions on what is going on when a group of friends is travelling together reveals that shared goals and common knowledge are not everything that is happening *in the world* when this joint action is occurring. What is happening *in the world* is relevant here because Gilbert is reflecting on the ontology of social phenomena. She claims that something more *in the world* is going on besides motivational and cognitive states. Even if the conditions of common goal and common knowledge are satisfied, “a crucial feature of going for a walk together will be lacking”<sup>33</sup>. This feature must be thought as *constitutive* of the social phenomenon. What is missing is a proper kind of content *de jure*. Keeping with Gilbert's preferred example, when John and Mary are in agreement walking together, be whatever logical conditions needed for this joint action fulfilled, John has more than a prudential reason to rebuke Mary if she *breaks the deal*. Suppose Mary suddenly stops and turns back, without any explanations. If John's and Mary's goals are completely out in the open, if Mary stops out of the blue and turns back, it is in John's interest to rebuke Mary, and we may say that this prudential reason obliges John to do so if he wants to have his goals fulfilled (of course John and Mary should monitor the situation and do whatever is needed to keep the situation according to their interests). But looking closer, it seems that John has more than a prudential reason to rebuke Mary, it seems that he is *entitled* to do so. Mary is not just obliged by her self-interest to keep walking with John (as she prefers to), she also *has an obligation* to keep acting jointly *if they have agreed to*. We may say that John has the right to, or is entitled to, rebuke Mary if the joint action is frustrated by Mary's fault, the same applying to Mary. The crucial claim of Gilbert's argument is that if the existence of such *entitlement* is not established, there can be no *joint* or *shared* action in any important sense.

<sup>32</sup> An action can be loosely joint if it consists only of an aggregate of individual actions establishing no cognitive relation between them. Searle offers an informative picture: “Imagine that a group of people are sitting on the grass in various places in a park. Imagine that it suddenly starts to rain and they all get up and run to a common, centrally located shelter. Each person has the intention expressed by the sentence “I am running to the shelter.” But for each person, we may suppose that his or her intention is entirely *independent* of the intentions and behavior of others. In this case there is no collective behavior; there is just a sequence of individual acts that happen to converge on a common goal. Now imagine a case where a group of people in a park converge on a common point as a piece of collective behavior. Imagine that they are part of an outdoor ballet where the choreography calls for the entire corps de ballet to converge on a common point” (SEARLE, John. “Collective Intentions and Actions”. In: Cohen; Morgan; Pollack, 1990, pp. 402-403, my italics). “Independent” here may be read as *cognitively independent*, but this is not Searle's own reading.

<sup>33</sup> GILBERT, Margaret. “Walking Together: a Paradigmatic Social Phenomenon”. *Midwest Studies in Philosophy*, vol. XV. Notre Dame: University of Notre Dame Press, 1990, p. 3.

Some qualifications are needed here. Although Gilbert's argument descends from contractualist views of rights and obligations concerning cooperation, it is not easily subject to the same sort of critique to which traditional contractualist views are. Take, for instance, Rawls's argument that one is obliged not to free ride in a cooperative situation: "when a number of persons engage in a *mutually advantageous* cooperative venture according to rules, and thus restrict their liberty in ways necessary to yield advantages for all, those who have submitted to these restrictions *have a right* to similar acquiescence on the part of those who have benefitted from their submission"<sup>34</sup>. Rawls's argument is not of the same nature of Gilbert's claims, since the principle of fairness alluded here is ultimately based on what prudence would determine us to agree. Russell Hardin accuses Rawls to deduce rights and obligations from conventional expectations<sup>35</sup>. The idea behind his critique is that the obligation not to free ride in a cooperative venture is generated by one's expectation that other players will be cooperating *under the assumption* that one is cooperating as well (conform our previous analysis). Prudence alone is the source of obligation here. But Hardin's critique does not threaten Gilbert's claims, since the sort of rights and obligations which constitute a joint activity are, according to Gilbert, not deduced from conventional expectations. The important consideration here is that Gilbert's analysis of the normative content of joint action is immune from critiques that are Humean in nature. The obligations that emerge in joint action are not *prudential obligations*, generated from the mutual advantage in keeping oneself bound to the original agreement. In conditional cooperation, each agent unilaterally pre-commits oneself to perform under a justified expectation that the other is unilaterally pre-committed to perform as well. Agents feel obliged to bind themselves in such a way for prudential reasons, and the normativeness of self-binding is essentially based on these prudential reasons. On the other hand, in joint commitments agents simultaneously and interdependently commit themselves to a *joint goal* generating a *plural subject* which is the bearer of the common goal. The proper sort of normativeness that emerges here cannot, according to Gilbert, be captured by mere prudential reasons. Each actor that composes a plural subject feels that he or she has an obligation to others and have a right to other's actions as well. The intuition is that if a group of friends is travelling *together*, it is just *wrong* for one of them to *unilaterally* change his or her destination. It is not only wrong because he or she would not be acting prudentially (presuming that he or she *prefers* to travel together with the group), but also and more important because in acting together people are jointly committed to one another, they are bound to one another in such a way that is *constitutive* of the joint action.

<sup>34</sup> RAWLS, John. (1971). "A Theory of Justice". Cambridge Mass.: Harvard University Press, 1999, p. 96, my italics.

<sup>35</sup> Hardin, Russell. "Collective Action". Baltimore, Maryland: The Johns Hopkins University Press, 1982, p. 179.

### III. Conclusion

I think I am now in a position to summarize my claims. I am presuming that there is a difference between the situation in which a group of friends agrees to travel together by car, and a situation in which a business man agrees to pay a professional driver in order to have the driver taking him to somewhere, and the driver accepts this transaction as well. There is an important sense in which these two actions are *joint* or *shared*, a sense that is not present in the example of the lone (coordinating) drivers. Joint or shared action is something more than acting in coordination. This sense is captured by the way we promptly say of those two inaugural examples that “*they* are doing such and such”. But what makes the two first cases distinct from each other? First, the *kind* of commitments involved in each situation. The group of friends *join forces in seeking a common goal*. They simultaneously and interdependently commit themselves to the pursuit of this goal. They are *jointly committed* to this goal. A proper sort of rights and obligations emerge here as an irreducible social phenomenon. Such is the kind of commitment involved here. On the other hand, when the business man contracts the services of the professional driver, although *they* are travelling together, one is *self-bound or pre-committed* to the joint action. They indeed *share an action* (what precisely distinguishes this case from the lone drivers’ case), but their commitment is *a prudential stance they take towards their preferences over time*. They are bound by expectations of conditional cooperation, and the proper sort of rights and obligations generated here are only due to prudential reasons. Their preferences are *idiotic* as opposed to the *joint* preferences of the group of friends.

Another important difference concerns the *role* that commitment to agreements play in each scenario, and precisely here I am in little disagreement with Lewis. Lewis argues that agreements may cause actions to reach coordination, but once caused by agreements they no longer depend on them. They “bear no trace of their origin”<sup>36</sup>. But if we do not ascribe to agreements a *constitutive* role in spite of a *causal* role in some joint actions, we will not be able to trace the origin of the commitments which distinguishes cases exemplified by two friends walking together, or having a conversation together, or travelling together, as opposed to cases like a commercial transaction. In all these social situations, people are acting jointly, sharing an action in some very important sense. But in some cases people's commitments to the agreements in carrying out the joint action will be *constitutive* of the action itself, in spite of merely carrying it out. The difference between our two inaugural examples purports to show that.

### References

ARENDDT, Hannah. (1958) “The Human Condition”. Chicago: The University of Chicago Press, 1998.

<sup>36</sup> LEWIS, David. (1969). “Conventions: a philosophical study”. Cambridge, Mass.: Harvard University Press, 2002, p. 84.

<i>intuitio</i>	ISSN 1983-4012	Porto Alegre	Vol.10 – Nº.1	Julho 2017	p. 107-122
-----------------	-------------------	--------------	---------------	---------------	------------

- COHEN, Philip R.; MORGAN, Jerry; POLLACK, Martha E. (eds.). "Intentions and Communication". Cambridge, Mass.: MIT Press, 1990.
- ELSTER, Jon. "Ulysses and the Sirens: a theory of imperfect rationality". *Social Science Information*, 1977.
- \_\_\_\_\_. "Ulysses Unbound: studies in rationality, precommitment, and constraints". Cambridge: Cambridge University Press, 2000.
- GILBERT, Margaret. "On Social Facts". Oxford: Bodleian Library, 1978.
- \_\_\_\_\_. "Walking Together: a Paradigmatic Social Phenomenon". *Midwest Studies in Philosophy*, vol. XV. Notre Dame: University of Notre Dame Press, pp. 1-14, 1990.
- \_\_\_\_\_. "Is an Agreement an Exchange of Promises?". *The Journal of Philosophy*, vol. 54, no. 12, pp. 627-649, 1993.
- \_\_\_\_\_. "Searle on Collective Intentions". In: Tsohatzidis, Savas L. (ed.), 2007, pp. 31-48.
- HAMPTON, Jean. "Hobbes State of War". *Topoi*, 4, pp. 47-60, 1985.
- HARDIN, Russell. "Collective Action". Baltimore, Maryland: The Johns Hopkins University Press, 1982.
- HAYEK, F. A. (1979). "Law, Legislation and Liberty". London: Routledge and Kegan Paul, 1973.
- HOBBS, Thomas. (1651). "Leviathan", edited by M. Oakeshott. London: Macmillan, 1962.
- HUME, David. (1740). "A Treatise of Human Nature", edited by L. A. Selby-Bigge (2nd edition). Oxford: Clarendon Press, 1960.
- LEWIS, David. (1969). "Conventions: a philosophical study". Cambridge, Mass.: Harvard University Press, 2002.
- LOCKE, John. (1690). "Second Treatise On Government". Edited by C.B Macpherson. Indianapolis: Hackett Publishing Company, 1980.
- POPPER, Karl. "The Open Society and Its Enemies" (2 volumes). London: Routledge, 1945.
- RAWLS, John. (1971). "A Theory of Justice". Cambridge Mass.: Harvard University Press, 1999.
- SEARLE, John. "Collective Intentions and Actions". In: Cohen; Morgan; Pollack (eds.), 1990, pp. 401-415.
- TAYLOR, Michael. "The Possibility of Cooperation". Cambridge: Cambridge University Press, 1987.
- TSOHATZIDIS, Savas L. (ed.), "Intentional Acts and Social Facts: Essays on John Searle's Social Ontology". Dordrecht: Springer, 2007.
- ULLMANN-MARGALIT, Edna. "The Emergence of Norms". Oxford: The Clarendon Press, 1977.
- WEBER, Max. (1922). "Economia e Sociedade". Vol. I. Brasília: Ed. UNB, 1994.

Recebido em: 02/04/2017

Aprovado para a publicação em: 29/05/2017

<i>intuitio</i>	ISSN 1983-4012	Porto Alegre	Vol.10 – Nº.1	Julho 2017	p. 107-122
-----------------	-------------------	--------------	---------------	---------------	------------