A convention or (tacit) agreement betwixt us

Article · March 2012

Source: OAI

CITATIONS
2

READS
138

4 authors:

Giulia Andrighetto
Italian National Research Council
98 PUBLICATIONS **748** CITATIONS
SEE PROFILE

Luca Tummolini
Italian National Research Council
63 PUBLICATIONS **1,115** CITATIONS
SEE PROFILE

Cristiano Castelfranchi
Italian National Research Council
401 PUBLICATIONS **11,827** CITATIONS
SEE PROFILE

Rosaria Conte
Italian National Research Council
167 PUBLICATIONS **4,471** CITATIONS
SEE PROFILE

A convention or (tacit) agreement betwixt us

Giulia Andrighetto, Luca Tummolini, Cristiano Castelfranchi, Rosaria Conte

Institute of Cognitive Sciences and Technologies
Via San Martino della Battaglia, 44
00185, Roma, Italy
{giulia.andrighetto; luca.tummolini; cristiano.castelfranchi; rosaria.conte}@istc.cnr.it

Abstract. The aim of this paper is to show that conventions are sources of tacit agreements. Such agreements are tacit in the sense that they are *implicated* by what the agents do (or forbear to do) though without that any communication between them be necessary. Conventions are sources of tacit agreements under two substantial assumptions: (1) that there is a salient interpretation, in some contexts, of every-one's silence as confirmatory of the others' expectations, and (2) that the agents share a value of not hostility. To characterize the normativity of agreements the Principle of Reliability is introduced.

1 Introduction

Conventions are social means for the sake of common ends. A common end needs not be a desire we pursue together (i.e. our joint desire to meet each other). A set of desires that are jointly co-realizable may suffice (i.e. our self-regarding desires to avoid collisions in traffic): coincidence of interests is, at least, agreement in desires[1]. Conventions describe a way to behave in recurrent situations, which is sufficient to obtain something we all want but which is at risk because of our reciprocal interference. Conventions are not necessary means though. They are arbitrary since some other way to behave might serve the same purpose. That is, our common interest (our ends in agreement) is to be fulfilled if our desires for the means are also in agreement, when at least another possible arrangement is foreseeable. To be useful, conventions should be stable: when established, conventions perpetuate themselves. And they are so because it is in the best interest of all of us to keep acting as we do, if the others do the same. Moreover this fact, as all the rest, is common knowledge between us, so much that, if one bothered enough to reason from the perspective of another fellow, it would discover that conformity to the convention is in the best interest of all the others and so be assured that the regularity will keep on.

[11] Two agents "agree in desires if exactly the same world would satisfy the desires of both; and a world that satisfies someone's desires is one wherein he has all the properties that he desires de se and wherein all the propositions hold the he desires de dicto. Agreement in desire makes for harmony" [20]. On the distinction between attitudes de dicto and attitudes de se see [18].

Dagstuhl Seminar Proceedings 09121
Normative Multi-Agent Systems
http://drops.dagstuhl.de/opus/volltexte/2009/1919

The idea that conventions are a peculiar kind of regularity in behaviour along these lines has been forcefully defended by David Lewis [15] [17], whose theory is considered by him as analogous to the one sketched by Hume in the *Treatise* while discussing the origin of justice and property.

According to this view, conventions *describe* a self-enforcing behavioural pattern; do they *prescribe* it too?

Many critics of Lewis' theory of conventions have been sceptical about his analysis, precisely because it seems that Lewis has missed the normative component. One way to put the critique being that conventions are not mere regularities but rules, not only regularities *de facto* but also regularities *de jure* [24] [8] [21]. Telling the truth when one is speaking in English is not only something that we *usually* do, it is something we *ought* to do. And the same is true for all the conventions we are parties of. Conformity to our conventions is not just what we happen to do, is something that is "required" from us.

Though often not acknowledged, Lewis' theory is able to readily accommodate these critiques. It is explicitly stated, in fact, that: "any convention is, by definition, a norm which there is some presumption that one ought to conform to (…) it is also by definition a socially enforced norm: one is expected to conform, and failure to conform tends to evoke unfavourable responses from others" [15].

What kind of norm any convention is, however, is not immediately clear.

Lewis suggests that there may be all sorts of reasons why, for any *particular* convention, one ought to conform to that particular regularity. If the convention originated by an exchange of promises, then one ought to act also to keep the promise; if the convention is also a social contract, then one ought to reciprocate the obtained benefit. Notwithstanding so, there are also *general* reasons why one ought to conform which are valid for any regularity that qualifies as a convention, for any population relative to which the convention exists, and for any situation the convention applies to.

Such general reasons derive from the fact that by conforming to a convention one acts in one's own best interest, and, at the same time, in a way that answers to others' preferences, *when they reasonably expect one to do so*. Both acting in one's own best interest and in the way that is in the interest of others (when they reasonably expect one to do so) are something that, according to Lewis, "we do presume, other things being equal, that one ought to do". If the former is a requirement of instrumental rationality, the latter stems from a moral principle that is, somehow, acknowledged by us. But is it so?

Alice has a good reason to expect Bob to do an action because John told her so. She completely trusts John; hence Alice has a reason to believe what John says. She really wants Bob to behave in that way and she reasonably expect him to behave so. Is this sufficient for Bob to be required to do the action in question? If Bob is not in any way responsible for what Alice believes, why ought he do that action?

Similarly, one can be reasonable in expecting conformity to a certain convention given widespread conformity in the population (e.g. it is reasonable to expect the next driver to keep the right given one's experience with what this population of drivers usually do) even without any direct experience with those of the others one is now dealing with (e.g. one's expectation about what the next driver will do is not grounded in one's experience with that driver). How is it, then, that such anonymous agent is

responsible for expectations he has not induced? Though our intuition tells us that any anonymous driver ought to conform to the convention that prevails in that population, it is not evident why he is so bound since he bears no immediate responsibility for what anyone reasonably expects from him.

In order to clarify what kind of normativity characterizes any convention, in this paper we will argue that *conventions are sources of agreements*, though it is not necessarily by agreement that a convention is established. That a convention is an agreement is usually considered as a platitude, so much that once the notion of convention is understood, it is thereby clarified in which sense a behavioural regularity is also an agreement. Agreements however are not only agreements in desire that as a consequence produce regularities in behaviour. Agreements are specific kinds of *social relationships* between the agents, and are created with the aim to produce such agreements in desires (see Sections 4 and 6). Agreements are considered by Lewis as a means to produce a system of mutual expectations [15], but what is important for us, is that the converse also holds: a system of mutual expectations of the kind presupposed by a convention is a source of agreements. This suggestion, however, seems to be counter-intuitive given that conventions are typically maintained without the need of any communication between the parties. If this is true, how can agreements be established without communication? How can conventions be real agreements and not a way to behave *as if* we have agreed though we didn't?

It has been Hume's suggestion that a convention is an "agreement betwixt us, though without the interposition of a promise". The aim of this article is to clarify what kind of agreement is established, once a convention is in place. By doing this, the peculiar normativity of conventions will be also analysed. The normativity of conventions is the same normativity of agreements, because conventions become agreements, *tacit agreements* but agreements nonetheless.

2 From preferences to reasons to conform

Let's first rehearse what a convention is.

Few years after his first contribution on the topic [15], Lewis amended his original analysis, by offering the following definition [17]:

> A regularity R, in action or in action and belief, is a *convention* in a population P if and only if, within P, the following six conditions hold:
> 1. Everyone conforms to R.
> 2. Everyone believes that the others conform to R.
> 3. This belief that the others conform to R gives everyone a good and decisive reason to conform to R himself.
> 4. Everyone who believes that at least almost everyone conforms to R will want the others, as well as himself, to conform.
> 5. R is not the only possible regularity meeting the last two conditions. There is at least one alternative R' such that the belief the others conformed to R' would give everyone a good and decisive reason to conform to R' likewise.

3

6. Finally, the various facts listed in conditions (1) to (5) are matters of common (or mutual) knowledge.

This definition is meant to capture the core of our common concept of convention whereby we are ready to acknowledge that a practiced regularity in acting or in acting and believing (condition 1), that everybody expects widespread conformity to (condition 2), that is arbitrary (condition 5) but serving our common ends (condition 4), and that perpetuate itself and it is stable because it is openly known that past conformity gives everyone a reason to go on conforming (condition 3 and 6), is what we would indeed consider one of our conventions[2].

Lewis has amended his 1969 analysis in several ways, but one change was particularly relevant to his original target, that is, the explanation of what convention underlies the use of a certain language by a population. Since clause (3) was originally formulated in terms of a conditional *preference* for conformity, the only acceptable regularities were in action alone: it makes no sense to prefer to believe something, since you cannot choose what to believe. As a consequence the convention governing the use of a language was characterized as a *convention of truthfulness* in that language, whereby only speakers conform to the convention, and, by doing so, coordinate with past speakers who truthfully used that language in the past [15]. Differently, the amended definition makes room for a regularity in *action and belief* to count as a convention since others' conformity provides one with a *reason* either to do or to believe something. The formulation in terms of reasons for conformity (instead of preferences for conformity) opens the way for coordination between speakers and hearers so that the convention whereby a population uses a language becomes a *convention of truthfulness and trust*, that is, a regularity in which conformity for speakers means to do something (i.e. speak truthfully) and for hearers to believe what speakers say since both share an interest in communicating and each other conformity is a practical or an epistemic reason to conform.

3 Generalizing Lewis: trust by convention

Once, however, convention is defined in this way it is also clear that trust, properly defined, is not peculiar of conventions of language alone.

Trust, in fact, is both a state of mind and a behaviour [3] in which an agent expects and wants that another agent does something, relies on this agent to behave in this way, and does in fact delegate the fulfilment of one's own desire to another agent. By trusting another agent, one makes oneself vulnerable; one exposes oneself to the risk that the other will not behave in the expected way and so frustrating one's desires.

Crucial for trusting is *reliance* on an agent for something, and not just reliance on something happening [13]. When we rely on something happening, say that the train will arrive on time, we assume that it will happen (usually because we believe that it

[2] Many of course have challenged this analysis under several different aspects. Here we will just assume it as correct, and focus on how, within such a framework, the normativity of conventions can be accounted for. For a critical assessment of Lewis' theory see [8]. For a recent account see [10].

will happen) and plan or intend accordingly. Differently, when we trust an agent we rely on him *as* an agent, that is, as an autonomous entity driven by his own beliefs and desires that are his reasons either to believe or do something. That is, when we rely *on an agent* to behave in some way, we assume that such agent will behave in that way and we plan or intend on this basis *because the agent's behaviour is based on his reasons*, not just because he his coerced to behave in that way. If I coerce you into giving me your pocket, I rely on the fact that you will give me your pocket but I do not rely *on you* to give it to me; there's no question of trust in coercive interactions. By the same token, trust also presupposes that the trustee is not motivated by a hostile attitude towards the trustor, so much that the trustor at least believes or assume such non-hostility in those the trustor rely on [3]. Trust is a fundamental non-hostile attitude[3].

Trust is relative to a desire one is pursuing and whose fulfilment depends on another agent's behaviour[4]. Desires can be either epistemic desires (i.e. the desire to know something or to know whether something is true or not) or practical ones (i.e. the desire that the world be in some way). Correspondently, reliance on somebody to behave in some way can be either for an epistemic or practical desire. That is, if I epistemically rely on you, I rely on you to do something in order to fulfil an epistemic desire of mine, something that typically happens by way of communication[5]. In such a case, epistemic reliance entails that I assume that you will truthfully communicate with me because you are motivated (for some reason) to so act, and, on this basis, believing what you want me to believe. This is the kind of trust that Lewis had in mind, where trust is coming to believe something. On the other hand, when I practically rely on you, I rely on you to do something in order to realize a state of affairs that I desire. If I rely on you to drive on the right side of the road, such practical reliance entails that I assume that you will so drive since you have a reason to do such an action, and I will behave accordingly. In both situations, by coming to believe something or by acting on the basis of my expectation about you, I trust you.

Finally, one trusts on the basis of reasons. But what are the reasons to trust another? Sometimes trusting may be 'irrational', as when, by making oneself vulnerable, one thereby creates a selfish reason for another to exploit such vulnerability[6]. Other times, trust is perfectly reasonable as when one relies on another to do something simply because it is also in the interest of the other agent to act in that way. Even if in this case trust is reasonable and more secure, it is not of course without risks given that the other could simply change his mind and act differently.

What is then the relation between trust and conventions?

According to the definition of convention given above, in any convention, the agents do conform to some regularity, want the others to conform, expect future conformity of their fellows, and this belief is a reason to conform (i.e. a practical reason to do an action or an epistemic reason to believe something). Given this, it is clear that *any act of conformity to a convention is also an act of reliance on the others to*

[3] See Section 6 for the relevance of not being motivated by hostile attitudes.

[4] That is, the trusting agent believes to be dependent on another one to obtain something he desires [3],[4].

[5] Though this is not necessary, see Section 8.

[6] If one extends a loan to another assuming that the other will do his best to repay it, one also gives the other a selfish reason not to repay it; see [1].

5

conform: the reason to conform is also the reason to trust, to rely upon the others and doing something accordingly. Since, however, by conforming one trusts in others' conformity, that is, in their trust in oneself, *the regularities that count as conventions are regularities of reciprocal trust*. Moreover, since the expectation of conformity is a reason to conform, trust is based on trust: I have a reason to trust you if you trust me and you have a reason to trust me if I trust you.

More precisely: a regularity R in reciprocal trust in a population P is a convention if and only if the following six conditions hold:

1. Everyone conforms to R, that is, everyone reciprocally trusts each other.
2. Everyone believes that others conform to R, that is, everyone believes that the others trust in oneself.
3. This belief that everyone conforms to R (i.e. rely on each other) gives everyone a good and decisive reason to conform to R himself (i.e. to rely on the others). That is, the belief that everyone reciprocally relies upon each other is a reason for everyone to rely on the others. This reason can be for practical reliance, if conforming to R is a matter of reliance on the others to act in a certain way and acting oneself accordingly. This reason can be for epistemic reliance, if conforming to R is a matter of reliance on the others to act in a certain way and believing oneself accordingly. In the case of a regularity of practical reliance, some desired end may be reached by relying upon the others and acting accordingly, provided that the others also rely upon on each other; therefore he wants to rely on the others and act, if they so rely and act. In the case of a regularity of epistemic reliance, his beliefs together with the belief that the others practically rely upon himself are premises that deductively imply or inductively support a conclusion, and by believing this conclusion he would thereby conform to R (i.e. he would epistemically rely on the others).
4. Everyone who believes that the others conform to R (reciprocally trust each other) will want the others, as well as himself, to conform (i.e. to trust on oneself).
5. R is not the only regularity meeting the last two conditions. There is at least one alternative regularity R' in reciprocal reliance which would perpetuate itself instead of R.
6. The various facts listed above in conditions (1) to (5) are matters of common knowledge.

Any convention then is always a form of reciprocal trust, which is sustained by past reciprocal trust, and that breeds future trust. Such reciprocal trust is reasonable since by trusting we are able to agree in the choice of the means to fulfil each of our individual end. Reciprocal trust can originate in several different ways, for example by explicit agreement. However, a regularity of reciprocal trust qualifies as convention by the way it perpetuates itself, and not by the way it originates. There is *trust by convention* whenever it is our reciprocal trust that, together with our desires for the end, gives us a reason to keep on trusting.

Generalizing the definition in this way is faithful to Lewis' analysis because no modification or additional clause has been proposed. Whether it is also fruitful to understand the peculiar normativity of conventions will be explored in what follows.

Since we intend to argue that such normativity stems from agreements, in the next section we turn to this issue.

4 Agreements without promises

It is natural, and correct, to view the practice of promising as a social device for making agreements. It is also natural, but wrong, to consider agreements primarily as 'an exchange of conditional promises'[7]. Though it's true that such an exchange creates a binding agreement, even my unconditional promise to you is sufficient for creating an agreement between us on something I will do, no matter what. Mutual conditional promises may be the natural model for contracts, but they hardly are the general analysis of agreements, at least if, as Hume has suggested, agreements might exist between us without the interposition of any promise.

In the contract view, agreements create obligations (and rights) on the parties entering into it due to such exchange of conditional promises. However it is sensible to consider promises just as one possible way to create agreements. Another opportunity is to avail oneself of a suggestion of a third party that, if it meets the interests of all, might be jointly accepted. However, that an agreement be mutual is also dispensable. Giving permission, for instance, is a way to enter in an agreement, originating only unilateral obligations and rights. When an agent gives the permission to another to do something that he has the power to prevent, there is an agreement between the two that enables the latter agent to do some action. In this kind of agreement, an agent becomes obliged not to interfere with the other one, who at the same time acquires the right to act as agreed upon. But given that no promise has been formulated, where does exactly such normative consequences come from?

An answer to this question is postponed to the next two sections because it is useful, first of all, to clarify what an agreement *primarily* is.

When there is an agreement between some agents, say Alice and Bob, the *consent* of at least one of them is necessary. When one consents, one is consenting *somebody to something*. Hence consenting creates a social relation between at least two agents.

But what one is consenting to? Sometimes Bob consents Alice to do something, like when he consents her to use his car. Other times, Bob consents to do something himself, like when he accepts to pick the children from school. Other times it happens both that Bob consents Alice to use his car and that he gives her the keys. In all these situations, by consenting an agreement is established. In any case, consenting is related to the fulfilment of another agent's desire *that one can interfere with*. This desire might be to do something that one may impede (negative interference). Other times, the desire is that one does something to favour another one (positive interference). Or a combination of both at the same time, like when Alice's desire to do something depends on Bob creating some favourable conditions, something which she also desires. What is common to all these cases, it that an agent has the power to interfere somehow with another agent's desire, and when the former consents that the latter fulfils such desire, it is entailed that the former does not interfere negatively

[7] [26], see also [15]; [9] for a critique of this view.

7

with it or that he interferes positively with it. For simplicity, from here on, we will just mention the negative interference situation.

An agreement then creates a social relationship between the parties, and presupposes a pre-existing *asymmetrical* social relation of dependence between them. When there is an agreement at least one agent that could (has the power to) interfere, is not interfering.

However something stronger is needed to have a real agreement. Though it is true that a lion that is not hungry is consenting a gazelle to wander around him safely, the gazelle does not have the lion's consent to do so and there is no agreement between them to this purpose. The gazelle does better to be ready to run as soon as the lion manifests any change of mind; she may exploit such temporary loss of interest but not rely on the lion only because the lion does not have a desire to interfere with her. Differently, her reliance would be more justified if the lion could be able to communicate his decision (i.e. intention) not to interfere with the gazelle, that is, to express his *consent*. Hence, one's consent (not just a behaviour that happens to consent) to the fulfilment of a desire of another agent is there when *one has the intention not to interfere with such desire fulfilment*. To be able to formulate such an intention one obviously is to be able and in condition to interfere, thus this condition presupposes the truth of the former.

This unilateral consent, however, is still not enough. Suppose Alice and Bob live together, and Bob has bought a car. Though the car is legally owned by him, between them, Alice may not 'acknowledge' it as Bob's because she does not consider the matter of who uses it as entirely up to him; she does not consider this choice as depending on him alone. She knows that Bob has the keys, and that he has some sort of legal power to interfere with her free use of it (she could be charged of theft, for instance). Alice also knows that she has Bob's consent to using the car whenever she wanted to, but still she contests this power over her. In this case, though all the conditions above might be true (i.e. Alice objectively depends on Bob, and Bob has decided not to interfere with Alice), there is still no agreement between them. It might be said, that Alice uses the car despite the fact that Bob can (he has the power to) interfere with her. To have a full agreement, then, there must be also *an acknowledgement of the power of interference of the agent who, in fact, has such power*.

Suppose now this variation of the example. Though the car is legally Bob's, it is Bob that reject his own power over Alice, as far the use of the car is concerned. Alice may consider the matter of who uses the car as up to Bob, but Bob himself contests this fact. If Alice asks his permission to take the car, Bob replies that she does not have to ask for it, that it is her choice whether to take it or not. In this situation, again, there is no agreement between them that Alice uses the car, because, between them, though Bob has the power, he does not 'value' it.

What is, then, to value a power, and what relation does it bear with power acknowledgement?

Valuing one's power is not simply desiring to exercise it because it may happen that Bob can indeed desire to use it against Alice in a moment of sudden anger but then hates himself for such desire given that Bob in fact contests such an asymmetrical relation between them. Bob does not desire to desire in this way towards Alice, that is, he does not value his power over her. One values one's power when one *desires that the use of one's power is motivated by one's desire to do so*, that is, when

one desires that it depends on one's choice whether to interfere or not. Hence, in the latter example, despite the fact that Bob can in fact interfere with Alice and he could come to desire to exercise his power over her, he does not want to be so motivated, at least when it comes to have a power over her. This example makes clear that, to have an agreement, *the agent values the power he has over the other one*, the agent values the fact the he is able and in condition to interfere with the other[8].

Power acknowledgement, differently, is the *acceptance* of such power, that is, the decision to forbear to resist to the exercise of the power over oneself if the other wants to exercise it. Acknowledging the power of another makes manifest one's fundamental non-hostility towards the other: to be prepared not to pursue something if this happens to be against the other's desire. While in giving one's consent, one accepts something one can interfere with (i.e. intends not to use a power of interference one has), power acknowledgement, differently, is just the acceptance of the use of such power, that is, the intention not to resist to the decision of the other agent. Both *valuing one's power of interference* and *the acknowledgement of another agent's power of interference* are necessary conditions to enter in an agreement.

Consider now this example. Alice wants to use Bob's car tomorrow, she has his consent, and she acknowledges his power on this matter. Notwithstanding so, to be safe in case something happens, Alice books a taxi for tomorrow morning. Suppose that Alice is quite sure that, intending not to interfere with her use of the car, he will so behave. But still she is worried that something unexpected might turn out. Assuming a worst-case scenario (which she considers highly improbable anyway), Alice decides not to rely on Bob. It seems that, in this situation, if Alice does not *uptake* Bob's consent, no agreement between them has been established. And *uptake* precisely *is such reliance on one's part on another agent's intention not to interfere with one's desire fulfilment.*

Finally, even if such condition is needed, it is not in itself sufficient for creating an agreement. In fact, Bob may know that she has a very important meeting tomorrow and that she needs the car. To avoid creating any obstacle, Bob decides to refrain from taking the car in the morning but he does this without that she realizes this intention of not interference, hence she does not uptake his consent though she would in case she knew about it. While Bob's intention of not interference is present, her ignorance of such intention makes it the case that they have no agreement that Alice uses the car today, and she may decide to call the taxi. Knowing that another agent has the power of interfering with oneself, and knowing that the other intends not to exercise such power is needed to have an agreement. But, as it is standard in many social interactions, even such first-order knowledge isn't enough to have an agreement because Alice may know this fact while Bob does not know that she knows it and, on this basis, Bob may think she will act otherwise and so in the end deciding to pick the car on the assumption that Alice may have decided to call a taxi, and so on for all the levels. In any agreement, then, an epistemic condition is necessary, that is, there should be *common knowledge* of the intention to not interfere. The same reasoning supports also other epistemic conditions. An agreement, in fact, cannot be in place unless the agent, who is consenting to the other's desire fulfilment, knows about such desire in the first place. And again this fact must be out in the open by being common

[8] On valuing and second-order desires see [7] and [20]; for a critique see [31].

9

knowledge that an agent has the power to interfere with a desire of another one who in fact has such desire. The acknowledgment of such power made by the other agent also must be matter of common knowledge, given that an agreement basically is a way to obtain something one wants without coercing the other to do so. And, finally, both the valuing of one's power and that the uptake of the consent are again matter of common knowledge between the agents.

Let's take stock.

A social relationship between at least two agents is an *agreement* between them if and only if the following five conditions hold:

1. The agent having the power of interference intends (for some reason) not to interfere with the other agent's desire fulfilment (consent condition);
2. The agent having the power of interference values his own power (valuing one's power condition)
3. The agent, who is subject to interference, acknowledges the power of the other one, that is, he intends to refrain from pursuing his desire if the other desire that he so behaves (no coercion condition);
4. The agent, who is subject to interference, relies on the consent of the other one, that is, intends to pursue his desire on the assumption that the other one intends not to interfere (uptake condition);
5. All conditions above are common knowledge.

An agreement of this sort may be called unconditional, in the sense that one does not give one's consent on condition of another agent's consent. Differently, an exchange of conditional promises gives rise to a conditional agreement in which each consent is conditioned on the other. Contracts, for instance, are instances of conditional agreements.

Moreover, on this analysis, it is also evident that there can be agreements without promises. Agreements are particular kinds of social relations between the agents, and a promise is one possible way to establish such relations (see also Section 6). Other possibilities, such as a mere exchange of a request and an acceptance or a mere unilateral permission without any request, make it clear that no promise is indeed necessary.

5 The principle of reliability

All agreements have normative consequences, even those that are unconditional and established without the interposition of a promise. However, on the present analysis, an agreement is primarily a social relation characterized by specific motivational and epistemic conditions that are true of the agents entering into it, and so no normative relation has been so far mentioned. How, on this account, is it possible to explain the 'obligation' of the consenting agent, and the corresponding 'right' to do or to obtain what an agent has been consented to? Or, differently put, what is the wrong of infringing an agreement?

In our view, the wrong of violating an agreement not made through promises is of the same family of the wrong one would commit if the agreement were promise-based. Both situations, in fact, pertain to a more general kind of social interactions that are wrong in relation to "what we owe to each other when we have led them to form expectations about our future conduct" [27].

The moral Principle of Fidelity put forward by Thomas Scanlon in this seminal paper was intended to account for the wrong of breaking a promise, and, as such, may be too strong for the kinds of agreement without promises we are after. However Scanlon has also insisted on several moral principles bearing family resemblances with each other given that all are related to the elicitation of expectations in others. To account for the normativity of agreements without promises the so-called principle of Loss Prevention could be enough [27]. This principle requires that *one that has intentionally or negligently led someone to expect that one will follow a certain course of action, and has reason to believe that that person will suffer significant loss as a result of this expectation if one does not fulfil it, must take reasonable steps to prevent that loss, that is, he ought to warn, fulfil the expectation or compensate*.

The fact that the principle is not just to prevent another agent's desires frustration but losses, indicates that some form of reliance is presupposed for the principle to be applicable. Suppose, in fact, that Bob had, somehow, led Alice to expect that he won't take the car tomorrow morning, say because he knows that she heard him accepting a lift from a colleague on the phone. Bob knows that she cares about this fact given that she needs the car. Still Alice decides not to rely on Bob as for having the car at her disposal tomorrow, and, to be completely safe, she books a taxi. Knowing this, Bob is under no obligation towards Alice, not even to warn her that in the end he will take the car. Though taking the car might be something she desires more than just taking whatever means of transportation, the frustration of this desire of her is not a loss Alice incurs with, it is not something she has and she wants which she is deprived of, hence the principle of Loss Prevention does not apply. Under this respect, even in the case that she had relied upon Bob and decided not to call the taxi, the very fact that the desire is frustrated when Bob instead took the car is not a real loss[9]. However in relying on Bob, Alice has in fact lost something she had before her. She has paid some costs, opportunity costs as the economists call them, which are the available alternatives of actions she had and which she has renounced to pursue by counting upon Bob's car being available.

At least for the aim of this article, then, the way Judith Thomson has defended a similar principle seems better suited [29]. Thomson, in fact, argues for the validity of a Word-Giving Thesis in which, when an agent invites another one to rely on the truth of a certain proposition, which invitation the latter agent accepts (or uptakes), then the latter agent acquires a claim (i.e. a right) against the former one to its being true. This way of formulating the moral principle bears two main advantages over Scanlon's: firstly, it makes explicit the relevance of reliance or uptake in the process, and, secondly, it generalize it towards whatever proposition one may rely upon beside those that refer to an action one will do in the future.

[9] Though it can be so when I consider the desire to have the car not as something I am to achieve but as something already achieved and to be protected, see for this possibility and its psychological plausibility [22].

11

However, though one can induce reliance, one can *allow* reliance as well and in such a way to have normative consequences.

Consider again the example above: Alice heard Bob's conversation with somebody else and, as a consequence, she comes to believe that Bob will not take the car and she relies on it. In this case, Bob has unintentionally induced in Alice some kind of reliance. We have suggested that by acting on these expectations about Bob, she will incur in some losses, and so the principle of Loss Prevention might apply. But is it so? After all, such induced reliance in this case is not intentional; can Bob be responsible for Alice's unilateral decision to rely upon him in this situation? It seems correct to say that though her reliance has been only involuntary induced, at least Bob has *allowed* her to rely on him. More precisely, in fact, *to allow a belief or an action is to have the power to disconfirm another's belief (which is a reason to believe something else or to act in some way) and to forbear to disconfirm it*. If hearing what Bob has said on the phone is a reason for Alice to believe that he will not take the car tomorrow, then this belief is obviously something that Bob can disconfirm. By not disconfirming such belief, Bob is also allowing her to believe in this way. Granted this, as such this form of allowing is still not sufficient for an agent to acquire a claim against another one. Suppose in fact that, immediately after having realized her reliance, Bob tells her that what is true is just that he does not confirm that he will take the car (which is the same of not disconfirming the belief that Bob will not take it) and nothing more than that. Can Alice hold him responsible for her losses if in the end he decides to take the car despite her unilateral reliance? It seems not. Suppose differently that just after his conversation on the phone and knowing that she needs the car, Bob turns to Alice and say 'yes, you heard correctly. I won't take it!'. By confirming a belief that he has unintentionally induced in her, Bob then become obliged towards her to warn in case he changed your mind, or, if it's too late, to do as expected or to compensate. Because such confirmation of the belief logically entails the absence of a disconfirmation, even in this case Bob has allowed her to believe something, though not passively (i.e. by forbearing to disconfirm it) but actively (i.e. by confirming it). It is this form of 'active' allowing that is necessary for the moral principle to apply when one does not induce intentionally reliance in others[10].

Finally, there are also cases in which one actively allows other agents' reliance on oneself *that one has not in any way induced*.

Suppose for example that Alice believes that Bob will not take the car tomorrow because John told her so, and that she relies on him for having the car. Bob knows about all this and he allows her to believe it (i.e. he forbears to disconfirm such belief). If she just act on this basis, and she does not know that Bob knows about her reliance, it seems that at most Bob should warn her if the belief is false, but if this is so, it is just out of sheer altruism[11]. If, differently, Bob has confirmed this expectation of

[10] Scanlon's principle of Loss Prevention indeed mentions also leading expectations negligently, besides doing it intentionally. However, negligence implies having not paid *due* care to avoid such reliance, and so it cannot be evoked to explain, without circularity, a principle which normatively demands such behaviour. Differently, our notion of active or confirmatory allowing has not such problem.

[11] The reason why common knowledge of another's forbearance to disconfirm one's beliefs may change the situation will be discussed in the Section 7.

her, for instance by nodding, Bob has actively allowed her to rely on him, and, from there on, he is responsible for her possible losses even if he has not induced that belief in the first place. Again, when it's too late for warning, Bob ought to fulfil the expectation or compensate.

To sum up, according to the view adopted in this paper, an agreement has normative consequences because the agent consenting another one to fulfil his desire is *either intentionally inducing or actively allowing uptake* on the consent is concerned (i.e. reliance that the former one intends not to interfere with such desire fulfilment), and, by doing so, undertakes a *duty of reliability* against the other one and creates a corresponding *right to rely*. Reliability is normatively required to prevent losses caused by intentionally inducing or actively allowing such reliance. One way in which such principle can be explicitly formulated is the following: *if one intentionally induces or actively allows another agent to rely on the truth of a certain proposition, then the latter one acquires a right to reliability (i.e. to be warned if the proposition turns to be false, or, in case the proposition is about the future action of the former one and it is too late for warning, a right that the former one acts so as to make the proposition true or to be compensated for the incurred losses)*. For these reasons, we name such a principle: the *principle of Reliability*.

6 The normativity of agreements and the value of not having hostile attitudes

Thus, by establishing an agreement between the agents at least one of them intentionally induces or actively allows the uptake of the other one. From this it follows that the uptaking agent acquires a right to rely on the other one. But what exactly he has a right to rely on?

We have suggested above that if there is an agreement between Alice and Bob that Alice will take his car tomorrow, his taking the car is doing something wrong. Under this perspective, by uptaking the consent, one acquires a right on a certain behaviour: i.e. that the other does not interfere with his desire fulfilment. But it seems even more than this: even if Bob does not take the car but afterwards he manifests some uneasiness because she has taken it, it seems again that Bob is doing something wrong. If agreements were there only to rule behaviours, what Bob has done should be enough for complying with its terms. However it looks like that it is not.

To understand why it is so, and what the peculiar normativity of an agreement is, consider the difference between the mere fact that some agents agree in their desires and the fact that there is a social relation of agreement between them.

When they agree in desires "the same world would satisfies the desires of both" [20] possibly without any social relation between them whatsoever. Differently, when there is an agreement there is also a social relation between them that aims precisely to create such agreement in desires *but facing the fact the things could have been different*. In fact, as we have noted above, an agreement presupposes an asymmetrical relation of power and dependence between the agents so that one can influence the fulfilment of the desires of the other.

13

However by acknowledging such power, one also signals one's basic *non-hostility*, that is, one's desire not to be motivated to do an action against the desire of an agent that has and values his power of interference. Correspondently, for whatever reason an agent decides to do so, by giving the consent, an agent also signals that the fulfilment of such desire 'agrees' with his own desire in the present conditions (i.e. the agent for some reason desires not interfere with the other). Thus, *an agreement results in the fact that the agents mutually know that their actual desires agree*: they are jointly co-realizable and they are so without any coercion.

Consider now the principle of Reliability. The similar principle of Loss Prevention is justified for Scanlon on a contractualist basis by the fact that "it is not unreasonable to refuse to grant others the freedom to ignore the losses caused by the expectation they intentionally or negligently lead others to form" [27]. One reason to refuse such freedom is readily available if *the agents share a value of not being motivated by hostile attitudes*[12]. In fact, ignoring such losses, when one has intentionally induced or actively confirmed an expectation on oneself, would be tantamount to be motivated by a hostile attitude: either one desires that the other incurs in those losses or at least lacks the desire that the other does not incur into them. Let's assume, then, that our agents share this value so that the principle of Reliability, as it has been here formulated, would just follow.

According to Lewis' dispositional theory [20], this would be a value *de se*, that is, *a property that the agents are disposed to desire to desire (i.e. to value) under ideal conditions*. The value of not being motivated by a hostile attitude amounts, then, to the fact that, if the agents are under ideal conditions, they are disposed to desire to desire to have such a property. Moreover, given that being motivated by a hostile attitude is being motivated to frustrate the desires of another agent, the compliance with such value requires them to revise their possible first-order hostile desires in a way that would inevitably result in the creation of harmony in the population, that is, in desires that agree.

Sharing the value however does not necessarily mean that the agents *will* behave according to what is required of them in the present conditions. It would of course in case they were in ideal ones, but no one is a saint, that is, no one lives always up to one's values. However, between agents that enter in a social interaction, such value can at least ground the *presupposition* that the other fellows will be so motivated, otherwise the best one can do is to avoid any possible contact with them.

What is then, on this basis, the peculiar normativity of agreements as social relations?

Recall that in giving one's consent, one induces or actively allows another agent's reliance on the consent, that is, not just on the observable behaviour of not interference but, more specifically, on the decision not to interfere. Moreover, given the details of the social interaction between the agents, it is also manifest that the decision is based on the fact that the desire of the consenting agent agrees with that of the other. It is on the decision based on this 'agreeing' desire that the other rely upon, if he wants to be non-hostile with the other. As a consequence, and given the principle of Reliability, those who uptake a consent acquire a *right to such decision of not inter-*

[12] Another reason would be available to them if they shared a value of assurance as Scanlon suggests [27].

ference based on an agreement in desires. Thus, the consenting agent that is willing to enter an agreement is not only obligated not to interfere (i.e. not to take the car himself), he is in fact bound *not to change his mind* otherwise the basic non hostile attitude of the agent would be frustrated: he is bound not to come to desire to interfere with the other. *When giving one's consent, one is obliged to keep one's desire in agreement with the other*. For this reason, it turns out that it is illegitimate even the expression of Bob's uneasiness with what Alice has done since such reaction on his part would signal that Bob has indeed changed his mind.

Is the other also bound similarly? We think so. In fact, the consent is given, and the decision is taken, on the assumption that the other agent has the desire in question (i.e. she wants to take the car): Bob relies on this fact and Alice has induced him to so rely. Hence Alice is bound too not to change her desire, on pain of being hostile with Bob, given that the opportunity costs he has paid to eventually decide not to interfere with her would then become just induced losses.

Hence, even in an unconditional agreement as this one, there are reciprocal obligations and reciprocal rights. *By establishing an agreement between them, the agents become reciprocally obligated and entitled to keep their mutually known desires in agreement*[13].

Does this entail hence that an exchange of promises has indeed occurred? No. By promising one creates the expectation that the promisor will do an action in the future *unless the other consent to not doing so* [27]. When giving one's consent without the interposition of a promise a timely warning can still be enough to release oneself from an obligation, at least when the other has not lost valuable alternatives to satisfy his desire. Agreements not based on promises are just weaker than agreements based on promises. They aim to create and protect desires that agree, and they do so for agents that share the value of not having hostile attitudes.

7 The ambiguity of silence and tacit confirmation

Now, suppose that it is common knowledge between Alice and Bob that Alice wants his car tomorrow morning, and that Alice believes that he will not take it because tomorrow is Monday, and on Mondays Bob never takes it (maybe just because it is his habit to act in this way or because the traffic on Monday mornings is more intense than in the other days and Bob hates to be stuck in traffic). Given that she believes that he has his own reasons for not taking the car, and she knows that he usually act in this way on Mondays, it is reasonable for her to expect Bob to behave in this way this Monday too (i.e. she believes with some probability that this will happen). Alice so believes this that she relies on him for not taking the car, and she decides to go to the meeting with his car. All above being common knowledge between them, she also observes that Bob has kept silent about the truth of this belief until Monday morning. However, just when the time has come, Bob decides to take the car, say because it

[13] More precisely, the obligation is to keep one's *first-order* desires in agreement. Such an obligation can be seen as a reason for all the parties to the agreement to have a second-order desire that their first-order desires keep motivating their behaviour. Those second-order desires would motivate the agents to do whatever they can to avoid to revise their first-order desires.

15

happens that today he needs the car for some unanticipated errands. Has Bob done something wrong?

It is foreseeable that having incurred in some losses, Alice may resent Bob's late decision, and she may even protest about such sudden change of mind. But, is she entitled to anything? Is Bob under a sort of duty towards her? In case she thought that way, Bob could legitimately claim to have not given her any consent to use the car, not even acted in order to make her believe something about him, that is, not even implicitly consenting her to something. So why would be Bob responsible for her losses? In the end, he has not intentionally induced any reliance on himself nor he said 'yes' or any other kind of confirmation because, by assumption, no communication between them has occurred.

Granted this, however something strange has indeed happened.

The closer she come to the fulfilment of her expectation, *the more she feel sure about such fulfilment and entitled towards the other acting as expected*. It is a fact that, though Bob knew about her belief, he kept silent until the moment has come, that is, that Bob has not disconfirmed her belief.

Suppose that Alice has interpreted this silent behaviour as a *confirmation* of her belief that Bob won't take the car, and then she has felt that such confirmation has somehow entitled her to have the car. But what kind of confirmation is this given that they do not communicate? Is it reasonable to read the other's silent behaviour in this way? And how can the omission of a disconfirmation create duties and rights?

To understand this issue more clearly, suppose that Alice is a Bayesian rational agent, that is, suppose that H_i is her hypothesis that Bob will not take the car tomorrow that is characterized by a subjective probability $p(H_i)$, representing her degree in belief in H_i. Because beliefs are represented by a well-defined additive probability function [27], her degree of *dis*belief in H_i is given by 1- $p(H_i)$. We can imagine such beliefs be warranted by inductive reasoning in which Alice has acknowledged that there is a pattern governing Bob's behaviour such that, almost on every Monday Bob does not take his car or, simply, that not taking the car on Monday is his best choice given his desire not be stuck in traffic.

Suppose that given Alice's concern on what Bob will do this Monday, she starts looking for additional evidences for her belief that he won't indeed take the car. Assuming, as we have done above, that everything is common knowledge between them, she happens to notice that Bob keeps silent about the truth of this belief she has about him, though he knows that she has decided to rely upon him.

The observation of silence, from a Bayesian perspective, can be treated as a 'datum' S for determining whether Alice's belief about Bob is true or false. Hence, by applying the Bayes' theorem, the belief can be updated accordingly. Moreover, such update of H_i must be determined relative to its complement $\neg H_i$, as the usual formula makes clear:

$$\frac{p(H \mid S)}{p(\neg H \mid S)} = \frac{p(S \mid H)}{p(S \mid \neg H)} \bullet \frac{p(H)}{p(\neg H)}$$

As a Bayesian rational agent, Alice is interested in the impact of the fact that Bob is silent on her belief that Bob will not take the car tomorrow, which amounts to calculating the probability that her belief is true, given that she has observed his silence. To do this, as a Bayesian rational agent, she needs to compute the *posterior* (i.e. the

odds that H_i is true in light of what is known after the observation of S) that equals the *likelihood ratio* (i.e. the second term from the right representing the information value of S with respect to the truth of H_i) multiplied for the *priors* that H_i and $\neg H_i$ are true before the observation of S. In such an inference, in case the probability of observing S when H_i is true differed from when is not true, the likelihood ratio would be different from 1, and the posterior would also differ. In particular, the datum (i.e. Bob's silence) favours the hypothesis H_i when the posterior odds are greater, and this happens when the conditional probability of his silence given that Alice's belief about him is true is larger than the conditional probability of Bob's silence given that her belief about Bob is false. In such a case, it is said that the observation of S is *diagnostic* of or *confirms* H_i and not $\neg H_i$.

Silence clearly is ambivalent evidence in that there are both reasons for believing that it supports Alice's belief about Bob (if Bob does not want to take the car, he does not inform Alice that he will instead take it) as well reasons to believe that it can *dis*-confirm my belief: it may be possible that Bob could not reach her in time or that he has forgotten her desire to have the car, or that he simply does not care about Alice enough to let her know something relevant for her, or that Bob wants to harm her on purpose and so on. Whether the evidence is relatively more confirmatory than not is a contingent matter, and depends on the ratio between the *known* conditional probabilities of observing silence on condition that my belief is true or false. If she is Bayesian rational agent, she compares these information values before updating her belief.

There are, however, (psychological) reasons to believe that Alice, as all of us, is not so rational.

It is in fact one of the "best known and most widely accepted notion of inferential error" [6] that human reasoning gives undue weight to evidence that supports one's beliefs while discounting evidence that would tell against it, and this tendency is called *confirmatory bias*[14]. A confirmatory bias can be discovered in many different situations in which one assesses the truthfulness of one's beliefs. However the scientific evidence is particularly vivid when one is both *concerned* in what one believes (the so-called motivated confirmation bias) and the evidence one is evaluating is *ambiguous* (i.e. it is partly supportive and partly not *without exactly knowing how much it is so*). In this kind of situations, there is a very strong tendency to interpret information in ways that are partial to one's beliefs, and in particular, in ways in which the positive side of the evidence is overemphasized.

On the basis of these empirical facts, it is seems plausible to assume that there is an analogously strong tendency *to read other's silence*, in the kind of situations we are interested in, *as a positive evidence for one's belief*. In fact, the ambivalence of silence would not be too much of a problem if silence were not often an *ambiguous* evidence in that one is not so sure on how to assess such ambivalence, whether the positive support to one's hypothesis is more likely than the negative one (Ellsberg 1961). In the case at hand, ambiguity about the evidential value of silence can be seen as a form of uncertainty about the relative conditional probabilities of $p(S|H_i)$ and $p(S|\neg H_i)$. The agent does not know what the likelihood ratio is because it is as if he considered as reasonable, in the present circumstances, more than one distribution of

[14] See [23] for a review of the relevant psychological literature; see [25] for a mathematical model, though focussed on a different aspect of the confirmatory bias.

17

conditional probabilities of observing silence, given that the hypothesis is true or false.

If we accept the confirmatory bias, it may be suggested that, in contexts where we already entertain the relevant belief, we update it by adopting the *best* expectation that could be associated with the observed evidence, which is the one that would *confirm* the belief already accepted. In other words, silence regarding one's belief, that is, the forbearance to disconfirm such belief means, for the agent holding the relevant belief, that the other one will act *as expected*.

To interpret silence in this way, one must think that the other is not hostile towards oneself, otherwise, if this were not the case, if he believed in the other's hostility, then the negative side of evidence would be maximally relevant. However, as we have assumed above, such non-hostility is a reasonable presupposition for agents that interact with each other. Under this presupposition of non-hostility, it is reasonable to consider that the 'natural' meaning of silence is confirmatory.

It is then understandable why the more Alice is close to fulfil her desire that Bob does not take the car, the more she is *sure* that he will not take it. Supposing that she has checked upon him several times until Monday morning arrives, each time Bob's silence has confirmed her belief possibly up to certainty.

So far so good for the expectation that Bob will not take the car becoming firmer (i.e. confirmed). But what about the fact that she also feels *entitled* that he does not take it?

First of all, given that the confirmatory meaning of silence is salient between them (Bob is a confirmatory agent just as Alice is) and he knows that he has not disconfirmed a belief she had about him, Alice has reasons to believe that Bob cannot but assent to her interpretation (at least from the perspective of bounded rationality): that Bob's silence means that Bob will act as expected is 'natural' or *salient* in this context (i.e. it is the obvious interpretation for confirmatory and non-hostile agents). If Bob has reasons to *assent* to Alice's belief, he has reasons to believe that it is reasonable to believe something in those circumstances and so he has reason to believe that he has as a matter of fact confirmed Alice's belief about him. If the salience of precedence suffices to justify the commonality of our beliefs in future conformity to a convention [15], the salience of silence might justify a mutual belief in the occurrence of confirmation.

One relevant consequence of such common knowledge is that, though at the beginning Bob were just 'passively' allowing Alice to believe something about him, under these conditions of common knowledge of his confirmation, the allowing becomes 'active'.

Moreover, given what we have discussed in Section 5, this is sufficient for the Principle of Reliability to apply, giving rise to Bob's duty of reliability and to Alice's corresponding right to rely. And from this it follows that her possible protest or resentment cannot but be *entitled* simply because she has a *right* that he does as expected, that is, that he is reliable.

8 Tacit agreements: when the agreement is implicated

Even if agreements can be established without promises, usually other kinds of speech acts are employed to create the required epistemic conditions behind them. For instance, for an agent to consent another one to something that is desired, the former needs to know about such desire in the first place. Usually, the latter communicates the desire simply by informing, or by formulating a request or, sometimes, by proposing an exchange, and so aiming to offer a reason to motivate the former acceptance. A conditional promise is first of all a way to influence such acceptance by offering some incentives. Similarly, the consent must be mutually known between the parties, and, to this end, one's the intention not to interfere with the other's desire fulfilment is usually communicated. This is often done through explicit communication, that is, by conventionally signalling one's agreement through nodding or using verbal communication.

However, I can inform you about my desire just by taking the keys of your car, knowing that you are looking at me, and that you will infer the desire behind my behaviour. Analogously, by acting in order to remove an obstacle for me or by avoiding creating one, you can communicate with me without language, gestures or other conventional means. In fact, practical actions (or forbearances) done with a communicative intention (i.e. practical actions done also because another agent while 'reading' such behaviour will believe something) might suffice to send a message. Elsewhere, we have argued for the importance of this kind of communication that we name 'behavioural implicit communication' [2] [30]. Here, we just confine ourselves to suggest that this form of communication through practical actions and their effects might support the creation of agreements that can be dubbed, for this reason, *implicit agreements.* When there is an implicit agreement between some agents, the one having the power to interfere with the other can implicitly give his consent by acting with the intention to refrain from interfering, knowing that the other understands what is happening. Those that are qualified as 'tacit' are often instances of agreements established, silently, via implicit communication.

Notwithstanding so, if there are cases in which it is already common knowledge between the agents that one of them wants something, even implicit communication may be useless; similarly for the consent, the uptake, and all the conditions that need to be commonly known for an agreement to be in place.

But how is it possible that all these epistemic conditions be satisfied, without either promises or any other kind of communication between the parties? Or, in other words, *how is it possible to have agreements without communication?*

Recall the necessary and sufficient conditions to have an agreement discussed in Section 4. One prominent clause is the so-called 'consent condition'. In the way it has been formulated, such condition does not require any communication. In fact, having another agent's consent just entails that the agent with the power to interfere, indeed, intends not to interfere. However, often, one does not only consent to something, but one also *gives one's consent*, which necessarily is the communication of such decision of non-interference, via the usual Gricean mechanism [11]. One can give one's consent without verbal or gestural communication, but at least implicit communica-

19

tion is necessary. However, though one cannot be *given* the consent without communication, one can *have* the *tacit consent* without any communication.

Consider again the example discussed in the previous section.

It has been shown that when the parties consider silence as a confirmatory device for the beliefs on the truth of which one relies, the confirming agent becomes obliged to be reliable, even if no communication has occurred between them. In the example, Bob become obliged not to take the car tomorrow, given his tacit confirmation of a belief Alice had about him. However the mere fact of not taking the car, and as a consequence of not interfering with her is not in itself sufficient for Alice to have his consent. According to the analysis developed in Section 4, if one has a consent then the other agent has the intention not to interfere with him, that is, the consent implies that *the content of the intention refers to another agent*. Differently, the intention behind the behaviour that contingently happens not to create obstacles for another agent needs not be so. Indeed, in the example, the decision not to take the car on Monday is motivated either because that is Bob's habit on Mondays or because it is the best option he has to avoid being stuck in traffic.

However, as noted in Section 5, once the principle of Reliability applies, one incurs in a 'directed' obligation, rather than an unqualified one: Bob is *obliged towards* Alice not to take the car, and Alice has a *right against* Bob to this behaviour. Therefore, such a directed obligation is not simply to avoid taking the car, but, more precisely, to forbear to do what would, in this context, prevent her to fulfil her desire that Bob does not the car, which amounts to being obliged to not interfere with such desire fulfilment.

Granted this, is it true that Bob's silence also means that he intends not to interfere with Alice, i.e. that she has his consent?

Recall that Bob's silence is confirmatory of her belief about him under the presupposition of non-hostility; otherwise the disconfirmatory reading of the evidence would be maximally relevant. The presupposition that Bob desires not to be moved by a hostile attitude, however, amounts to assuming that the principle of Reliability is actually followed.

To see why it is so, consider Lewis' analysis of the kinematics of presuppositions in a conversation [19]. According to Lewis: "presuppositions evolves according to a rule of accommodation specifying that any presuppositions that are required by what is said straightway come into existence, *provided that nobody objects*" [19]. Though presuppositions are almost always approached in the contest of communication, the fact that social interaction, even tacit as in this case, may have the same properties and consequences of linguistic exchanges and proper conversations is explicitly endorsed by some pragmatists [14]. If a presupposition of reciprocal non-hostility, possibly grounded in a shared value of not being motivated by hostile attitudes, is reasonable, then what is required 'by what is done' when one is in a social interaction with another agent becomes immediately into existence. That is, it becomes common knowledge between the agents that both of them share a value *de se* not to be motivated by hostile attitudes. In the present context, Bob's violation of the principle of Reliability would amount to actively allowing that Alice incurred into losses, and he would be indeed hostile towards her. If we accept that there is such a presupposition of non-hostility in the background of this kind of interactions, then we are also accepting that there is a shared assumption between the agents that principle of Reliabil-

ity is indeed followed. Under these conditions, and given that Bob's silence confirmed the belief that he will not take the car, and that he is consequently obliged not to interfere with her, his silence *also* means that he intends not to interfere with Alice or that she has his consent that her desire is fulfilled. More precisely, *if in this context one's silence "naturally" means one's confirmation* [11], *it also "implicates" one's consent* [12]: that an agent intends not to interfere with another one or that the latter has the consent of the former is an "implicature" of such tacit confirmation because it is required that the former agent has such an intention in order to preserve the shared assumption that he is not hostile towards the other, or, which is the same, that he is not violating the principle of Reliability since this is something that the latter is assuming the other is not doing[15].

To sum up, given a shared assumption of non-hostility and thanks to the process of tacit confirmation, Alice knows that Bob also has a sufficient reason, a normative reason, for consenting her to something that she wants, that is, he desire that Bob does not take the car this Monday. Under the same assumption of non-hostility, which in this context amounts to the assumption that the principle of Reliability is followed, she also has reason to believe that Bob intends not to interfere with her since, by being silent, he implicates that she has his consent. Moreover, given that the assumption of non-hostility is shared by the agents, and that both the tacit confirmation and the normative consequences are common knowledge, it is also commonly known that Bob's silence means (implicates) his consent. It is this kind of consent that we consider a *tacit consent*, that is, consent without any communication between the parties, which is tacit in the sense that is *implicated* by something your are doing and from what is already commonly known and assumed by the agents. As a consequence that Alice has such tacit consent is also commonly known without having been manifested in any way, that is, without Bob giving it to her.

Let's now consider conditions 2 and 3: the valuing one's power and the no coercion conditions.

An agreement between them that Bob does not take the car entails also (1), that he desires that it is his desire to use or not to use the power over her to move him to act and (2) that Alice acknowledges this power over her as far as this issue is concerned, that is, she intends not to oppose Bob's decision to interfere with her desire fulfilment.

However, there has been no deliberation to consent her to something in the first place, and Bob's tacit consent is just implicated by something he did. So, how can such consent be compatible with Bob valuing his power?

This is the same objection put forward by Hume against Locke's famous justification of political authority. Hume in fact in his *Of the Original Contract* has resisted the claim that such authority is the product of a tacit consent whereby "the subjects have tacitly reserved the power of resisting their sovereign" on the account that, "an

[15] 'Implicatures', like presuppositions, are usually approached in the context of conversation, a situation in which we use language for common aims in a way that, as Grice has suggested, is governed by a Cooperative Principle. However Grice notoriously claimed also that the principle and the related maxims apply to cooperative contexts that are not communicative [12]. The relation between Grice's Cooperative Principle and the weaker principle of Reliability exceeds the scope of this contribution and are left for future research.

21

implied consent can only have place, where a man imagines, that the matter depends on his choice", that is, where a man imagines that by desiring to interfere, he would thereby have successfully exercised his power. Whether this is so in relation to political authority is not of our concern here, but still for an agreement to be in place such condition, or better, conditions 2 and 3 of an agreement must be met.

We have argued above that the consent is normatively required by the fact that Bob has actively allowed Alice's reliance. Even if it is required, this does not mean that the consent has been coerced or that no other alternative was indeed possible. In fact, *if he had not confirmed her belief, she would have accepted his decision to act in ways that interfered with her desire fulfilment.* The truth of this counterfactual, together with the fact that Bob has indeed confirmed her belief about him are also sufficient to guarantee that, though she does acknowledge his power over her in this context, she is now entitled to fulfil her desire. But how can the agents mutually know that such a counterfactual is true of them?

Simply because the shared assumption of non-hostility requires it too. Suppose in fact that Bob thought differently. Bob imagines that even in case he hastened to disconfirm Alice's belief, she would have pursued her desire in any case. This belief is incompatible with the truth of proposition that Alice values non-hostility as much as Bob does. Given that there was indeed an alternative to what has happened (Bob could have disconfirmed her belief but he didn't) Bob has to assume, if the shared assumption is to be considered true, that she would have behaved in non-hostile way. Hence, both conditions 2 and 3 are also satisfied, or better implicated, by what it is already common knowledge between them.

Moreover since both the fact that he is moved by a desire not to interfere with her and that she acknowledges his power are implicated on the background of what they already commonly know, both conditions are common knowledge, or at least potentially so.

Finally, for the social relation between the agents to qualify as an agreement, as already argued, the agent having the consent needs to uptake it (condition 4) and this fact must be common knowledge between the parties.

At first glance it may seem that this condition is already established because, in the example, Alice is in fact already relying on Bob not taking the car tomorrow. However, the uptake of an agreement is not just reliance on another's behaviour that happens not to interfere with one's desire but is, more specifically, reliance on the other's *intention* not to interfere with such desire fulfilment (see Section 4); to have an agreement one does not merely rely on another's behaviour, one relies on an intention, that is, one uptakes a consent.

Since however, in the example, condition 1 is satisfied, Alice also has the opportunity to rely on his intention not to interfere with her desire fulfilment, and not simply on his observable behaviour. But how can such uptake on her part be common knowledge between them?

Suppose that she does not in fact uptake the tacit consent. She can do this for, at least, two very distinct reasons[16]. She can consider that he is not trustworthy enough, in the sense that, though he now desires not to interfere with her, she believes that he will indeed change his mind on this issue. Differently, despite the fact that Alice be-

[16] We thank Maria Miceli for clarifying the relevance of this distinction.

lieves in Bob's trustworthiness, she simply does not want to take his car anymore: it is Alice who has changed her mind. Both state of affairs are however incompatible with the shared assumption of non-hostility. Let's consider the latter first. If eventually Alice does not desire to take his car anymore, then, since he has decided not to interfere with her, Bob will in incur into losses (i.e. the opportunity costs Bob has already paid) given that he is relying on the fact that she has this desire. In fact, just as Bob's silence, her silence too is a continuing confirmation of a belief of his: the expectation that she still desires something from him. Thus, she has also actively allowed him to rely on something and, as a consequence, he has now acquired a right to the truth of this proposition, for the same reasons discussed above. If it is now too late for a warning, either she ought to compensate for the losses or she ought to fulfil his expectation, that is, Alice has to keep her desire in agreement with Bob's. Thus, her silence, like his, has in this context a natural or salient meaning: it means a confirmation that Alice still desires what Bob expects her to desire. On the other hand, given that both agents are presupposed to value non-hostility, Alice possible distrust in Bob is incompatible with his actual being non-hostile because by believing that he *will* change his mind, she would also believe that he *will* be hostile with her. And this is something that is ruled out by our shared assumption, or at least, it is something that is to be considered as false in order not to violate it. As a consequence, if Alice's silence naturally means that she still desires what he expects her to desire, and having common knowledge of the tacit consent, then Alice's silence means also, or better implicates, that she relies on his consent. This is what is implicated in order not to violate the shared assumption of non-hostility. Because this fact follows from something we already commonly know and assumed, it is again something that we commonly know.

Let's take stock. Though agreements are very often based on communication, there is a kind of agreement that is not based on any form of communication, not even implicit. It is for this kind that we reserve the name of *tacit agreement*. Crucial for the establishment of tacit agreements is the fact that *there is a salient interpretation for one's silence when it is common knowledge that an agent reasonably expects and wants something from another one or has a right to obtain*. It is due to the salience of silence as a confirmatory device that we tacitly, and often involuntarily, become obliged to be reliable. To account for such normativity the *prima facie* plausibility of a principle of Reliability has been invoked. Under a presupposition that the agents share a value *de se* of not being moved by hostile attitudes, there is also an assumption that the principle of Reliability is actually followed. As a consequence a tacit confirmation also means one's tacit consent, or better, it 'implicates' such consent. Though implicated, such consent is not however coerced because it is also implicated that things could have been different, and this counterfactual possibility is matter of common knowledge. Finally, once an agent has another's consent, it is again the salience of silence that guarantees that the last condition for an agreement is satisfied, that is, those who have the tacit consent tacitly confirm that they keep their desires in agreement and, on this basis, implicate their uptake. Tacit agreements are agreements without communication, and are established necessarily by the tacit confirmation of the involved parties. Tacit agreements are potential agreements in the sense that there are reasons to believe that all the conditions for an agreement are fulfilled and this fact is accessible to the parties, at least if they bothered to think hard enough. Tacit

23

agreements remain potential as long as everything goes smoothly, that is, for example, if the agent who is in fact tacitly consenting, also acts as expected for whatever reason. They become actualized and operative agreements when one, willing to act against what the tacit agreement mandates, cannot but acknowledge that the consent, the uptakes and all the other conditions do in fact hold, that is, cannot but assent that a real agreement is in place. Finally tacit agreements, as all agreements, create reciprocal obligations and rights in the parties entering into them to keep their desires in agreement, that is, after an agreement is in place no unilateral change of mind is legitimate anymore.

9 Conventions are tacit unconditional agreements

If, following Hume's suggestion, conventions are agreements, and given that conventions persist without the need of communication, they are agreements without communication, that is, tacit agreements.

Consider a convention to drive on the right sustained by an interest in avoiding collisions.

As we have proposed in Section 3, conventions are regularities of reciprocal trust, hence, in the example, agents in the population regularly rely on the others to drive on the right: everyone assumes that the other will drive on the right and acts accordingly, that is, he himself drives on the right. Given that a convention presupposes an agreement in desire for some ends (our agreement in desiring not to collide), the expectation of reciprocal reliance is a reason for everyone to rely on each other so that, in this way, also our desire for the means (each desire to drive on the right in order to avoid collisions) are in agreement too.

Trust, as we have suggested in Section 3, is a fundamental non-hostile attitude on the part of the trustor: an agent relies on another to do an action that stems from his motivation, without any coercion. The reason why each relies on the others when they are parties of a convention is that each one expects the others to rely on oneself in the same, non-hostile, way. Moreover in order to trust everyone has to assume such non-hostile attitudes in the trustees. Suppose, then, as we have done in Section 6 that the agents in the population share a value of not being motivated by hostile attitudes, something that, of course, would promote the disposition to trust each other. Suppose also, as we have done in Section 7, that the agents have a bias for confirmation.

Under these two assumptions, and given that a convention exists in a population, each time two or more agents interact with each other in a situation that is governed by the convention, if they keep silent about the expectation of reciprocal reliance that they mutually know to have, each of them confirms their reasonable expectations about each other, *even if their mutual expectations of reciprocal reliance are not grounded in direct experience*; the agents might have never met before. By being confirmatory, each actively allows reliance on the truth of such expectation of reciprocal reliance. As a consequence, each also acquires both a *right* that the other rely on oneself, and an *obligation towards the other* to rely on the other one. Each agent has now a right that the other drives on the right (i.e. has a right to be trusted) and an obligation to drive on the right himself (i.e. ought to trust the other one).

Moreover, for the same reasons discussed in Section 8, on the basis of a presupposition of reciprocal non-hostility, each silence also "implicates" each consent, that is, that each intends not to interfere with the desire fulfilment of the other one. Given that in a convention, all the agents desire conformity of all the others, the tacit consent is the decision not to interfere with this desire of one's own conformity. And since conformity of others to a convention amounts to that fact that the others do rely on oneself, in the example, one's silence implicates one's tacit consent to all the others that one has decided not to interfere with their desires to rely on them. In a convention, each also tacitly consents to trust the others.

Moreover, in any convention it is the individual interest of each agent to conform, that is, everyone trusts the others because it is in the interest of everyone not to collide with the others, and so to rely on the others by driving on the right. Everyone's desire for the means stems from everyone's motivation not to collide. This very basic capacity (or power) of instrumental rationality is something that everyone values and everyone acknowledges to the others. If one had known that was not in the interest of the others to drive on the right, that is, to rely on oneself, one would have acted accordingly. This much is granted both by the fact that the agents are in a coordination problem [15], and in order to preserve our presupposition of non-hostility.

Finally, each uptakes such tacit consent of the other by tacitly confirming, firstly, that others' trust on oneself is still something one desires, and, secondly, by implicating that one does rely on such trust on oneself of the others and will act accordingly. That the uptake holds is required again by the presupposition of non-hostility, and has the consequence that each does not only trust the others, but also rely on the trust of the others on oneself.

Each time the agents, ignorant of each other's identities as they may be, do meet and keep silent about each other mutual expectations of reciprocal reliance establish or implicate a *tacit agreement to trust each other*. Since the tacit agreement is implicated by one's own silence *both as a trustor and as a trustee* the agreement is reciprocal: there is a tacit agreement between the interacting agents the both trust and are trusted by the other one. The tacit agreement is unconditional because the tacit consent are not conditioned one on the other; differently they are implicated by the presupposition that the agents are non-hostile, or, in the specific context, that the principle of Reliability is followed. Finally, the normativity of conventions is that of the tacit or implicated agreement: by tacit agreeing to trust each other *everyone is obliged to keep one's desires for the means in agreement with the other and has a right that the others do the same*.

10 Why conventions are tacit agreements

A regularity is a convention for the way it persists, not for its origins. In convention, one conforms if the others conform because it is in one's interest to conform. Since the stability of conventions is guaranteed by this specific motivational structure (i.e. their pre-existing agreement in desiring some end) together with common knowledge of all the conditions specified in Section 2, individual instrumental rationality alone suffices to stabilize it. Then, why should a convention be also a tacit agreement? Isn't

25

is only just an additional pressure that is made redundant by the reasons the agents already have for acting as they do? What is the role of obligations and rights in conventions?

Though it is true that conventions are stable for these reasons, the fundamental condition that ensures stability is that the agents agree in desiring jointly co-realizable ends. But what is there to guarantee that they will keep doing so? After all a common interest needs not be some ultimate end that we will invariably pursue forever. The ends we agree in desiring are often just means for some further ends we have. All instrumental desires cease to be motivationally effective, once the end in light of which we pursue the means has been either fulfilled or abandoned. Suppose Alice and Bob have a common desire to meet each other one day during the week and they fulfil their desires following the convention to go at the movie together every Wednesday. Suppose also that Bob is secretly in love with Alice, and hopes that by recurrently meeting him she will fall in love too. Differently, for Alice, Bob is just a friend that she is keen to meet, and nothing more. This Wednesday, at the end, Bob realizes how desperate his situation is, how impossible it is that his love will be ever reciprocated, and he abandons his plan to seduce Alice altogether. If he suddenly revised his recurrent end to meet with Alice, there would no motive at all to still pursue the means of going to the movie with Alice that night. Still however, by not showing up, Bob would do something wrong and against Alice, something that, notwithstanding his feelings, he may wish to desire not to be moved to do.

In other words, since all the parties to a convention conform (trust) on the assumption of the trust of others, agents need protection and assurance against the mutability of interest that might compromise each individual project. Since the kind of common interest presupposed by a convention may be as volatile as any other end we pursue, everyone would be at risk if everyone were free to change one's mind without taking into account the other in any way. Obligations act as further assurance in case one was to change his desires by entitling possible influencing actions (e.g. punishment by reproach), which can motivate the others beside their current desires.

Conventions tend to reproduce agreement in desiring arbitrary means from agreement in desires for the ends. However, by also being sources of tacit agreements between the agents, the arbitrary means are turned into ends to be pursued unless one is able to warn the other in time or is prepared to compensate for possible losses.

11 Conclusion

In his paper on causation, Lewis noted that Hume has defined a causal succession "twice over" [16][17]. The aim of this article is to suggest that something similar has occurred when Hume defined a convention as: "a general sense of common interest, which sense all the members of society express to one another, and which induces them to regulate their conduct by certain rules. […] When this common sense of interest is mutually expressed, and is known to both, it produces a suitable resolution

[17] Hume defines a causal succession both as a succession that institutes a regularity and by way of a counterfactual analysis. The two notions are to be kept separated, see Lewis (1973).

and behaviour. And this may properly enough be called a *convention or agreement betwixt us, though without the interposition of a promise*; since the actions of each of us have a reference to those of the other, and are performed upon the supposition, that something is to be performed on the other part" (Hume, *A Treatise of Human Nature*, III.ii.2, emphasis added).

That convention can be seen as tacit agreements is often suggested, and is considered as tantamount to the analysis offered by Lewis. However, what Lewis has shown is that, in certain conditions, an agreement in desires for the means might stem from our independent agreement in desires for the ends. However an agreement in desires is not the same as an agreement between the agents in that the latter, but not the former, is a social relationship between the agents. The fact the there is an agreement between the agents entails that their relationship is also a normative relationship. Whereas their mere agreement in desires may not have such consequences.

In this paper we have shown that the normativity of conventions is the normativity of tacit agreements, that is, that the agent becomes bound to keep their desires for the means in agreement, and by becoming so bound they are assures the other will not change their minds without some concern for their fellows.

The agreements that stem from conventions are tacit in the sense that they are implicated by what the agents do (or forbear to do) though without any communication between them is necessary. In order for this be possible we have offered two substantial hypotheses: (1) that there is a salient interpretation, in some contexts, of everyone's silence as confirmatory of the others' expectations, and (2) that the agents share a value of not being motivated by hostile attitudes, ad, on this basis that their interaction are regulated by a presupposition that the principle of Reliability is followed. If the former hypothesis is compatible with many available empirical data about human decision-making (Section 7), the latter is matter of future research.

References

1. Bacharach, M. & Gambetta, D. (2000) Trust in signs. In Karen Cook (Ed) *Trust and Social Structure*. New York: Russell Sage Foundation.
2. Castelfranchi. C. (2006) From conversation to interaction via behavioral communication. In S. Bagnara and G. Crampton-Smith (Eds.*) Theories and Practice in Interaction Design*, pp. 157-179. New Jersey (USA): Erlbaum.
3. Castelfranchi, C. & Falcone, R. (in press) *Trust Theory: Structures, Processes, Dynamics*. Wiley & Sons.
4. Conte, R. & Castelfranchi, C. (1995) *Cognitive and Social Action*. London: UCL Press.
5. Ellsberg, D. (1961) Risk, ambiguity, and the savage axioms, *The Quarterly Journal of Economics*, 75(4), pp. 643-669.
6. Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*.. Hillsdale, NJ: Erlbaum.
7. Frankfurt, H. (1971) Freedom of the Will and the Concept of a Person, *Journal of Philosophy*, 68, pp. 5-20
8. Gilbert, M. (1983) *On Social Facts*. London and New York: Routledge.

27

9. Gilbert, M. (1993) Is an agreement an exchange of promises?, *The Journal of Philosophy*, 54 (12), pp. 627-649.
10. Gilbert, M. (2008) Social convention revisited. *Topoi*, 27(1-2).
11. Grice, P. (1957) Meaning, *Philosophical Review*, 66, pp. 377-388.
12. Grice, P. (1989) *Studies in the Ways of Words*. Cambridge (MA): Harvard University Press.
13. Holton, R. (1994) Deciding to trust, coming to believe, *Australasian Journal of Philosophy*, pp. 63-76.
14. Levinson, S.C. (1995) Interactional biases in human thinking. In E. Goody (Ed.) *Social Intelligence and Interaction*, pp. 221-260. Cambridge: Cambridge University Press.
15. Lewis, D. (1969) *Convention: A Philosophical Study*. Cambridge (MA): Harvard University Press.
16. Lewis, D. (1973) Causation, *Journal of Philosophy*, 70, pp 556-67.
17. Lewis, D. (1975) Languages and Language. In K. Gunderstone (Ed.) *Minnesota Studies in the Philosophy of Science*, vol. VII, University of Minnesota Press (reprinted in his Philosophical Papers, volume 1, pp. 163-188).
18. Lewis, D. (1979a) Attitudes de dicto and de se, *The Philosophical Review*, 88(4), pp. 513-543.
19. Lewis, D. (1979b) Scorekeeping in a language game, *Journal of Philosophical Logic*, 8, pp. 339-359.
20. Lewis, D. (1989) Dispositional Theories of Value, *Proceedings of the Aristotelian Society, Supplementary Volumes*, 63, pp. 113-137.
21. Marmor, A. (1996) On convention, *Synthese*, 107, pp. 349-371.
22. Miceli, M. & Castelfranchi, C. (2002) The mind and the future. The (negative) power of expectations, *Theory & Psychology*, 12(3), pp. 335-366.
23. Nickerson, R.S. (1998) Confirmation bias: a ubiquitous phenomenon in many guises, *Review of General Psychology*, 2(2), 175-220.
24. Postema, G.T. (1982) Coordination and convention at the foundation of law, *The Journal of Legal Studies*, 11(1), pp. 165-203.
25. Rabin, M. & Schrag, J.L. (1999) First impressions matter: a model of confirmatory bias, *The Quarterly Journal of Economics*, 114(1), pp. 37-82.
26. Robins, M. (1984) *Promising, Intending and Moral Autonomy*. Cambridge: Cambridge University Press.
27. Savage, L. J. (1954) *The Foundations of Statistics*. New York: John Wiley & Sons.
28. Scanlon, T. (1990) Promises and Practices, *Philosophy and Public Affairs*, 19, 199-226.
29. Thomson, J.J. (1990) *The Realm of Rights*, Cambridge (MA): Harvard University Press.
30. Tummolini, L. & Castelfranchi C. (2007) Trace signals: The meanings of stigmergy. In D. Weyns, V. Parunak, and F. Michel (Eds.) *Environments for Multi-Agent Systems III*, number 4389 in Lecture Notes in Artificial Intelligence, pp. 141-156, Berlin/Heidelberg: Springer-Verlag.
31. Watson, G. (1975) Free agency, *Journal of Philosophy*, 72, pp. 205-220.