**Multilevel conflicts over forest governance under a game-theory perspective: the Great Bear Rainforest case examined.**

Sara Lorenzini[1],

## Abstract

*Forests are multifunctional collective goods (Frischmann, 2012; Chiappero, Von Jacobi and Fabbri, 2015) that exist at multiple spatial, temporal and jurisdictional scales (Geores, 2003). Because of the wide range of goods and services they provide, forests governance is usually at the root of multiple-level conflicts (Matiru, 2000; Bennett et al., 2001; Nelson et al., 2009; White et al., 2009) that, however, can be turned into a positive force for institutional innovation if properly handled.*

*By studying the Great Bear Rainforest (GBR) case (Armstrong, 2009; Cashore et al., 2011; Howlett et al., 2009; Raitio, 2012; Raitio and Saarikoski, 2013; Sranko, 2011) under a game-theory perspective, this paper aims at creating a model to demonstrate how cross-scale (Berkes, 2002) institutional interactions can become an opportunity to develop integrative conflict management solutions. Specifically, the focus is on the ex-post phase, that is compliance with the agreement.*

*With this objective, I apply Degli Antoni and Sacconi's model (2013) and demonstrate that as in their case, ex-post conformity can be explained by (a) strong stakeholders coming into play and (b) introduction of psychological utility (Geanakoplos, Pearce and Stacchetti, 1989; Rabin, 1993) that can shape conformity preferences (Grimalda & Sacconi, 2005;2007; Cecchini Manara & Sacconi, 2019b;2021). Additionally, through an extension of their model, I also show the crucial role of weakest parties' organizational capacity and mobilization. Importantly, these elements together allow for a new Psychological Nash Equilibrium to emerge that is not just environmentally conscious but also socially fair.*

[1] PhD Candidate in International and Public Law, Ethics and Economics for Sustainable Development at the University of Milan, Department of Italian and Supranational Public Law. Via Festa del Perdono 7, 20122 Milan. Email: Sara.lorenzini@unimi.it.

## Introduction

Forests are multifunctional infrastructures (Frischmann, 2012) that provide a wide range of goods and services, from the local to the global scale. They act as carbon sinks, they are home to biodiversity and they play a role in water and soil preservation. They are resources that also contain other resources such as food, medicinal plants, wood fuels and timber, which make them essential in both livelihood and economic terms. Social and cultural services' provision is another relevant function of forests, often contrasting with commercial activities.

Because of these multiple benefits, different actors have divergent interests over the same forest, which usually entails an increasing competition for appropriation. While specific conflict dynamics differ from one context to the other, similarities can be drawn by undertaking a definition of forests as collective goods (Chiappero, Von Jacobi and Fabbri, 2015) existing at multiple spatial, temporal and jurisdictional scales (Geores, 2003) and embodying multifunctionality.

Collective goods are Common Pool Resources (CPRs) (Ostrom & Ostrom, 1997; Ostrom, Gardner and Walker, 1994) that refer to specific collectivities placed at higher or lower levels, for which reason the protection of one collectivity may harm another one (Chiappero, Von Jacobi and Fabbri, 2015).

Especially in the case of large-scale areas such as the Amazon, it is particularly difficult to exclude non-contributors from using a forest or its resources, because of technical issues and high transactions costs. At the same time, if mis-managed, privately appropriated or over-exploited, a single collectivity's harvesting can subtract products from others' collectivities, making the resource rival. Importantly, as living organisms that interact with the physical environment as an interdependent system (Odum, 1971), forests are rival both in their stock – the forest itself – and in their flow– the derived assets – and the two aspects are strictly interrelated. Forest' subtractability reduces its capacity to generate benefits and, *vice versa*, flow's subtractability may intensively squander the forest.

Therefore, non-excludability becomes a property that must be preserved rather than an obstacle to efficient management; this point is important to move from simply studying forests' objective nature to consider normative arguments for a proper governance solution.

Indeed, due to the stock-flow and the components' interdependence, for a collectivity to be able to benefit from a specific good or service, other forests' functions shall be preserved as well. Moreover, for some collectivities such as Indigenous communities, accessibility to forests and their products is a matter of accessibility to something that has moral value, so that forests are also goods that "express benefits which are functional to the exercise of fundamental rights as well as to the free development of the person"[2]. Thus, non-excludability can be intended as an ethical property that must be preserved as precondition for the accessibility to an entitlement to perform valuable functionings according to what valuable mean for the collectivity itself (Sen, 1980, 1999, 2009; Anderson, 1999, 2010; Fia & Sacconi, 2018).

**Governing forests as global commons: multiple-level conflicts under a Game-Theory perspective**

As multifunctional resources existing at multiple scales, forest governance and management are usually at the root of multiple-level conflicts. Natural resources conflicts can be defined as "disagreements and disputes over access to, and control and use of, natural resources" (Matiru, 2000:1). They arise when users do not make the same use of the forest, have competing demands or different management priorities, and those of some groups are excluded from policies, programs and projects (Matiru, 2000) while at least one of the stakeholder collectivities is seen as affirming its interests at the disadvantage of one or more other groups (Bennett et al., 2001). Involved parties might also have the same claims but different views on how the resource should be distributed to meet their needs (White et al., 2009).

A collectivity's aspirations for using and managing a certain resource depends on its perception of it (Godelier, 1981), including higher-level principles and core values as well as technical, causal arguments (Sranko, 2011). For instance, part of the literature studies conflicts or cooperative situations over natural resources' use in terms of different feelings of psychological ownership, i.e a state in which individuals or collectivities perceive a certain target as being "their own" possession even thought it might not be so in legal terms (Matilainen & Lähdesmäki, 2009; Matilainen et al., 2017). Moreover, actors located at different scales might have different sources of behaviour and knowledge, different regulatory and policy systems and different socio-economic and political settings (Young, 2006). As a case in point, forest-dependent peoples and groups might have a higher discount rate than corporate actors or national governments when considering present costs *v.* future benefits and resource status (Scott, 1953). Similarly, Indigenous and local communities usually resort to traditional

[2] Draft Dlegation Law to the Government for the reform of Chapter II, Title I, Book III of the Civil Code, by the Rodotà Commission, 14 June 2007, Art. 3.1.c.

experiential knowledge and informal rules, while national and local administrations tend to privilege scientific forestry and formal, legally defined rules (Young, 2006).

Whatever their nature, scale, geographical scope and time coverage, conflicts imply an array of costs and risks, at the point that they can be described as "a lose-lose situation and a manifestation of governance failure" (Eckerberg & Sandström, 2013:1). The result is a negative impact on both efficiency and fairness: indeed, beyond being at the roots of forests deterioration and depletion, conflicts may result in unequal access to or abuse of the good by one collectivity on behalf of another. These dynamics, in turn, contribute to shape asymmetries in power and wellbeing (Chiappero, Von Jacobi and Fabbri, 2015).

However, if properly handled, conflicts situations can also give rise to innovative institutional solutions and collaborative governance forms. Since the root of these issues usually lays in decision-making problems over joint production of a number of competitive Ecosystem Services (Nelson et al., 2009)[3], transforming disputes is more about coordinating several interests and objectives rather than optimizing a single objective function. Accordingly, to take a game-theoretical perspective in analysing these interactions seems particularly insightful, in so far as game theory suggests that games representing social situations are essentially not pure-conflict, zero-sum games, but mixed interest ones (Harsanyi 1977). This category includes both cooperative and non-cooperative games. Concerning the latter, social situations mixing conflict and cooperation are represented by coordination games, where various coordination solutions (Nash equilibria) are mutually more advantageous than coordination failure outcomes but a conflict remains about which of the many coordination solutions to select, given that some players prefer one coordination solution whereas others prefer another one. Cooperative games, instead, are games in which the assumption is made that any achieved agreement will be enforced, so that the bargaining problem is about the selection of one amongst the many possible bargaining results or agreements. Consequently, the game represents at the same time the common interest to avoid the (mutually destructive) non-agreement exit and hence to achieve an agreement, but also a conflict of interest among the bargaining parties over which of the many possible bargaining results to agree upon.

[3] Clearly, material aspects (production of ESs) of the conflict are related to social and political ones, so that the game's outcome matters in terms of efficiency, equity and justice.

Given these premises, Game Theory (GT) constitutes a useful tool to study conflicts over forests. GT studies the mathematical modelling of strategic interactions among decision makers, especially when actions taken by a certain player affect others through social and environmental externalities (Parrachino et al., 2006). Indeed, a game can be described as any situation involving at least two players who recognize that their payoffs are influenced not only by what they decide but also by other players' strategy (Hargreaves Heap et al., 1992). Its basic elements are the players, their respective strategies and consequent payoffs, available information and specific rules.

The basic assumption of GT is instrumental rationality (Soltani et al., 2016; Osborne & Rubinstein, 1994) and its key concept is strategic interaction, in both its objective and subjective meanings. According to the former, consequences of A's decisions depend on what B does and *vice versa*. The latter instead refers to the fact that A expects rational decisions from B and she predicts that B will try to predict what she predicts and so on. Thus, reciprocal expectations also play a crucial role, because they represent interdependency: taking a simple game between two agents A and B, not only A's actions depend on B's ones, but A also has specific beliefs that B reasons similarly to her, which means A believes B knows that her own actions depend on A's actions. Thus, A's actions depend not only on what she thinks B will do but also on what she thinks B thinks about what A does, and *vice versa*.

However, because of players' inability of predicting others' behaviour despite assuming that others are rational and their rationality is common knowledge, there is the possibility of an infinite regression that means strategic uncertainty.

Undertaking a GT perspective, one might say that non-cooperation in conflicts over natural resources leads to a worse solution in terms of both players' payoffs – i.e pareto efficiency – and environmental quality. Indeed, while cooperative outcomes might bring better results, several problems exist in reaching them, such as lack of enforcement mechanisms and equilibrium property, social dilemmas (Zara et al., 2006), coordination problems and multiplicity of equilibrium solutions, and the need to select one of then consistently with some principle of distributive justice.

This paper applies GT mathematical modelling to the analysis of the historical agreement reached over the Great Bear Rainforest (GBR), the largest of undeveloped temperate rainforest in the world, situated in British Columbia (BC). Because this type of agreements over forests governance can be considered as legal institutions over deforestation, the overall aim of this work is to use GT to explain their ratio, when do they work and under which conditions compliance is feasible.

In particular, I start by introducing the case to explain how, after twenty years of conflict, in 2006 the multiple partners and stakeholders involved in this "War in the Woods" (Cashore et al., 2011) managed to reach a settlement by creating a shared vision and important commitments by all the parties (Section 1). Then, by using GT tools, I present an analytical framework of the story and a simplified representation of how the conflicting situation has been turned into collaboration through various steps (Section 2). Finally, the focus is on the *ex-post* (Gauthier, 1986) compliance with the agreement, on which the current literature is more silent (Section 3).

To model this stage, I take from Degli Antoni & Sacconi's work (2013) on "Social responsibility, activism and boycotting in a firm–stakeholders network of games with players' conformist preferences". Specifically, I take their model of a psychological game representing the relationship between a firm and its strong and weak stakeholders to demonstrate that as in their case, *ex-post* conformity can be explained by (a) strong stakeholders coming into play and (b) introduction of psychological games (Geanakoplos, Pearce and Stacchetti, 1989; Rabin, 1993) in which conformity preferences can be elicited by an impartial agreement and induce compliance equilibria (Grimalda & Sacconi, 2005;2007; Sacconi & Faillo, 2010; Cecchini Manara & Sacconi, 2019b;2021). Additionally, through an extension of the model, I also show the crucial role of weakest parties' organizational capacity and mobilization. Importantly, these elements together allow for a new Psychological Nash Equilibrium (PNE) to emerge that is not just environmentally conscious but also socially fair.

## 1. The Great Bear Rainforest Case

The "War in the Woods" started in the 1980s, when some of the First Nations (FNs) inhabiting the area together with environmental movements and grassroots organizations (ENGOs), started protesting against logging activities carried out by BC timber industries in old-growth forests.

For its part, the BC Provincial Government (ProvGov) tried to introduce a collaborative planning system - the Land and Resource Management Planning (LRMP) - but neither the FNs nor the ENGOs accepted to participate in the process. On the contrary, the former decided to pursue the legal path by bringing cases to Courts, while the latter started a big environmental campaign addressed at the BC forest industries and their international customers (Cus).

These strategies brought to a change in ProvGov and logging companies' attitude respectively. By 2001, the government began to sign protocols with the FNs so to enter into a "Government to Government" (G2G) relation, by which a commitment was undertaken for shared-decision making and negotiation of agreements over land use planning and resource management. The logging industries, instead, moved from an "attack and defend" (Raitio, 2012) strategy to preserve the status quo, towards searching for mutually acceptable solutions (Armstrong, 2009).

In the meanwhile, the coastal FNs, who had until then worked mostly independently of each other to promote their rights and title, started to establish formal coalitions such as the Coastal First Nations Turning Point Initiative (CFN) created by the end of 2001. At the same time, the relation between ENGOs and FNs became stronger: the former recognised their need to take FNs' concerns seriously while the latter realized they could benefit from environmental groups' market campaigns. Indeed, the campaigns contributed to put the region on the global agenda and created the visibility and pressure FNs needed to have more influence on the ProvGov. It is against this background that a critical meeting took place in 2001, where the FNs expressed their interest for ecological sustainability in the region, as long as the ENGOs were willing to recognize their rights and title and to take the issue of FNs' human well-being on their agenda (Raitio, 2012).

A concrete outcome of FNs' increased influence was the establishment of a new category of protected areas called Conservancies, which, in contrast to Class A Parks (hereafter Protected Areas or PAs), allow for the FNs' traditional uses (Howlett et al., 2009:389).

Moreover, as a result of increased international visibility, negotiations between ENGOs and timber companies brought to an agreement over basic principles for joint solutions' development and the suspension of respective activities that were at the roots of the conflict (Armstrong, 2009). Through the so-called Joint Solutions Project (JSP), companies and ENGOs agreed on the implementation of an Ecosystem-Based Management (EBM) plan for commercial forests and on the creation of the Coast Information team, an independent team of experts that would have provided knowledge and learning all along the process (Raitio, 2012).

In the end, the entire situation went again under the ProvGov's control, which brought key stakeholders to the formal planning table. As Raitio (2012) well explains, the ProvGov itself was strongly dependent on forest industries for income and employment and thus wished parties could arrive at a permanent decision to be communicated to markets. On the other side, industries, ENGOs and FNs needed certainty and cabinet's legitimacy, leadership and expertise. Moreover, failure to reach a consensus would have induced the risk of a unilateral decision by the ProvGov itself. The process culminated into BC Premier Gordon Campbell's announcement, in February 2006, that an historical agreement had been reached over the GBR. Full implementation of the Coast Land Use Decision and the system of EBM was finally reached in 2009 (Armstrong, 2009).

This case is an example of a complex conflict involving multiple partners and stakeholders, which was turned into collaboration thanks to a dialectical process that changed actors' perception of their interdependence as well as their incentive structures. In particular, the story shows the relevance of (a) weakest parties' ability to organise and mobilise, and (b) involvement of international customers and other nonstate actors operating at various scales (Bulkeley, 2005).

Indeed, FNs on the one hand and ENGOs on the other were able to form coalitions that strengthen their positions *vis-à-vis* other actors. Moreover, their strong mobilisation changed other players' best response, allowing them to develop more inclusive decision-making processes and reach their desired outcomes. Thus, the case demonstrates how cross-scale interactions (Berkes, 2002) can be turned into the basis for a new collaborative and integrated governance solution between public and private actors whereby no relevant stakeholders is excluded from certain governance rights and freedoms over the forests.

Moreover, the gist of this paper is that – as later explained – the process bringing to actual compliance with the provisions of the agreement can be captured in terms of a psychological game. Despite a form of agreement based on reciprocal understanding was missing at the beginning, so that each party could not develop necessary trust to believe that the other(s) will not take advantage of special relations to betray the deal, the agreement can later make parties develop preferences and beliefs that are necessary for its stability. This is true for both the agreement between NGOs and FNs on making common cause and, above all, for the final one. In other words, it is possible to see how (a) the agreement between ENGOs-FNs coalition on the one hand and E on the other, especially thanks to the involvement of external stakeholders

(Cus), and (b) the role of government as the one who places negotiating rules to the agreement, can influence the activation of dispositions to conformity.

## 2. From conflict to agreement: the GBR case explained through Game Theory

The very opening stage of the story (Phase 1), when there is no agreement, can be represented through three separate games. The first (Game 1.1) is between ENGOs (Player 1) and FNs (Player 2). By assumption, different ENGOs are considered to be one unique actor and different FNs another one. Both of them are rational decision makers who maximize their own objective function and the game is a simultaneous move game.

ENGOs' interest is mainly an environmental one: they want the creation of PAs where extractive activities are forbidden and forests remain undisturbed enough to regenerate and sustain wildlife population. FNs, instead, have traditionally inhabited the forest and used it for nearly every aspect of their everyday life, such as producing clothes and utensils, building shelters and religious objects or making ceremonies. Thus, they fear the creation of Class A Parks because of the evictions they would face. Instead, their most preferred outcome is the creation of Conservancies, where extractive activities are banned to protect and maintain both natural environments and biological diversity on the one hand and FNs' social, ceremonial, cultural and recreational values on the other. FNs, indeed, would have some rights over Conservancies.

In principle, the two players could arrive at a mutually advantageous deal. Indeed, even though FNs claim their rights over forests instead of a pure environmentally-protectionist deal as ENGOs would like, both of them want timber companies' activities to be reduced so that the GBR regenerative capacity is preserved. For instance, they might agree on a compromise solution entailing smaller PAs and Conservancies. Moreover, by converging on their strategies and allying among themselves, the two players might increase their bargaining power *vis-à-vis* more powerful actors like the ProvGov and the E.

Despite the possibility to behave cooperatively in order to reach an outcome of mutual advantage on some term of compromise amongst their preferences, the parties may be not ready to enter an agreement in the proper sense, i.e., to explicitly endorse an agreement on some common plan to be executed later on. A true agreement may appear as not credible to them since it is not guaranteed or protected enough by either legal constrains or any other enforcement mechanism and parties might not have enough foresight to see long-term effects of a collaboration between them positive enough to justify the costly development of such guarantees. Moreover, they may not be capable or simply not have the opportunity to identify the relevant terms of compromise by means a proper agreement understood as a process of discussion and deliberation such that parties reciprocally and effectively take the other's point of view in order to identify a common plan of action that satisfies whichever claims. And, if this expression may seem requiring too much with reference to many concrete bargaining situations, you should consider that the solution of bargaining games in general imposes an assumption of symmetry entailing that each party assumes that the counterparty will never concede something in order to reach an agreement that the first party herself would not never concede in the same situation (Harsanyi, 1977).

As later shown, if certain qualifications held, such processes of agreement are capable to develop preferences supporting conformity with the agreement itself. However, at this point of the story the involved parties do not have the opportunity to enter such a process, which means they would be compelled neither by external constraints nor by endogenous motivations to conform with the agreement. As a result, each player has in her mind a certain representation of the game (Aoki, 2001) according to which she would get a higher payoff from deviating when the counterpart instead believes that the first would cooperate. This corresponds to the pessimistic framing of the interaction situation, despite a mutual advantage can be obtained by cooperating according to some terms of compromise, each agent nevertheless has an incentive to deviate from the cooperative course of action if she believes that the counterparty will cooperate. For instance, if FNs trust that ENGOs will promote both PAs and Conservancies in their contracts with E, and therefore decide to delegate a part of their lobbying action to ENGOs, then ENGOs would in fact have an interest in promoting just PAs, i.e. in violating the agreement. *Vice versa*, if FNs acquire ENGOs' trust and ENGOs renounce to come to separate agreements with the other stakeholders, then ProvGov FNs would have an incentive to call only for Conservancies in their private negotiations with the. This is even more true if FNs are afraid that ENGOs do not actually stick to the agreement.

Given the premises, agents find themselves playing a non-cooperative PD. Each player must decide separately, and without the other knowing, whether to ally (A) or not (NA) with the counterpart before making any agreement with other stakeholders. Both are aware of their actions' consequences, represented by their respective payoffs in the matrix below (Table 1.1); the first number in each cell represents ENGOs' payoff while the second represents FNs' one.

**Table 1.1**

| | | FNs | |
|---|---|---|---|
| | | **A** | **NA** |
| **ENGOs** | **A** | (4;4) | (0;5) |
| | **NA** | (5;0) | (1;1) |

Importantly, what is at stake here is the decision between actually fulfilling a verbal agreement making it effective (cooperating or A) or not (defecting or NA). The former would be the best strategy for both players, since they would obtain more bargaining power against firms and ProvGov's interests. However, each of them can also decide not to ally (NA) and try to arrive to a separate agreement alone with other stakeholders.

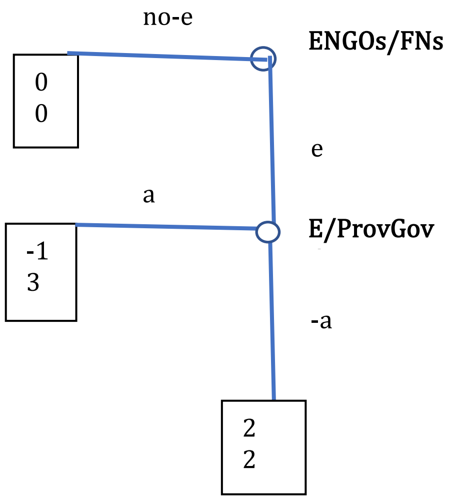
Because of the mental representation players have in their minds, defecting from cooperation is the only rational conduct for both players, even though a mutually advantageous agreement is evident and possible by playing (A;A). As a result, (NA;NA) becomes an equilibrium and it is in dominant strategies.

Later on, finding themselves in a situation of weakness and loneliness, each of them decides to pursue a separate strategy from the other: ENGOs start dealing with timber industries (E) while FNs enter into negotiations with the ProvGov. Both interactions can be modelled as non-simultaneous Trust Games (TG) where Player 1 (ENGOs and FNs) plays first and Player 2 (E and ProvGov) plays after. Games are represented in Figure 1.2 in their extensive form, while Table 1.3 summarises the normal form. Payoffs are shown for Player 1 and Player 2 respectively.

**Table 1.2**

| | | E / ProvGov | |
|---|---|---|---|
| | | a | -a |
| **ENGOs / FNs** | e | (-1;3) | (2;2) |
| | -e | (0;0) | (0;0) |

**Figure 1.3**




In general, a TG is a way to represent contractual relations or relations of authority whereby a player decides to give (or not) trust to the other party and legitimise its action by delegating her to take a decision, while the latter can either abuse or give Player 1 back a part of the surplus generated by their relation. In some cases, but not always, this implies accepting a formal subordination: the TG catches a common frame between contractual relations of delegation and authority relations.

In Game 1.2, the ENGOs can decide either to enter (e) or to stay out (-e): if ENGOs decide to enter the relation and give trust to E, the latter can abuse (a) or not (-a), by respecting or not its commitments. Trust has a value for E because the situation is like a trustor/trustee relation in which by entering the relation, the former delegates discretion to manage something of value – i.e. a capital of credibility – in exchange for the latter's commitment that she will manage it according to some common endeavor or interest of the two side. However, there is also the possibility that the trustee will pursue her own interest at the detriment of the trustor's desires. In other words, by trusting E, the ENGOs concede some discretion in the management of a capital of credibility that E may spend in market relations with its consumers as an element that facilitates consumers' decisions. Such discretion may be spent according to some common understanding, but it may also turn out to be exploitative allowing E to take a unilateral advantage in contradiction with the commitment undertaken with ENGOs.

In parallel, in Game 1.3, FNs can decide whether to delegate authority to the ProvGov (e) or not (-e), where this means giving or not trust by voting it as potential representative of FNs' interests. Since the ProvGov is the *de jure* owner of the forest, it may decide to abuse (a) and give more licenses to extract timber than the ones allowing FNs to pursue the functionings they value the most (Sen, 1980, 1999, 2009). Or, it might decide not to abuse (-a) and respect the electoral promises. On their part, E and ProvGov are interested in extracting as much as possible (E) and collecting fees and other voters' support (ProvGov).

In this case, the TG is a one-shot game, which means stakes are high but do not carry further repercussions like the possibility that the opponent defects in the last move or game to maximize its utility[4]. Moreover, players' motivations are only material, which means they are only interested in monetary payoffs.

In light of these elements, the TG has one unique equilibrium in (-e;a): Player 1 does not trust Player 2 and, expecting the latter will abuse, it decides to stay out. Clearly, the outcome is Pareto-inefficient because both players could be better off by playing (e;-a) (Fudenberg & Levine, 1989; Fudenberg & Tirole, 1991; Güth, Schmittberger & Schwarze, 1982; Kreps, 1990).

At a certain point of this history, however, relations between ENGOs and FNs start to change and, as a consequence, those between ENGOs-E and FNs-ProvGov as well (Phase 2). As Aoki's definition of games as "shared-beliefs-cum-equilibrium-summary" (Aoki, 2001:232) implies, games are mental representations exposed to limited rationality. Thus, under certain trigger mechanisms or elicitation, players' frame of the game can change. For instance, if conducted in a proper way (Habermas, 1981), communication can play a role in this shift.

In the case at issue, once ENGOs and FNs start communicating between them, agents are induced to re-assess and revise their own model of subjective game; thus, they possibly activate a new repertoire (Dosi and Marengo, 1994) of action choices (Aoki, 2001). As explained above, the former realize they need to take FNs' concerns seriously, while the latter recognize ENGOs' support can increase the case's visibility at the international level and have a strong reputational effect on businesses and the ProvGov, especially thanks to international customers' involvement. Because of this, players might change their own cognitive representation of the game's structure they play: this has an effect on their respective inference

[4] Clearly, E and ENGOs could also be seen as playing an iterated game, but in this case there is not so much to be taken from repetition because, at the beginning, E does not perceive ENGOs' cooperation as really valuable as social capital, i.e. E does not consider ENGOs as really able to influence its reputation – and therefore its market. The same is true for FNs-ProvGov relations: at first, the ProvGov does not consider particularly important FNs' vote, especially if FNs are not able to communicate with other parts of the electorate, which would be possible only by allying with ENGOs.

about the consequences of their choices (Aoki, 2001) and the payoff-dominant solution by agreement might become salient (Table 1.4).

**Table 1.4**

| ENGOs \ FNs | A | NA |
|---|---|---|
| **A** | (6;6) | (0;5) |
| **NA** | (5;0) | (1;1) |

Such forms of learning might let cooperation to emerge: both ENGOs and FNs start perceiving that by allying (A;A), they can obtain a mutual advantage. Thus, the game passes from being perceived as a PD to be seen as a Stag-Hunt game where players' utility from defecting is lower and the payoff dominant solution can prevail (Aoki, 2001; Cecchini Manara & Sacconi, 2019a). This outcome generates a shift in other players' incentives as well.

Indeed, businesses and the ProvGov recognize FNs and ENGOs' increasing bargaining power, visibility and influence not only in BC but worldwide. On the one hand, parallel negotiations between ENGOs-E and FNs-ProvGov go on: as Armstrong (2009:6) writes, "part of the New Relationship forged between the government of British Columbia and FNs is a commitment to creating new mechanisms for negotiating government-to-government agreements for shared decision-making in relation to land use planning and resource management". These G2G agreements aim at integrating the LRMP with FNs' knowledge, values and social-economic aspirations. Concerning ENGOs-E relation, instead, they undertook a joint commitment to suspend their market campaigns targeting companies operating in the region and hold up logging in more than one hundred intact landscapes in the GBR. Moreover, they agreed on basic principles upon which the so-called Joint Solutions Project (JSP) would have been built (Armstrong, 2009). On the other hand, the ProvGov retakes control over the entire process and the story (Phase 3) can be described as a cooperative game with mixed interests (Harsanyi, 1977:110) or what Luce & Raiffa (1957:1) define as a conflict of interests.

This process can be seen as the institutional trigger for a change in the game's perception. Previously, parties could expect to derive a mutual advantage from forming coalitions and pursuing bilateral relations but there was no attempt to come to a broader agreement between multiple stakeholders, i.e. a kind of social contract constituting common principles. This is up to the government to manage it. In turn, the possibility of coming to a multi-stakeholder agreement changes the perception of the other games: for instance, it comes to FNs and ENGOs' mind that the alliance is mutually advantageous because there is going to be a table where their increased bargaining power can actually count. Similarly, ENGOs and E can make agreements in the wider context, so that they do not see the problem as a TG anymore.

As in a conflict of interests, in principle players have a common interest in agreeing on a forest governance structure and they could promote it by coming to terms on a joint strategy for mutual gain. However, each of them has a certain preference over the specific terms of this structure and no full control over its implementation, because the resulting agreement is influenced by other players' decisions and assumes significance only if each participant can reasonably expect others' compliance. Thus, the situation is a mix of risk and uncertainty (Luce & Raiffa, 1957) that can be modelled in three ways: as (1) a bargaining game favoured by the government who allows other players to arrive at a solution they would have not arrived to; as (2) a problem of coalition formation; or as (3) an arbitration game where the government is the "social arbiter" of a deliberative process that can be compared to deliberation for an impartial agreement (Luce and Raiffa, 1957:9). In all these cases, the result will be the payoff vector of the game that maximizes the *n*-person Nash Bargaining Product (Harsanyi, 1977; Luce and Raiffa,1957).

Like in a bargaining game, players (ENGOs, FNs, ProvGov, E) must choose, from a payoff space P, a payoff vector $U = (u_{ENGOs}, u_{FNs}, u_E)$ that corresponds to a specific management structure to distribute rights over and utilities from the forest. The choice of vector U is such that the net product of the *status quo* $\prod_{j=1,2,3}^{n}(u_i - d_i)$ is maximized. Since not reaching the agreement means players would obtain the conflict payoff C without effectively satisfying their own interests, they have an incentive to arrive at it. Because of partially similar interests, players might also decide to form a big coalition and act as one bargaining unit. Indeed, by working jointly, each of them would obtain a higher payoff than by not cooperating (Harsanyi, 1977).

On the other hand, instead of attempting to come to an agreement by bargaining among themselves, players can decide to refer the case to an arbitrator, which would be the ProvGov. While the latter might follow its own value judgment, no solution can be imposed, because the agreement requires all parties' consent (Luce & Raiffa, 1957). Therefore, the ProvGov would work as an impartial outsider and guarantor that, by considering each player's preferences and possible strategies, suggests a conciliative and fair solution of the conflict promising it will enforce it.

Once an agreement is reached, however, questions on *ex-post* compliance with it remain open. Is the agreement self-enforcing? While the existing literature on the GBR case is largely focused on how the parties came to the agreement, conformity has received less attention. The last section of this paper is an attempt to formalize it through an extension of Degli Antoni and Sacconi's work on Corporate Social Responsibility (2013).

## 2. Compliance with the agreement: conformity preferences favouring the emergence of socially-fair behaviour.

The compliance or implementation problem deals with *ex post* rationality (Gauthier, 1986): it asks how can an agreement generate motivational causal forces strong enough to induce its execution in situations where at least one party could abuse it and obtain a better outcome in material terms. While in my work this question is referred to compliance with an agreement on a forest governance structure, the same issue has been largely discussed in the literature on Corporate Social Responsibility (CSR).

Interestingly, Degli Antoni and Sacconi (2013) explain conformity through the introduction of what behavioural economists call motivational complexity. In practice, their model considers agents who do not act solely on a self-interest basis but also have a sense of justice (Rawls, 1971) that translates into psychological utility for conformity with a principle of value, called T. Once conformity preferences are introduced, the game becomes a psychological game and new PNE can emerge, possibly more socially responsible. Moreover, the authors' model assigns a relevant role to the involvement of firms' strong stakeholders that can shape firms' incentives.

In the next paragraphs, I briefly introduce Degli Antoni & Sacconi's work (2013) explaining why I have decided to use it (par. 2.1) and how I have extended it so that it can better represent the GBR case (pars. 2.2 and 2.2.1). Finally, I conclude by analysing how the introduction of psychological utility in the players' utility functions can change the outcome of the game (par. 2.3).

*2.1Psychological Trust Game and conformity preferences: explaining ex-post compliance*

In their paper, Degli Antoni & Sacconi (2013) start by considering a formal model of a network (Lippert and Spagnolo, 2011) whereby an enterprise E is linked with a strong stakeholder ($S_s$) and two weak Stakeholders ($S_{w1}$ and $S_{w2}$) who are linked with other agents. While E and $S_s$ have a mutual relationship and prefer to cooperate forever, the relation between E and its weak stakeholders is unilateral and deficient: $S_{w1}$ and $S_{w2}$ have a high discount factor for future cooperation and thus an interest in repeated cooperation but E does not have the same material incentives, so that it has an interest in defecting. Taking this as a point of departure, the authors then set up their own model to analyse the games played within the network. Specifically, they keep the relation between E and $S_w$ as a PD, but they devise a three-players game G (Figure 2, Degli Antoni & Sacconi, 2013:224) to model the relation between E and its $S_s$.

In G, E and $S_s$ are active players while $S_w$ is a dummy one. $S_s$ has the faculty of deciding between entering (e) or not (-e) into a cooperative relation with E considering whether the latter will abuse (a) or not (-a). If $S_s$ plays e, both E and $S_s$ can act Fairly ($F_E$ and $F_{Ss}$) or Unfairly ($U_E$ and $U_{Ss}$) towards the dummy player, which means they must decide whether to collude and appropriate the entire surplus or to save a part for the $S_w$ who would then be able to use it in the following PDs with E.

Interestingly, Degli Antoni & Sacconi (2013) arrive at demonstrating that, as long as agents are endowed with only material preferences, (i) E and $S_s$ will always collude and (ii) E has no incentive to cooperate with its $S_w$. By introducing conformity preferences instead, mutual cooperative and fair behaviour can become endogenously sustainable: precisely, (i) $S_s$ might have an incentive to boycott firms that do not cooperate with $S_w$ and (ii) both $S_s$ and E might have incentives to cooperate with the $S_w$ in the long term.

Precisely, conformity preferences enter the picture when a) an agreement is reached in a pre-play communication phase over a distributive justice principle T, which for instance is a principle of equal treatment for the supply chains of companies like E, and b) this agreement activates reciprocal conformity beliefs. This last point refers to the fact that, in the same game, immediately after the agreement, both $S_s$ and E by default believe that the counterpart will adopt a strategy that fully realizes the distributive principle or it is strongly near to do so - as it would be crazy to have just agreed on a principle expressing a common intentionality but not having the same intention to do so even in the immediate, before any negative experience with implementation.

Once these two conditions held, a disposition to conformity to the principle T is activated that in the model is represented by the appearance, in E and $S_S$' payoffs, of a positive parameter of psychological utility $\lambda$ that adds to the material payoffs. If this parameter is high enough to constitute a sufficient incentive for effective compliance with the agreement, compliance strategies are a new equilibrium of the game.

**Figure 2**

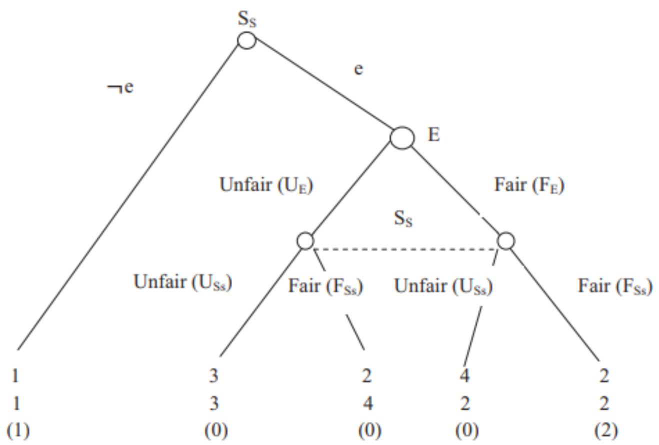



**Fig. 2.** The extensive form of game *G* representing the relationship between the firm and the strong stakeholder (the numbers in the column represent the payoffs obtained by $S_S$, $E$, $S_W$ respectively).

This model constitutes a good starting point to represent compliance with the agreement over the GBR case, for different reasons. First, it catches the essential role activism and boycotting play in shifting firms' incentives towards adopting socially responsible behaviours. Secondly, the distinction between $S_S$ and $S_W$ and the introduction of a dummy player are relevant to capture the different types of relationship E has with international customers, ENGOs and FNs. Thirdly, conformity preferences and psychological utility allows new PNE to emerge in the GBR case too, which – as I will demonstrate – are not only environmentally-fair but also socially-fair. This specification is of particular significance for the discourse on forests governance and sustainability, where false "green" solutions such as the creation of PAs are more and more criticized as new forms of racism and colonialism in the name of "environmental protection" (Mombeshora & Le Bel, 2009; Wilkie et al., 2006).

However, to properly represent this latter point and the whole GBR case, the model needs to be extended. In particular, as the next paragraph shows, the modified TG is a four players game where both E and $S_s$ can decide to behave unfairly or fairly but with an additional distinction between a mere environmentally-fair and a socially-fair strategy.

### *2.2 Extending the model*

As in the original model, my game is a kind of TG where there is a firm – the timber companies E operating in the GBR – that is linked with strong and weak stakeholders. Specifically, the $S_s$ is embodied by the intermediate Customers (Cus), that is pulp and other industries that transform the wood to sell final products in the market. They have a major interest in the running of the firm and, being indispensable for E's surplus production, they can influence it significantly (Freeman et al., 2010). FNs, instead, are $S_w$ for E: they directly participate E's transactions instead of simply undergoing external effects but because E could substitute them at affordable costs, they are not irreplaceable. Moreover, they are represented as dummy players, which means the game's outcome is insensitive to their strategy and they have no role in the agreement's implementation, even if its payoffs are determined by the others' choices and, therefore, represented. ENGOs, on their part, have a crucial role in the game: by extending Degli Antoni & Sacconi's model and considering ENGOs after they have made an alliance with FNs in the previous phases, I claim here that they are have the faculty of letting or not the game begin by entering (e) or not (-e) the agreement, where this means complying or contravening. Indeed, as in the real story, as long as they decided not to sit at the negotiating table, no agreement was reached and the conflict lasted for about 20 years.

In other words, ENGOs are not in a position to abuse but they have the power to decide not to conform to the agreement and negatively influence the possibility for $S_s$ and E to take advantage from their commercial partnership. This element catches the idea of boycotting, which means that $S_s$ depends on the interaction with ENGOs and only then, they can influence the exchange with E by exiting from the agreement. However, ENGOs' capacity of threatening the *status quo* or letting the game begin comes only as a result of their previous alliance with FNs.

This said, if ENGOs decide not to conform (-e) and go back to lobbying and campaigning against E, the *status quo* is kept and everyone obtains 1 because the positive payoffs associated with the agreement cannot be reached. If, instead, ENGOs decide to enter (e), the situation passes into $S_s$' hands that can in turn decide between entering (e) or not (-e) into relationship with the firm, by keeping commercial partnerships and buying timber from it or boycotting.

Again, if $S_s$ stays out the *status quo* is kept while if Ss enters, the game passes into E and $S_s$' hands who must then decide how to deal with the dummy stakeholder and the ENGOs. As in the original model, E and $S_s$ can collude and appropriate all the agreement's surplus as business partners or moderate their appropriation, but – choosing to play fairly – they must also decide whether to pursue just a green-fair (gf) strategy or a socially-fair (sf) one. Table 2.1 explains what each strategy means for each of the two players.

**Table 2.1**

| Strategy / Player | nf | Gf | sf |
|---|---|---|---|
| E | extract wood as much as possible | extract less and not extract in PAs because of environmental policies | extract less and not extract in PAs and Conservancies because of social-environmental policies |
| $S_s$ | buying as much as possible wood from E | buying less wood because of environmental policies (boycotting) | buying less wood because of socially and environmental policies (stronger boycotting) |

Then, depending on what E and $S_s$ play, each of the four players obtain a certain final payoff. In mathematical terms, final payoffs and their distribution depend on the total surplus S generated by compliance with the agreement. Given the amount X of extracted timber, S is calculated by summing up the monetary values *P* (profit) that E and $S_s$ obtain from extraction and transformation of timber in final commodities to be sold on the marketplace, to the positive externality U generated by reducing X. Thus,

$$S = P + U$$

where P is derived as the difference between the revenue V and production costs C, both of them assumed to be linear positive function of the quantity extracted X.

In general, the concept of *externality* refers to the fact that the production or consumption activities of a certain agent have a negative or positive effect on another, even though the latter does not receive any compensation for that (negative externality) or does not pay the cost of the activity (positive externality). In the case of timber companies' activities, one might be induced to think only in terms of negative social and environmental externalities. However, I here refer to U as positive externality because it represents the result of reducing the amount X of extracted timber. Thus, while V and hence P - until marginal costs do not equate marginal revenue - is an increasing function of X, U is a decreasing function of it. Importantly, U includes both social and environmental externalities, while the distribution is visible by looking at ENGOs and FNs' final payoffs. Depending on whether E and $S_s$ follow a gf or a sf strategy, a reduced X translates into more environmental protection (gf) and governance rights (sf) that go to the benefit of ENGOs and/or FNs.

More into details, E obtains P by selling wood to the $S_s$ while the latter derive it from the industrial transformation of wood into goods that are then sold to final consumers. Depending on E's and $S_s$' strategy, there are three levels of X, $X_{max} > X_{gf} >$ and $X_{sf}$ - corresponding to nf, gf and sf respectively – which result into $P_{max} > P_{gf} > P_{sf.}$

Table 2.2 represents the game in its normal form, while Figure 2.3 presents it in its extensive form; final payoffs are written for ENGOs, S, E and FNs respectively.

**Figure 2.3**

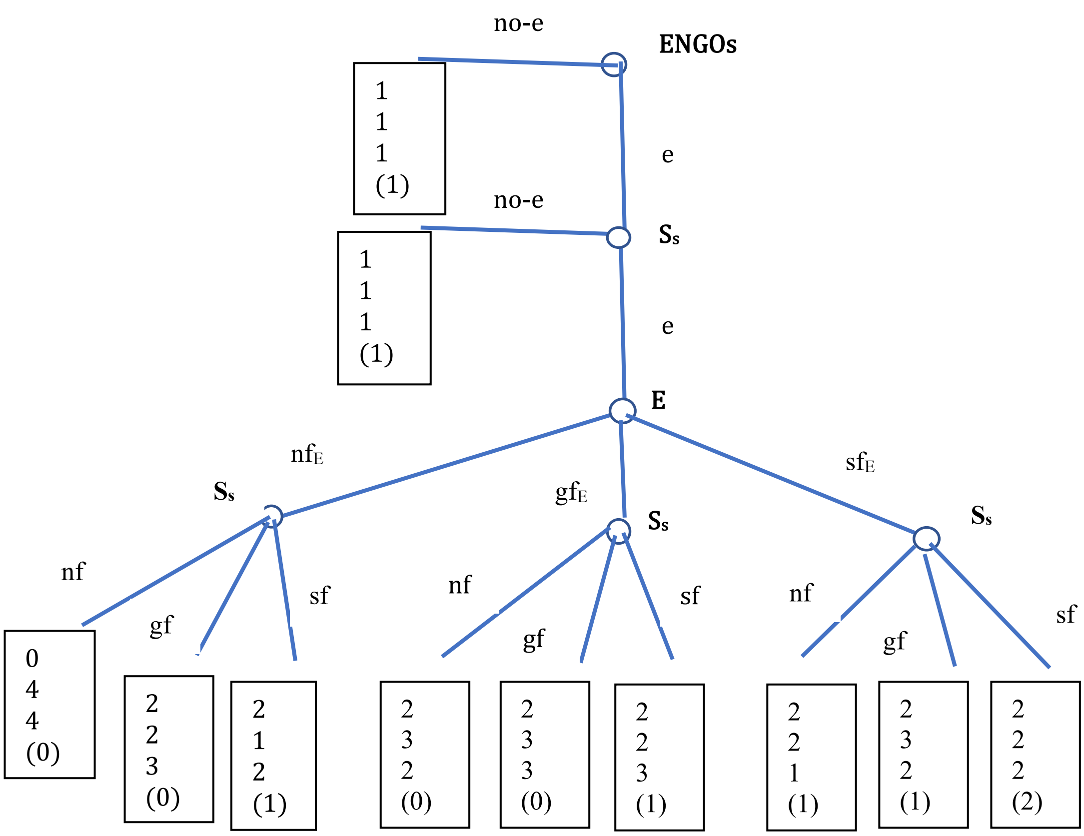


**Table 2.2**

**ENGOs** — 1;1;1;(1)

| **E** / $S_s$ | $nf_E$ | $gf_E$ | $sf_E$ |
|---|---|---|---|
| nf | 0;4;4;(0) | 2;3;2;(0) | 2;2;1;(1) |
| gf | 2;2;3;(0) | 2,3,3,(0) | 2;3;2;(1) |
| sf | 2;1;2;(1) | 2;2;3;(1) | 2;2;2;(2) |
| no-e | 1;1;1;(1) | 1;1;1;(1) | 1;1;1;(1) |

The model assumes that the sum between P and U is equal to 8 in the three main outcomes – when any extraction's reduction is excluded and profits are maximum, when there is a mutual restriction of extraction for environmental reasons that satisfy ENGOs but leave profits at a second best level and when there is an agreement of maximum reduction of the extraction that satisfies both environmental and FNs' rights protection. Thus, I exclude that there can be an improvement in terms of Pareto optimality, that is associated with environmental policies aimed at reducing wood extraction. In this context, compliance with the agreement is strictly contrary to E and $S_s$' interests.

If both E and $S_s$ decide to play nf, they obtain the maximum monetary value from extraction – $P_{max}$ equal to 8 in the numerical example – and appropriate it all. No positive externality U is produced and both ENGOs and FNs are left with a payoff equal to 0. When, instead, both E and Ss play gf, they renounce to a part of the market so that the total monetary value they obtain – $P_g$ equal to 6 – is reduced and equally distributed among them. A positive environmental externality U = 2 is generated that, however, goes only to ENGOs' benefit. It is only when both E and $S_s$ play sf that the total surplus – S equal to 8 in the numerical example – is equally distributed among the four players. Indeed, E and $S_s$ reduce X to $X_{sf}$ and equally split the monetary value among them – $P_{sf}$ equal to 4. The resulting environmental and social externality U = 4 goes at the advantage of both ENGOs and FNs.

However, the monetary value P obtained from extraction X is not always equally divided between E and Ss; indeed, when the two agents do not play the same strategy, one of them loses a higher market share or incurs in higher indirect costs. Thus, its payoff will be lower.

Take the case in which E plays $nf_E$ while $S_s$ plays gf or *vice versa*. Here, E does not comply with the agreement and extracts as much as it can, while $S_s$ adopts an environmentally-friendly strategy and renounce to buy part of the timber from E. Thus, E does not have to assume the direct costs of implementing a gf strategy but loses part of its market. However, because it can sell part of what is not bought by $S_s$ to $S_s$' competitors, its loss in only partial. Thus, E takes a higher portion – equal to 3 – of the resulting monetary value P equal to 5, while $S_s$, who incurs both direct costs and loss of competitiveness, is left with a payoff of 2. The contrary occurs when E plays $gf_E$ and $S_s$ plays nf: the former will face both direct costs from reducing X and loss of competitiveness because $S_s$ will buy wood from other timber companies, while the latter will have to renounce only to a part of its market. In both cases, the positive externality will be equal to the case in which both players play gf - U equal to 2 - and go entirely to ENGOs' benefit.

The same reasoning can be applied to the case where E plays $nf_E$ while $S_s$ plays sf or *vice versa*. However, here X is further reduced and so is P – equal to 3 in the numerical example – so that, in the first case, E gets 2 and $S_s$ gets 1 and *vice versa*. Because X and P are further reduced and social effects add to environmental ones, the resulting positive externality is higher – U equal to 3 – and ENGOs and FNs take 2 and 1 respectively, since ENGOs gain from both gf and sf while FNs gain only from the latter.

Finally, for the same reasons, when E ($S_s$) plays gf while $S_s$ (E) plays sf, the former takes a higher portion than the latter – equal to 3 and 2 respectively in the numerical example – and ENGOs obtain an higher portion of U.

Given this matrix and considering only material utilities, two Nash equilibria emerge.

- In a first hypothesis, ENGOs decide to stay out because they anticipate that by playing e, $S_s$ will play (e;nf) and E will play $nf_E$. Because the resulting distribution would be (0,4,4,(0)), ENGOs' best response is not to enter, which gives a payoff of (1,1,1,1).
- By looking at the normal form of the sub game between $S_s$ and E – where it is possible to assume a simultaneous move game – another possible Nash Equilibrium is in (e-gf; $gf_E$).
- First, assuming ENGOs have entered at the very beginning of the game, if $S_s$ plays (e, gf), E's best reply is to play $gf_E$ and *vice versa*: this is a weak Nash Equilibrium, because for each player the gf strategy is better than playing sf but indifferent to play non-fairly in

terms of payoffs. Thus, gf ($gf_E$) is the best response to the other player playing $gf_E$ (gf) but it is so in *weak sense* because there are indifferences with other strategic alternatives. However, this is a Nash equilibrium and, in fact, it is also true that in case a player replied to gf by adopting his nf strategy, the other player response would shift to nf, and so the result would therefore degenerate to the other equilibrium (e-nf ; $nf_E$). Considering now the ENGO's initial choice, in the case $S_s$ and E play (e-gf; $gf_E$), ENGOs' would confirm its strategy and play e.

Two observations are important. First, the extensive form of the game permits us to see that the second equilibrium does not satisfy the perfectness in the subgames condition. Assume players are playing the equilibrium (e; e-gf; $gf_E$) but E makes an error (trembling hand) and, with a small probability of 0.1, it plays $nf_E$ instead of $gf_E$. Should then $S_s$' equilibrium strategy be destabilized? Should $S_s$ shift to play nf in its second move? The answer is yes. Indeed, by confirming its gf strategy, $S_s$ will expect to obtain 0.1(2) + 0.9(3), which gives a lower payoff than what it would obtain by shifting to a nf strategy, which gives 0.1(4) + 0.9(3). Likewise, if E thinks $S_s$ will change its strategy with a probability of 0.1, going on with playing $gf_E$ will induce an expected payoff of 0.1(2) + 0.9(3) against 0.1(4) + 0.9(3) if it plays $nf_E$. Thus, the equilibrium is not stable in the subgame with respect to a disturbance in rationality and it is based on indifference and not payoff dominance, which confirms it is weak compared to (e-nf, $nf_E$). Expected payoffs of non-fair strategies under the hypothesis of trembling hand are strictly dominating over green-fair strategies. Thus, the equilibrium is not stable in the subgame with respect to a disturbance in rationality. This adds to what can be observed from the normal form, i.e. that this Nash equilibrium is weak. Secondly, (e, sf; $sf_E$) is the outcome with the highest Nash Bargaining product but it is not a Nash equilibrium of this non cooperative game, since for both Ss and E, gf ($gf_E$) is the best reply to the other player's playing sf ($sf_E$). Moreover, if we abstract from the dummy player FNs, the emerging equilibria are Pareto Optimal, which means that other players would be better by not caring about the weak stakeholder. The overall outcome can change only by introducing conformity preferences, that allow the emergence of new equilibria.

Overall, the mostly intriguing aspect of the game equilibrium set is that notwithstanding that it is a weak equilibrium, environmental collusion amongst E, $S_S$ and ENGOs is an available possibility of minor deviation from the agreement, that allows to avoid the stay-out equilibrium based on the prediction of complete defection on the part of E and Ss. This equilibrium witnesses the possible divorce between social justice and environmental protection which is the essential problem of environmental justice.

*2.2.1. Introducing psychological utility*

As in Degli Antoni and Sacconi (2013), the game I just presented is only the material basis for a psychological game PG where players have conformity preferences (Grimalda and Sacconi, 2005; 2007, Sacconi and Faillo, 2010) and related first and second-order beliefs.

The first step to model the PG is to identify the T value, that is the ideal principle of fair cooperation free and rational self-interested persons would accept in a pre-play communication stage where they reason "as if" they were under a veil of ignorance (Rawls, 1971). In our case, T represents an impartial agreement over forest governance, timber extraction and surplus' distribution.

By assumption (Aoki, 2001; Binmore, 2005; Sacconi, 2000, 2006a, 2010a), T is equal to the Nash Bargaining Solution (Nash, 1950;1953)

$$T(\sigma) = \prod_{i=1}^{n}(U_i - d_i)$$

Then, given T and because of their sense of justice (Rawls, 1971), players can obtain psychological utility by mutual compliance with T; this utility adds to the material one and modifies players' final payoffs. However, the incentive to conform is subject to players' mutual beliefs about reciprocal conformity; depending on this, new PNEs might emerge.

Thus, the second step is to include ideal utility and reciprocal beliefs in the agents' utility functions following the psychological game concept (Geanakoplos, Pearce and Stacchetti, 1989; Rabin, 1993). Each overall utility function $V_i$ is computed as

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

where $U_i(\sigma)$ represents the material utility player *i* attaches to the state of affair $\sigma$ while $\lambda_i$F[T(σ)] represents *i*'s conformity reason given her expectations of reciprocal conformity. $\lambda_i$ is an exogenous motivational parameter that measures a player *i*'s desire to conform with T while the function F[T($\sigma$)] captures beliefs' role in affecting ideal utility. Thus, given a certain state of affair $\sigma$, a system of beliefs b which includes player *i*'s conditional conformity index $f_i$ and a reciprocal expected conformity index $f_j$, F[T($\sigma$)] measures an agent's conformity to the agreed principle T, given what she expects the other player will play and what she expects the other expects from her. As explained more into details in Appendix A, this can also be expressed as

$$V_i(\sigma) = U_i(\sigma) + \Psi_i(\sigma, T, \lambda, b)$$

or, more explicitly, as

$$V_i(\sigma_i, b_i^1, b_i^2) = U_i(\sigma_i, b_i^1) + \lambda_i[1 + f_j(b_i^1, b_i^2)][1 + f_i(\sigma_i, b_i^1)].$$

Given the theory, I now apply it to the case at point by reconsidering the games described in Figs. 2.2 and 2.3 and introducing the assumption that ENGOs, E and $S_s$ are endowed with conformity preferences. To do so, I need first to assess T values for each state of the game, as shown in Table 2.4. Precisely, they are here computed for each state of affair belonging to the subgame wherein only $S_s$ and E participate, because ENGOs have first chosen to enter.

As the table shows, for each strategy played by $S_s$, the corresponding $T^{MAX}$ values engendered by E's choices are $T^{MAX}$(e,nf) = 4 obtained by $sf_E$, $T^{MAX}$(e,gf) = 12 obtained by $sf_E$, $T^{MAX}$(e,sf) = 16 obtained by $sf_E$ and $T^{MAX}$ (no-e) = 1 obtained indifferently by any E's response. A symmetrical situation holds for the $T^{MAX}$ values associated to the strategies $nf_E$, $gf_E$ and $sf_E$ of player E obtained by player $S_s$' responses. Thus, the highest degree of distributive fairness is reached in (e, sf; $sf_E$) and is equal to 16 in the numerical example.

**Table 2.4**

| $\sigma$ | T |
|---|---|
| (e, nf; $nf_E$) | 0 |
| (e, nf; $gf_E$) | 0 |
| (e, nf; $sf_E$) | 4 |
| (e, gf; $nf_E$) | 0 |
| (e, gf; $gf_E$) | 0 |
| (e, gf; $sf_E$) | 12 |
| (e, sf; $nf_E$) | 4 |
| (e, sf; $gf_E$) | 12 |
| (e, sf; $sf_E$) | 16 |
| (-e; $nf_E$) | 1 |
| (-e; $gf_E$) | 1 |
| (-e; $sf_E$) | 1 |

Using the formula, it is possible to compute $S_s$' overall conformity indexes and overall utility function for each state of affair; Tables 2.5 and 2.5A summarize the results again with respect to the subgame that is reached in case ENGOs enter and $S_s$ and E may then choose their moves. Appendix A contains details for calculation.

**Table 2.5**

| $\sigma$ | $(1 + f_{Ss})\ (1 + f_E)$ | F(T) |
|---|---|---|
| (e, nf; $nf_E$) | (1 - 1) (1 – 1) | 0 |
| (e, nf; $gf_E$) | (1 - 1) (1 - 1) | 0 |
| (e, nf; $sf_E$) | (1 - 1) (1 – 12/15) | 0.2 |
| (e, gf; $nf_E$) | (1 – 1) (1 – 1) | 0 |
| (e, gf; $gf_E$) | (1 – 1) (1 – 1) | 0 |
| (e, gf; $sf_E$) | (1 – 0,5) (1– 0,27) | 0.73 |
| (e, sf; $nf_E$) | (1 – 0) (1 – 0) | 0 |
| (e, sf; $gf_E$) | (1 – 0) (1 – 0,5) | 0.5 |
| (e, sf; $sf_E$) | (1 – 0) (1 – 0) | 1 |
| (-e; $nf_E$) | (1 – 0) (1 – 7/8) | 0.125 |
| (-e; $gf_E$) | (1 – 0) (1 – 11/12) | 0.08 |
| (-e; $sf_E$) | (1 – 0) (1 – 1) | 0 |

**Table 2.5A**

| $\sigma$ | $U_{Ss} + \Psi_{Ss} = V_{Ss}$ |
|---|---|
| (e, nf; $nf_E$) | 4 |
| (e, nf; $gf_E$) | 3 |
| (e, nf; $sf_E$) | $2 + 0{,}2\lambda_{Ss}$ |
| (e, gf; $nf_E$) | 2 |
| (e, gf; $gf_E$) | 3 |
| (e, gf; $sf_E$) | $3 + 0{,}73\lambda_{Ss}$ |
| (e, sf; $nf_E$) | 1 |
| (e, sf; $gf_E$) | $3 + 0{,}5\lambda_{Ss}$ |
| (e, sf; $sf_E$) | $2 + \lambda_{Ss}$ |
| (-e; $nf_E$) | $1 + 0.125\lambda_{Ss}$ |
| (-e; $gf_E$) | $1 + 0.08\lambda_{Ss}$ |
| (-e; $sf_E$) | 1 |

Through the same computation method, I compute E's overall utility function for each state of affair $\sigma$ of the subgame and hence the corresponding payoffs that are symmetrical to player $S_s$' ones, except for the strategy no-e. Indeed, if $S_s$ stays out, E obtains an overall utility equal to

$1+0{,}25\lambda_E$ from playing $nf_E$, $1+0.08\lambda_E$ from playing $gf_E$ and no ideal utility from playing $sf_E$. This might be interpreted as a reduced or a lack of psychological incentive for E to assume the costs that being environmentally or socially-fair implies if $S_s$, on its part, stays out of the game.

What emerges clearly from this analysis, albeit partial, is that depending on the absolute value $\lambda$ assumes, the calculus of players $S_s$ and E's best responses may change with respect to the game where only material payoffs were considered.

However, because psychological payoffs as illustrated so far pertain only to combinations of strategies in the subgame, deviation and conformity indexes as well as the resulting psychological utilities with various weights assumed by $\lambda$ should be interpreted as follows: when one of the players (E or $S_s$) predicts that the other ($S_s$ or E) will use a certain strategy *a* and thus responds with a strategy *b*, this occurs on the hypothesis that ENGOs have already played e. Thus, there is no need to compute the same indexes with respect to ENGOs' choice not to enter.

On the other hand, it is necessary to consider ENGOs' indexes for its two choices (e) and (no-e) with respect to each conceivable combination of $S_s$ and E's strategies. This holds both in the case $S_s$ and E are actually called into play by ENGOs' choice to enter and in the case that the no-entry choice by ENGOs does not allow $S_s$ and E's strategies to become effective.

Accordingly, I now consider all the possible state of affairs $\sigma$ of the overall game. At any rate, in order to favour the reader, I will continue to present them as cells of a matrix in which columns entries represent ENGOs' alternative choices to enter or not to enter while rows entries display each possible combination of $S_s$ and E's strategies considered as conjoint events, as indicated by the symbol of conjunction "&".

To start, let's now compute (Table 2.6) T values for each strategy followed by ENGOs under the hypothesis that each combination of $S_s$ and E's strategies is generated by their choices.

**Table 2.6**

| ENGOs / $S_s$ &E | e | no-e |
|---|---|---|
| (e,nf & $nf_E$) | T=0 | T=1 |
| (e,nf & $gf_E$ ) | T=0 | T=1 |
| (e,nf & $sf_E$) | T=4 | T=1 |
| (e,gf & $nf_E$) | T=0 | T=1 |
| (e,gf & $gf_E$) | T=0 | T=1 |
| (e,gf & $sf_E$) | T=12 | T=1 |
| (e,sf & $nf_E$) | T=4 | T=1 |
| (e,sf & $gf_E$) | T=12 | T=1 |
| (e,sf & $sf_E$) | T=16 | T=1 |
| (no-e & $nf_E$) | T=1 | T=1 |
| (no-e & $gf_E$) | T=1 | T=1 |
| (no-e & $sf_E$) | T=1 | T=1 |

Secondly, in Table 2.7, I compute

- for each strategy of ENGOs, conformity indexes $(1 + f_{ENGO})$ conditional on the hypothesis that $S_s$ and E use each of their strategies' combinations considered as conjoint events;
- indexes of expected and jointed conformity $(1 + f_{Ss,E})$ of players $S_s$ and E as estimated by ENGOs considering the hypothesis that $S_s$ and E use each of their combination under the second order hypothesis that ENGOs uses each of its two strategies alternatively;
- ENGOs' composite conformity indexes - conditional and expected - with respect to each possible combination of strategies by $S_s$ and E.

**Table 2.7**

| ENGOs / $S_s$ &E | e | | | no-e | | |
|---|---|---|---|---|---|---|
| | $(1 + f_{ENGOs})$ | $(1 + f_{Ss,E})$ | F(T) | $(1 + f_{ENGOs})$ | $(1 + f_{Ss,E})$ | F(T) |
| (e,nf & $nf_E$) | (1 – 1) | (1 – 1) | 0 | (1 – 0) | (1 – 0) | 1 |
| (e,nf & $gf_E$) | (1 – 1) | (1 – 1) | 0 | (1 – 0) | (1 – 0) | 1 |
| (e,nf & $sf_E$) | (1 – 0) | (1 – 0.75) | 0.25 | (1 – 1) | (1 – 0) | 0 |
| (e,gf & $nf_E$) | (1 – 1) | (1 – 1) | 0 | (1 – 0) | (1 – 0) | 1 |
| (e,gf & $gf_E$) | (1 – 1) | (1 – 1) | 0 | (1 – 0) | (1 – 0) | 1 |
| (e,gf & $sf_E$) | (1 – 0) | (1 – 0.25) | 0.75 | (1 – 1) | (1 – 0) | 0 |
| (e,sf & $nf_E$) | (1 – 0) | (1 – 0.75) | 0.25 | (1 – 1) | (1 – 0) | 0 |
| (e,sf & $gf_E$) | (1 – 0) | (1 – 0.25) | 0.75 | (1 – 1) | (1 – 0) | 0 |
| (e,sf & $sf_E$) | (1 – 0) | (1 – 0) | 1 | (1 – 1) | (1 – 0) | 0 |
| (no-e & $nf_E$ ) | (1 – 0) | (1 – 0.94) | 0.06 | (1 – 0) | (1 – 0) | 1 |
| (no-e & $gf_E$) | (1 – 0) | (1 – 0.94) | 0.06 | (1 – 0) | (1 – 0) | 1 |
| (no-e & $sf_E$) | (1 – 0) | (1 – 0.94) | 0.06 | (1 – 0) | (1 – 0) | 1 |

Thirdly, taking into account that ENGOs' psychological utility depends on a) their conditional conformity index to the alternative possible combinations of strategies that $S_s$ and E could choose and b) their expectation about $S_s$ and E's conformity through their different possible combinations of strategies in the hypothesis that ENGOs use the entry strategy or not, I consider (Table 2.8) the psychological payoffs that ENGOs obtain if

- they chose to enter, making several alternative assumptions about the possibility for $S_s$ and E to use each of their combinations of strategies, which become actually usable only after ENGOs' choice to play e;
- they chose the no-entry strategy under the hypothesis that the $S_s$ and E still chose to use some of their strategies' combination that would have become actually effective only if ENGOs had decided to enter and that, as a consequence, is made unfeasible by ENGOs' decision to stay out, with related consequences on ENGOs' conformity and $S_s$ and E's expected conformity level.

**Table 2.8**

| ENGOs / $S_s$ &E | e | no-e |
|---|---|---|
| (e,nf & $nf_E$) | 0 | $1+\lambda$ |
| (e,nf & $gf_E$) | 2 | $1+\lambda$ |
| (e,nf & $sf_E$) | $2+0{,}25\lambda$ | 1 |
| (e,gf & $nf_E$) | 2 | $1+\lambda$ |
| (e,gf & $gf_E$) | 2 | $1+\lambda$ |
| (e,gf & $sf_E$) | $2+0.75\lambda$ | 1 |
| (e,sf & $nf_E$) | $2+0.25\lambda$ | 1 |
| (e,sf & $gf_E$) | $2+0.75\lambda$ | 1 |
| (e,sf & $sf_E$) | $2+\lambda$ | 1 |
| (no-e & $nf_E$) | $1+0.06\lambda$ | $1+\lambda$ |
| (no-e & $gf_E$) | $1+0.06\lambda$ | $1+\lambda$ |
| (no-e & $sf_E$) | $1+0.06\lambda$ | $1+\lambda$ |

Finally, I consider psychological payoffs that the three players obtain under the hypothesis that ENGOs do not enter. It is important to remember that conformity indexes in psychological payoffs are conditional with respect to the possible combinations of strategies that all players could use. Thus, in this case, the conformity of ENGOs' no-e strategy depends on various hypotheses about the combination of strategies that $S_s$ and E would have used if ENGOs' choice not to enter had made it unfeasible. Moreover, the conformity of $S_s$ and E's combination of strategies depends on the T value determined as an effect of ENGOs' choice not to enter that impedes their choices to be effective, with respect to the alternative in which the same choices could have real consequences.

Clearly, psychological payoffs obtained as a result of ENGOs' no-entry choice are the same for each player, given that their material payoffs (1,1,1,1) are identical as is their degree of conformity resulting from ENGOs' decision not to enter. In Figure 2.9, these payoffs are represented in the upper matrix of the normal form of the game. The lower matrix instead represents the psychological payoffs when ENGOs decide to enter and $S_s$ and E's combination of strategies is therefore effective. Overall, Fig. 2.9 represents the psychological payoffs of the new psychological game PG'' and illustrates strategies' combinations of all players, for which the psychological parameter λ enters their payoffs.

Importantly, this transformation of the original game shows that players' utilities depend not only on objective consequences of choices but also on beliefs that agents have about others' behaviour and respective consequences. Thus, these beliefs also must be included in the model.

**Fig. 2.9**

| $S_s$ & E / ENGOs | e,nf & $nf_E$ | e,nf & $gf_E$ | e,nf & $sf_E$ | e,gf & $nf_E$ | e,gf & $gf_E$ | e,gf & $sf_E$ | e,sf & $nf_E$ | e,sf & $gf_E$ | e,sf & $sf_E$ | no-e |
|---|---|---|---|---|---|---|---|---|---|---|
| no-e | $1+\lambda$<br>$1+\lambda$<br>$1+\lambda$<br>(1) | $1+\lambda$<br>$1+\lambda$<br>$1+\lambda$<br>(1) | 1<br>1<br>1<br>(1) | $1+\lambda$<br>$1+\lambda$<br>$1+\lambda$<br>(1) | $1+\lambda$<br>$1+\lambda$<br>$1+\lambda$<br>(1) | 1<br>1<br>1<br>(1) | 1<br>1<br>1<br>(1) | 1<br>1<br>1<br>(1) | 1<br>1<br>1<br>(1) | $1+\lambda$<br>$1+\lambda$<br>$1+\lambda$<br>(1) |

**no-e**

**ENGOs**

**e**

| E / $S_s$ | $nf_E$ | $gf_E$ | $sf_E$ |
|---|---|---|---|
| e,nf | 0;4;4;(0) | 2;3;2;(0) | $2+0.25\lambda$; $2+0.2\lambda_{Ss}$;$1+0.2\lambda_E$; (1) |
| e,gf | 2;2;3;(0) | 2;3;3;(0) | $2+0.75\lambda$; $3+0{,}73\lambda_{Ss}$; $2+0.73\lambda_E$; (1) |
| e,sf | 2;1;2;(1) | $2+0.75\lambda$; $2+0.5\lambda_{Ss}$; $3+0.5\lambda_E$;(1) | $2+\lambda$; $2+\lambda_{Ss}$; $2+\lambda_E$;(2) |
| no-e | $1+0.06\lambda$;$1+0.125\lambda_{Ss}$;$1+0.25\lambda_E$;(1) | $1+0.06\lambda$;$1+0.08\lambda_{Ss}$;$1+0.08\lambda_E$;(1) | $1+0.06\lambda$;1;1;(1) |

Commenting the subgame represented by the inferior matrix of fig 3, the PG'' shows:

- $S_s$ obtains no ideal utility from playing (e, nf) given its first-order belief that E will play $nf_E$ or $gf_E$ and its second-order belief that E believes that $S_s$ will play (e, nf). Indeed, if E plays either $nf_E$ or $gf_E$, $S_s$ playing (e, nf) gives the worst T value – equal to 0 in the numerical example – with respect to the stay-out option – T equal to 1 – or the sf strategy option – T equal to 4 (for $nf_E$) and T equal to 12 (for $gf_E$).
- The same is true for $S_s$ playing (e, gf) given its first-order belief that E will play $nf_E$ or $gf_E$ and its second-order belief that E believes that $S_s$ will play (e, gf).

The effect of these minimizing choices with respect to the T value, is that the psychological parameter λ enters $S_s$ and E's payoffs when ENGOs decide to stay-out given they expect the other players chose these strategies' combinations. As explained below, the superior matrix of the game shows more into detail how the three players' payoffs change if this is the case.

Overall, these results are in line with the conformity preference model, according to which "staying out of an unfair cooperative relation can induce the relative best level of conformity if the other player's 'cooperative' choice is such that acceding to such a proposal of unfair cooperation or collusion would induce a lower level of implementation of the principle T" (Degli Antoni & Sacconi, 2011:184).

- $S_s$ obtains an intermediate value of ideal utility $\Psi_{Ss}$ equal to 0.73λ$s_s$ when it plays (e, gf) given its first-order belief that E will play $sf_E$ and its second-order belief that E believes that $S_s$ will play (e, gf) and $\Psi_{Ss}$ equal to 0.5λ$s_s$ when it plays (e, sf) given its first-order belief that E will play $gf_E$ and its second-order belief that E believes that $S_s$ will play (e, sf). Indeed, both cases bring to a T equal to 12, which is higher than T equal to 0 obtained when neither of them plays sf, T equal to 1 for the stay-out option and T equal to 4 for (e, sf; $gf_E$) and (e, gf; $sf_E$), but lower than T equal to 16 for (e, sf; $sf_E$).

Correspondingly, ENGOs obtain an intermediate value of ideal utility $\Psi_{ENGOs}$ equal to 0,75λ when their strategy is to enter.

- $S_s$ obtains the maximum value of ideal utility $\Psi_{Ss}$ equal to 1λ$s_s$, when it plays (e, sf) given its first-order belief that E will play $sf_E$ and its second-order belief that E believes that $S_s$ will play (e, sf). Indeed, this state of affair corresponds to the highest T value 16.

Correspondingly, ENGOs obtain the maximum value of ideal utility $\Psi_{ENGOs}$ equal to 1λ when their strategy is to enter.

- $S_s$ obtains low ideal utility $\Psi_{Ss}$ equal to 0.2λ$s_s$ from playing (e, nf) given its first-order belief that E will play $sf_E$ and its second-order belief that E believes that $S_s$ will play (e, nf) but no ideal utility from playing (e, sf) given its first-order belief that E will play $nf_E$ and its second-order belief that E believes that $S_s$ will play (e, sf). It is interesting to note that this occurs even though both (e, nf; $sf_E$) and (e, sf; $nf_E$) give T equal to 4, which means that there would be a certain degree of distributive fairness. In (e, sf; $nf_E$), this result might be interpreted as a lack of psychological incentive for $S_s$ to assume all the costs that being socially-fair implies if E, on its part, keeps acting unfairly.

Correspondingly, if they decide to enter, ENGOs obtain an intermediate value of ideal utility $\Psi_{ENGOs}$ equal to 0,75λ in the first case and no ideal utility in the second one.

- Finally, $S_s$ obtains a different ideal utility from staying out, depending on its first order belief about E's strategy and its second-order belief about what E believes $S_s$ will play. In the case $S_s$ plays no-e given its first order belief that E plays $nf_E$ and its second order belief that E believes $S_s$ plays no-e, $S_s$ obtains an ideal utility $\Psi_{Ss}$ equal to $0.125\lambda_{Ss}$. In the case $S_s$ plays no-e given its first-order belief that E plays $gf_E$ and its second-order belief that E believes $S_s$ plays no-e, $S_s$ obtains an ideal utility $\Psi_{Ss}$ equal to $0.08\lambda_{Ss}$. Finally, in the case $S_s$ plays no-e given its first-order belief that E plays $sf_E$ and its second-order belief that E believes $S_s$ plays no-e, $S_s$ obtains no ideal utility.

ENGOs, on its part, obtain an ideal utility $\Psi_{ENGOs}$ equal to 0,06λ when $S_s$ stays out regardless of what E plays. Thus, ENGOs would prefer $S_s$ to stay out than $S_s$ to enter and i) collude with E on a non-fair or green-fair strategies' combination; ii) play either nf or gf when E plays unfairly; iii) play unfairly when E adopts an environmentally fair strategy. Contrarily, ENGOs prefer all the other outcomes than $S_s$ staying out.

At this point, results for E's are symmetrical to those computed for $S_s$, with the only difference for the cases in which $S_s$ stays out. Indeed, when it predicts that $S_s$ stays out, E cannot do anything other than choosing one of its strategies to maximize the T value, but depending on which strategy it chooses, it will obtain a different psychological utility. Specifically, it will obtain a lower ideal utility from playing $nf_E$ - $\Psi_E$ equal to $0.25\lambda_E$ - than from playing $gf_E$ ($\Psi_E$ equal to $0.08\lambda_E$) and no ideal utility from playing $sf_E$. As said above, this might be interpreted as a reduced or a lack of psychological incentive for E to assume the costs that being environmentally or socially-fair implies if $S_s$, on its part, stays out.

By looking at the upper matrix of the normal form of the game, instead, it is possible to comment on ENGOs' psychological payoffs given their first-order beliefs about $S_s$ and E's conjoint choice. In particular, G'' shows that

- ENGOs obtain maximum ideal utility $\Psi_{ENGOs}$ equal to 1λ by staying out when they believe that if they enter, $S_s$ and E will then either i) collude on a non-fair strategies' combination or ii) play either gf and $nf_E$ or *vice versa*. Since, given the same expectations, playing e would instead give ENGOs no ideal utility, boycotting would be the best strategy. These, indeed, are the outcomes that minimize the T value to 0.
- ENGOs obtain maximum ideal utility $\Psi_{ENGOs}$ equal to 1λ by staying out when they believe that if they enter, $S_s$ and E will then collude on a green-fair strategies' combination.

Contrarily, given the same expectations, ENGOs would obtain no ideal utility from entering the game. This result is particularly interesting because it completely overturns the outcome of the game where only material utilities were considered: in that game, indeed, given the same expectation ENGOs enter the game and (e; e-gf; $gf_E$) is a - weak - Nash equilibrium of the game. It is therefore evident here how an agreement can activate reasons to start a relation whereby the surplus is created and shared in a fairer way.

- ENGOs obtain maximum ideal utility $\Psi_{ENGOs}$ equal to 1λ by staying out when they believe that, if they enter, $S_s$ stay outs. This is higher than what they would obtain by entering the game, that is $\Psi_{ENGOs}$ equal to 0,06λ, for any E's strategy.
- ENGOs obtain no ideal utility by staying out, if they expect i)E to play socially-fairly but $S_s$ to adopt either a nf or a green strategy, or ii)$S_s$ to adopt a sf strategy and E to adopts its environmentally or socially fair strategy. Since, instead, given the same expectations, entering would give them a ideal utility $\Psi_{ENGOs}$ equal to 0.25λ for (e; e-nf; $sf_E$), $\Psi_{ENGOs}$ equal to 0.75λ for (e; e-gf; $sf_E$), $\Psi_{ENGOs}$ equal to 0,75λ (e; e-sf; $gf_E$) and $\Psi_{ENGOs}$ equal to λ (e; e-sf; $sf_E$), here boycotting is not a preferred strategy for ENGOs. Indeed, in all these cases, ENGOs would maximize the T value by playing e rather than no-e.

Overall, it is possible to say that ENGOs obtain psychological utility from staying out if they predict that by entering, $S_s$ and E would then decide for a strategies' combination that determine low T values. Contrarily, as long as ENGOs predict that $S_s$ and E will determine high T values, their λ instead increases by playing e.

### *2.3 New Psychological Equilibria*

To sum up, when psychological utilities enter in the players' payoffs, the normal form of the game is modified as in Figure 3. Some important observations follow. First, ENGOs would be better off by staying out if $S_s$ and E collude on a non-fair strategies' combination; thus, when λ enters players' psychological payoffs, a PNE emerges where there was already a NE. This means that ENGOs impede the violation of the equity principle imposing a fairer outcome (1;1;1;1) and avoiding that $S_s$ and E together determine an high deviation index.

Secondly and differently from the game where only material payoffs are considered, (e; e, sf; $sf_E$) – corresponding to the maximum value of T – is now an additional PNE provided that players' motivational weight is high enough to counterbalance their temptation to adopt a nf or just gf strategy. In other words, the material utility each player derives from adopting a sf behaviour with the addition of the psychological parameter $\lambda$ must be higher than the material utility obtained from playing -e, nf ($nf_E$) and gf ($gf_E$)[5]. Moreover, this equilibrium is contingent on the existence of appropriate players' first and second order beliefs in mutual conformity with the ideal principle T. Thus, when λ is greater than -1, $\lambda_{Ss}$ is greater than 4 and $\lambda_E$ is greater than 2; ENGOs believe that $S_s$ and E play (e-sf; $sf_E$), $S_s$ believes that E plays $sf_E$ and E believes that $S_s$ plays (e;sf) and; each of them has a second (and higher) order belief that the other has exactly these beliefs, then (e, sf; $sf_E$) becomes a PNE. This is equal to say that ENGOs prefer not to boycott and $S_s$ and E prefer not to collude and ensure a fair division of the surplus that also goes at the benefit of FNs.

Importantly, once ENGOs enter the game because λ is greater than -1 and they have appropriate beliefs, if $\lambda_{Ss}$ is greater than 4 and $\lambda_E$ is greater than 2, (e, sf) and $sf_E$ constitute players' best replies also to the other playing $gf_E$ and gf. Precisely, $\lambda_{Ss}$ greater than 2 and $\lambda_{Ss}$ greater than 1,36 would be enough. As a result, the old environmentally but not socially fair equilibrium (e; e, gf; $gf_E$) is not an equilibrium anymore and a fairer and more cooperative one emerges.

Thirdly and related to this last point, the fact that, for certain values of λ, (e; e, gf; $gf_E$) is not an equilibrium outcome anymore, is confirmed by looking at the upper matrix of PG'', which shows that if $\lambda > 1$, ENGOs' best reply to an environmental collusion by the other players is to stay out of the game.

## Conclusions

After an introduction on why game theory can be used to study forest governance-related conflicts, where forests are considered as CPRs in terms of both descriptive and normative reasons, this paper has focused on the GBR case as an example of a long-lasting conflict that was turned into collaboration.

In particular, I have tried to depict the case using formal modelling that, by abstracting and simplifying the facts and the problem, is useful to extract the most conceptually relevant aspects. Then, following and extending Degli Antoni & Sacconi's (2013) model, I focused especially on the last stage of the case, that is to say compliance with the agreement on which, in fact, the current literature does not pay particular attention.

[5] Note that, since the material utility $S_s$ and E derived from playing nf is higher than the one derived from playing gf, computation can be done with reference only to the former.

The result confirms Degli Antoni & Sacconi's (2011) hypothesis that, assuming ENGOs enter the game, when E and $S_s$ make an effort to come to an agreement on general and abstract principles for forests' governance considering they could find themselves in the $S_W$'s position, they would formulate these principles by extending fair treatment also to the dummy player, who has no influence on the game. Thus, the resulting agreement will be consistent with the egalitarian solution (Binmore, 2005) that in my case corresponds to (e, sf; $sf_E$). This reasoning is in line with the conformity preference model (Grimalda & Sacconi, 2005;2007) according to which the motivational force able to counteract incentives to act unjustly comes from the sense of justice (Rawls, 1971) deriving from principles rationally agreed under a veil of ignorance.

Another crucial element emerging from the extension of model, is the essential role played by ENGOs' sanctioning behaviour associated with staying out, after having entered into a coalition with FNs. As seen, conformity preferences reinforce ENGOs' decision to stay out if they expect $S_s$ and E to collude on a non-fair strategies' combination and create an incentive to do so even when they expect $S_s$ and E to collude on a environmentally but not socially strategies' combination. These results show that ENGOs have the faculty to boycott and would actually impede the violation of the equity principle imposing a more cooperative and equitable outcome.

Therefore, the analysis is consistent with the idea that a SC is not just an *ex-ante* equilibrium selection device: indeed, it also shapes conformity preferences and, admitting proper reciprocal expectations, it creates endogenous motivations for *ex-post* compliance and stability. Said differently, collaboration on the basis of a fair principle for forest governance can be stable even though powerful actors and their stakeholders have no material interests in complying with the agreement. The disposition to conformity is particularly important because it works as an endogenous incentive that can induce ENGOs' boycotting and $S_s$' sanctioning behaviour with respect to E.

Finally, and particularly relevant for the discourse on natural resources management, sustainability and justice, the extended version of the model sheds light on (1) the crucial role of weakest parties' organizational and mobilization capacity and (2) the importance of distinguishing between conflict-management solutions that are purely environmental and those arrangements that, instead, display consideration for social justice as well.

Certainly, this work leaves a lot of space open for further research. For instance, it would be particularly interesting to develop, also on the basis of existing literature, a theory of compliance to derive institutional elements that can explain why certain regimes for forest governance emerge – or not – and how they are kept or abused. Then, real case-studies could be analysed to verify empirical elements that validate – or not – the theory. Moreover, considering the crucial role culture plays in this topic, it would be stimulating to investigate whether different $\lambda$ could have a – at least partial – cultural explanation. A comparative empirical analysis of conformity might present interesting results.

## Appendix A. Computation of psychological payoffs and ideal utility

The first step is to identify the value of T, that is the ideal principle of fair cooperation free and rational self-interested persons would accept under a veil of ignorance (Rawls, 1971). By assumption (Aoki, 2001; Binmore, 2005; Sacconi, 2000, 2006a, 2010a), T is equal to the Nash Bargaining Solution (Nash, 1950;1953).

$$T(\sigma) = \prod_{i=1}^{n}(U_i - d_i)$$

The formula says that, for each state of affairs $\sigma$, T is equal to the product of differences between the utility $U_i(\sigma)$ that each player $i$ could get from $\sigma$ and its reservation utility $d_i$ that is the utility she obtains if the agreement fails and agents find themselves in the *status quo*. $T^{MAX}$ is obtained when material utilities' distribution is in accordance with T.

Ideal utility is introduced here as distinct from the utility discussed till now, and it means that players may get ideal utility from mutual compliance and adherence to the principle, i,e. for the possibility of getting the principle implemented to some extent. Agents can get a positive ideal utility by mutual compliance with T and they will have an incentive to conform to it only insofar as they expect reciprocal conformity by other players. Thus, the second step is to include reciprocal beliefs in the agents' utility function following the psychological game concept (Geanakoplos, Pearce and Stacchetti, 1989; Rabin, 1993). By doings so, the overall utility function of each player $i$ results from the sum of her material utility $U_i$ and her ideal utility $\lambda_i F[T(\sigma)]$. In a formula:

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

$\lambda_i F[T(\sigma)]$ represents $i$'s conformity reason given her expectations of reciprocal conformity. $\lambda_i$ is an exogenous motivational parameter that measures player $i$'s desire to conform with T while the function $F[T(\sigma)]$ captures the beliefs' role in affecting ideal utility. Given a certain state of affair $\sigma$, a system of beliefs b which includes player $i$'s conditional conformity index $f_i$ and a reciprocal expected conformity index $f_j$, $F[T(\sigma)]$ measures an agent's conformity to the agreed principle T, given what she expects the other player will play and what she expects the other expects from her. In a formula:

$$F(T) = (1 + f_i\,(\sigma_i,b_i^1))\;(1 + f_j\,(b_i^1,b_i^2))$$

$f_i$ measures the extent to which, given what she expects $j$ will do, $i$ will comply with principle T. $f_j$ instead is an esteem of player $j$'s compliance with T, given what $j$ is expected to expect from $i$'s behaviour. Both indexes can take values that range from 0 (no conformity) to 1 (full conformity).

Given the state of affair $\sigma$, a first order belief $b_i^1$ that $i$ has in $j$'s behaviour and a second order belief $b_i^2$ about $j$'s belief in $i$'s behaviour, $f_i$ is computed as

$$f_i(\sigma_i, b_i^1) = \frac{\mathrm{T}(\sigma_i, b_i^1) - \mathrm{T}^{\mathrm{MAX}}(b_i^1)}{\mathrm{T}^{\mathrm{MAX}}(b_i^1) - \mathrm{T}^{\mathrm{MIN}}(b_i^1)}$$

where $\mathrm{T}(\sigma_i, b_i^1)$ is the value T assumes when player $i$ adopts strategy $\sigma_i$ and player $j$ plays what player $i$ believes ($b_i^1$) while $\mathrm{T}^{\mathrm{MAX}}$ ($b_i^1$) and $\mathrm{T}^{\mathrm{MIN}}$ ($b_i^1$) are the maximum and minimum values that T can assume depending on $j$'s strategy, given first and second order beliefs.

Similarly, given a first order belief $b_i^1$ and a second order belief $b_i^2$ about what player j believes about $i$'s strategy, $f_j$ is computed as

$$\tilde{f}_j(b_i^1, b_i^2) = \frac{\mathrm{T}(b_i^1, b_i^2) - \mathrm{T}^{\mathrm{MAX}}(b_i^2)}{\mathrm{T}^{\mathrm{MAX}}(b_i^2) - \mathrm{T}^{\mathrm{MIN}}(b_i^2)}$$

where $\mathrm{T}^{\mathrm{MAX}}$ ($b_i^2$) and $\mathrm{T}^{\mathrm{MIN}}$ ($b_i^2$) are the maximum and minimum values that T can assume depending on $j$'s first and second order beliefs.

Considering all these elements, it is now clear that the overall utility function $V_i$ is computed as a sum of player $i$'s material and psychological utility:

$$V_i\,(\sigma) = U_i\,(\sigma) + \Psi_i\,(\sigma, T, \lambda, b)$$

or, more explicitly:

$$V_i\,(\sigma_i, b_i^1, b_i^2) = Ui\,(\sigma_i, b_i^1) + \lambda_i[1 + f_j(b_i^1, b_i^2)][1 + f_i(\sigma_i, b_i^1)].$$

Now, considering the case at issue, table 1A summarizes the T value for each state of affairs, calculated as Nash Products – see the formula above - and taking (0;0;0;0) as the *status quo*.

**Table 1A**

| $\sigma$ | T |
|---|---|
| (e, nf; $nf_E$) | 0 |
| (e, nf; $gf_E$) | 0 |
| (e, nf; $sf_E$) | 4 |
| (e, gf; $nf_E$) | 0 |
| (e, gf; $gf_E$) | 0 |
| (e, gf; $sf_E$) | 12 |
| (e, sf; $nf_E$) | 4 |
| (e, sf; $gf_E$) | 12 |
| (e, sf; $sf_E$) | 16 |
| (-e; $nf_E$) | 1 |
| (-e; $gf_E$) | 1 |
| (-e; $sf_E$) | 1 |

By following the formula, it is now possible to calculate the two indexes and the overall index of conformity F(T) as well as the overall utility functions for player $S_S$.

For instance, $S_S$' deviation from full conformity for strategy (e, nf) is computed as:

$$f_{Ss}\,(e, nf; nf_E) = \frac{T(nf_E)-T^{MAX}(nf_E)}{T^{MAX}(nfE)-T^{MIN}(nf_E)} = \frac{0-4}{4-0} = -1$$

while E' expected deviation from full conformity for strategy $nf_E$ is computed as:

$$f_E\,(nf_E; e, nf) = \frac{T(nf_E;e,nf)-T^{MAX}(e;nf)}{T^{MAX}(e;nf)-T^{MIN}(e;nf)} = \frac{0-4}{4-0} = -1.$$

Thus, $S_S$' index of conditional conformity is equal to:

$$[1 + f_{Ss}(e, nf|f_E)] = 1 - 1 = 0$$

while the index of expected reciprocal conformity is:

$$[1+ f_E\,(nf_E|e, nf)] = 1 - 1 = 0$$

which entails $S_S$' overall conformity index equal to:

$$F(T) = (1 + f_{Ss})\,(1 + f_E) = (1\text{-}1)\,(1\text{-}1) = 0$$

and an overall utility function equal to:

$$V_{Ss} = U_{Ss} + \Psi_{Ss} = 4 + 0 = 4.$$

By applying the same computation method for each state of affair as well as to calculate E's psychological and overall utility functions, we obtain the new psychological game PG as presented in the text.

## BIBLIOGRAPHY


Aoki, M. (2001). *Toward a Comparative Institutional Analysis*. Cambridge, Massachussets: MIT Press.

Aoki, M. (2011). *Institutions as cognitive media between strategic interactions and individual beliefs*. In *Journal of Economic Behavior and Organization*, Vol. 79, pp. 20–34.

Armstrong, P. (2009). *Conflict resolution and the British Columbia's Great Bear Rainforest. Lessons learned 1995–2009*.

Bacharach, M. (2006). *Beyond Individual Choice. Teams and Frames in Game Theory*. Princeton University Press, Princeton.

Bennett, E., Neiland, A., Anang, E., Bannerman, P., Atiq Rahman, A., Huq, S., Bhuiya, S., Day, M., Fulford-Gardiner, M. and Clerveaux, M. (2001). *Towards a better understanding of conflict management in tropical fisheries: evidence from Ghana, Bangladesh and the Caribbean*. In *Marine Policy*, Vol. 25, pp. 365-376.

Berkes, F. (2002). *Cross-Scale Institutional Linkages: Perspectives from the Bottom Up*. In Ostrom, E., Dietz, T., Dolsak, N., Stern, P. C., Stonich, S. and Weber, E. *The Drama of the Commons*, pp. 233- 257, National Research Council. Washington DC, USA: National Academy Press.

Binmore, K. (2005). *Natural Justice*. Oxford: Oxford University Press.

Bulkeley, H. (2005). *Reconfiguring environmental governance. Towards a politics of scales and networks*. In *Political Geography*, Vol. 24, pp. 875–902.

Cashore, B., Hoberg, G., Howlett, M., Rayner, J., Wilson, J., (2001). In *Search of Sustainability: British Columbia Forest Policy in the 1990*. Vancouver, Canada: University of British Columbia Press.

Cecchini Manara, V. and Sacconi, L. (2019a). *Institutions, Frames and Social Contract Reasoning*. In *EconomEtica*, N. 71.

Cecchini Manara, V. and Sacconi, L. (2019b). *Compliance with socially responsible behaviour: reputation v. conformity*. In *EconomEtica*, N. 73.

Cecchini Manara, V. and Sacconi, L. (2021). *The motivational role of ex ante agreements: a psychological game approach*. Mimeo.

Degli Antoni, G. and Sacconi, L. (2011). *Modeling Cognitive Social Capital and Corporate Social Responsibility as Preconditions for Sustainable Networks of Relations*. In Degli Antoni, G. and Sacconi, L. (Eds.). *Social Capital, Corporate Social Responsibility, Economic Behaviour and Performance*, pp. 161-239. UK: Palgrave Millan.

Degli Antoni, G. and Sacconi, L. (2013). *Social Responsibility, activism and boycotting in a firm-stakeholders network of games with players' conformist preferences*. In *The Journal of Socio-Economics*, Vol. 45, pp. 216-226.

Derkyi, M., Ros-Tonen, M. A., Kyereh, B and Dietz, T. (2014). *Fighting Over Forest: Toward a Shared Analysis of Livelihood Conflicts and Conflict Management in Ghana*. In *Society and Natural Resources*, Vol. 27, pp. 281–298.

Deutsch, M., 2006. *Cooperation and competition*. In Deutsch, M., Coleman, P.T. and Marcus, E.C. (Eds.). *The Handbook of Conflict Resolution*, pp. 23-42. San Francisco, US: Jossey Bass.

Dosi, G., and L. Marengo. 1994. Some elements of an evolutionary theory of organizational competence. In England, W. (Ed.). *Evolutionary Concepts in Contemporary Economics*. Ann Arbor, Michigan, USA: University of Michigan Press, pp. 157–78. Reprinted in Dosi (2000):211.

Duncan Luce, R and Raiffa, H. (1957). *Games and Decisions. Introduction and Critical Survey*. Garden City, New York, USA: Dover Publications, Inc.

Eckerberg, K and Sandström, C. (2013). *Forest conflicts: A growing research field*. In *Forest Policy and Economics*, Vol. 33, p. 3-7.

Fia, M. & Sacconi, L. (2018). *Justice and Corporate Governance: New Insights from Rawlsian Social Contract and Sen's Capabilities Approach*. In *Journal of Business Ethics*, Vol. 160, pp. 937-960.

Frischmann, B.M. (2012). *Infrastructure - The Social Value of Shared Resources*. Oxford/New York, USA: Oxford University Press.

Fudenberg, D. and Levine, D. K. (1989). *Reputation and Equilibrium Selection in Games with a Patient Player*. In *Econometrica*, Vol. 57, No. 4, pp. 759-778.

Fudenberg, D., and Tirole, J. (1991). *Game Theory*. Cambridge, Massachusetts: MIT Press.

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

Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989). *Psychological games and sequential rationality*. In Games and Economic Behavior, Vol. 1, pp. 60–79.

Geores, M. E. (2003). *The Relationship between Resource Definition and Scale: Considering the Forest*. In Dolsak, N., Ostrom, E. and McKay, B. J. (2003). *The Commons in the New Millennium: Changes and Adapation*, pp. 77-97 Cambridge, Massachusetts: MIT Press.

Gilbert, M. (2013). *Joint Commitment*. New York, USA: Oxford University Press.

Godelier, M. (1981). *Instituciones económicas*. Barcelona, Spain: Editorial Anagrama.

Grimalda G. and Sacconi, L. (2005). *The Constitution of the Not-For-Profit Organization: Reciprocal Conformity*. In *Constitutional Political Economy*, Vol. 16, No. 3, pp. 249–276.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). *An experimental analysis of ultimatum bargaining*. In *Journal of Economic Behaviour & Organization*, Vol. 3, Issue 4, pp. 367–388.

Habermas, J. (1981). *The Theory of Communicative Action. Vol. 1. Reason and the Rationalization of Society; Vol. 2. Lifeworld and System*. Translated by Thomas McCarthy, Boston, Massachusetts, USA: Beacon Press.

Harsanyi, J. C. (1977). *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*. New York, USA: Cambridge University Press.

Heargraves Heap, S., Hollis, M., Lyons, B., Sugden, R. and Weale, A. (1992). *The theory of choice. A critical guide*. Cambridge, Massachusetts: Blackwell Publishers.

Kabir, K. H., Knierim, A. and Chowdhury, A. (2020). *No forest, no dispute: the rights-based approach to creating an enabling environment for participatory forest management based on a case from Madhupur Sal Forest*, Bangladesh. In *Journal of Environmental Planning and Management*, Vol. 64, No. 1, pp. 22–46.

Kreps, D. M. (1990). *Corporate culture and economic theory*. In *Perspectives on Positive Political Economy*, Alt, J. E. and Shepsle, K. A. (Eds.). Cambridge: Cambridge University Press, pp. 90–143.

Martinetti-Chiappero, E., Von Jacobi, N. and Fabbri, M. (2015). *Sustainability: Pillars of Social Sustainability*. Annale Fondazione Giangiacomo Feltrinelli 49, pp. 95-121.

Matilainen, A. and Lähdesmäki, M. (2009). *Nature-based entrepreneurship in private forests – The preconditions for the sustainable co-operation between private forest owners and entrepreneurs*. University of Helsinki Ruralia Institute, Report 48.

Matilainen, M., Pohja-Mykrä, M., Lähdesmäki, L and Kurki, S. (2017). *"I feel it is mine!". Psychological ownership in relation to natural resources*. In *Journal of Environmental Psychology*, Vol. 51, pp. 31-45.

Matiru, V. (2000). *Conflict and Natural Resource Management*. In Hart, N. and Castro, P. (Eds). *Sustainable Forest Management (SFM) Toolbox*, FAO. Rome, Italy.

Mombeshora, S. & Le Bel, S. (2009). *Parks-people conflicts: the case of Gonarezhou National Park and the Chitsa community in south-east Zimbabwe*. In *Biodiversity and Conservation*, Vol. 18, pp. 2601-2623.

Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D. R., Chan, K. M. A., Daily, G. C., Goldstein, J., Kareiva, P. M., Lonsdorf, E., Naidoo, R., Ricketts, T. H. and Shaw, R. M. (2009). *Modeling Multiple Ecosystem Services, Biodiversity Conservation, Commodity Production, and Tradeoffs at Landscape Scales*. In *Frontiers in Ecology and the Environment*, Vol. 7, No. 1, pp. 4-11.

Parrachino, I., Zara, S. and Patrone, F. (2006). *Cooperative Game Theory and its application to natural, environmental and water resources issues. 1. Basic Theory*. World Bank Policy Research Working Paper 4072.

Parrachino, I., Dinar, A. and Patrone, F. (2006). *Cooperative Game Theory and its application to natural, environmental and water resource issues. 3. Application to Water Resources.* World Bank Policy Research Working Paper 4074.

Osborne, M. J. and Rubinstein. (1994). *A course in Game Theory*. Massachussets Institute of Technology, Cambridge, Massachusetts: MIT Press.

Ostrom, E., Gardner, R. and Walker, J. M. (1994). *Rules, Games and Common Pool Resources*. Ann Arbor: University of Michigann Press.

Ostrom, E. & Ostrom, V. (1977). *A theory for institutional analysis of common pool problems*. In Hardin, G. and Baden, J. (Eds.). *Managing the Commons*, pp- 157-172. San Francisco, W. H. Freeman.

Rabin, M., 1993. *Incorporating fairness into game theory and economics*. In *American Economic Review*, Vol. 83, No. 5, pp. 1281–1302.

Raitio, K. (2012). *Discursive institutionalist approach to conflict management analysis — the case of old-growth forest conflicts on state-owned land in Finland*. In *Forest Policy and Economics*, Vol. 33, pp. 97-103.

Raitio, K. and Saarikoski, H. (2013). *Governing old-growth forests: the interdependence of actors in Great Bear Rainforest in British Columbia*. In *Society and Natural Resources*, Vol. 25, Issue 9, pp. 900–914.

Reiter, R. (1980). *A logic for default reasoning*. In *Artificial Intelligence*, Vol. 13, pp. 81–132.

Sacconi, L. and Grimalda, G. (2007). *Ideals, Conformism and Reciprocity: A Model of Individual Choice with Conformist Motivations, and an Application to the Not-for-Profit Case*. In Porta, P. L. and Bruni, L. (Eds.) *Handbook of Happiness in Economics*. Cheltenham Northampton, Massachusetts, USA: Elgar Publishing, pp. 532-570.

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

Satyal Pravat, P. and Humphreys, D. (2012). *Using a multilevel approach to analyse the case of forest conflicts in the Terai, Nepal*. In *Forest Policy and Economics*, Vol. 33, pp. 47-55.

Scott, A. (1953). *Natural resources: the economics of conservation.* Toronto, Ontario, Canada: University of Toronto Press.

Sen, A. (Ed.). (2009). *The Idea of Justice*. US: Harvard University Press.

Shahi, C. and Kant, S. (2007). *An evolutionary game-theoretic approach to the strategies of community members under Joint Forest Management regime*. In *Forest Policy and Economics*, Vol. 9, pp. 763–775.

Soltani, A., Sankhayan, P. L. and Hofstad, O. (2016). *Playing forest governance games: State-village conflict in Iran*. In *Forest Policy and Economics*, Vol. 73, pp. 251-261.

Sranko, G. R. (2011). *Collaborative governance and a strategic approach to facilitating change: lessons learned from forest agreements in South East Queensland and the Great Bear Rainforest*. In A journal for and about social movements, Vol. 3, No. 1, pp. 210-239.

Young, O. 2006. *Vertical interplay among scale-dependent environmental and resource regimes*. In *Ecology and Society*, Vol. 11, No. 1.

Zachrisson, A. and Beland Lindhal, K. (2013). *Conflict resolution through collaboration: Preconditions and limitations in forest and nature conservation controversies*. In *Forests Policy and Economics*, Vol. 13, pp. 39-46.

Zara, S., Dinar, A. and Patrone, F. (2006). *Cooperative Game Theory and its application to natural, environmental and water resource issues. 2. Application to Natural and Environmental Resources*. World Bank Policy Research Working Paper 4073.

White, R. M., Fischer, A., Marshall, K., Travis, J. M., Webb, T. J., Di Falco, S., Redpath, S. M. and Van der Wal, R. (2019). *Developing an integrated conceptual framework to understand biodiversity conflicts.* In *Land Use Policy*, Vol. 26, Issue 2, pp. 242-253.